Essays on travel mode choice modeling : a discrete choice approach of the interactions between economic and behavioral theories

Hélène Bouscasse

▶ **To cite this version:**

Hélène Bouscasse. Essays on travel mode choice modeling : a discrete choice approach of the interactions between economic and behavioral theories. Economics and Finance. Université de Lyon, 2017. English. NNT : 2017LYSE2106 . tel-01661687v2

HAL Id: tel-01661687
https://theses.hal.science/tel-01661687v2

Submitted on 12 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° d'ordre NNT : 2017LYSE2106

THESE de DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de

L'UNIVERSITÉ LUMIÈRE LYON 2

École Doctorale : ED 486 Sciences Économique et de Gestion

Discipline : Sciences économiques

Soutenue publiquement le 9 novembre 2017, par :

Hélène BOUSCASSE

Essays on travel mode choice modeling.

A discrete choice approach to the interactions between economic and behavioral theories.

Devant le jury composé de :

Véronique SIMONNET, Professeure des universités, Université Grenoble Alpes, Présidente
Laurent DENANT-BOEMONT, Professeur des universités, Université de Rennes, Rapporteur
André DE PALMA, Professeur des universités, École Normale Supérieure de Cachan, Rapporteur
Catherine VIOT, Professeure des universités, Université Lyon 1, Examinatrice
Michel BIERLAIRE, Professeur des universités, École Polytechnique Fédérale de Lausanne, Examinateur
Jacques-François THISSE, Professeur d'université, Université Catholique de Louvain, Examinateur
Patrick BONNEL, Docteur Ingénieur, École Nationale des Travaux Publics de l'État, Directeur de thèse
Iragaël JOLY, Maître de Conférences HDR, Institut National Polytechnique de Grenoble, Co-Directeur de thèse

Contrat de diffusion

Ce document est diffusé sous le contrat *Creative Commons* « **Paternité – pas d'utilisation commerciale - pas de modification** » : vous êtes libre de le reproduire, de le distribuer et de le communiquer au public à condition d'en mentionner le nom de l'auteur et de ne pas le modifier, le transformer, l'adapter ni l'utiliser à des fins commerciales.

LABORATOIRE AMÉNAGEMENT ÉCONOMIE TRANSPORTS | TRANSPORT URBAN PLANNING ECONOMICS LABORATORY

UNIVERSITÉ DE LYON

École Doctorale 486 Économie et Gestion

THÈSE

pour obtenir le grade de docteur délivré par

UNIVERSITE LUMIERE LYON 2

présentée et soutenue publiquement par

Hélène BOUSCASSE

le 9 novembre 2017

Essays on travel mode choice modeling

A discrete choice approach to the interactions between economic and behavioral theories

Jury

Patrick BONNEL,	Ingénieur en Chef HDR05, ENTPE, Université de Lyon	Directeur de thèse
Iragaël JOLY,	Maître de Conférences HDR05, Grenoble-INP	Co-directeur de thèse
Laurent DENANT-BOEMONT,	Professeur, Université de Rennes	Rapporteur
André DE PALMA,	Professeur, École Normale Supérieure de Cachan	Rapporteur
Michel BIERLAIRE,	Professeur, École Polytechnique Fédérale de Lausanne	Examinateur
Véronique SIMONNET,	Professeur, Université Grenoble Alpes	Examinatrice
Jacques-François THISSE,	Professeur, Université Catholique de Louvain	Examinateur
Catherine VIOT,	Professeur, Université de Lyon	Examinatrice

A Olivier, Nathanaël et Gabriel

Merci pour votre amour, votre patience et votre gentillesse

Remerciements

Les premières personnes que je tiens à remercier sont mon directeur de thèse, Patrick Bonnel, et mon co-directeur de thèse, Iragaël Joly. Je vous remercie de m'avoir accompagnée durant ces trois années de thèse et de m'avoir introduit à l'univers passionnant des transports. Vous avez su me faire bénéficier de votre expérience et de vos connaissances, m'encourager et m'appuyer lorsque j'en avais besoin, tout en m'accordant une grande confiance et liberté que j'ai beaucoup appréciées. Votre regard critique, vos remarques rigoureuses et conseils judicieux ont largement contribué à la réalisation de ce travail.

Je remercie ensuite les autres personnes avec lesquelles j'ai eu l'occasion de travailler pendant la thèse. Je pense notamment à Aurélie Bousquet qui m'a aidée sous SQL à personnaliser l'enquête. Et évidemment, mes remerciements les plus chaleureux vont à Matthieu de Lapparent dont les connaissances économiques et économétriques m'ont beaucoup apporté et à Jean Peyhardi grâce auquel j'ai beaucoup appris en statistiques et dont les conseils d'écriture ont été très utiles.

J'adresse mes sincères remerciements à Laurent Denant-Boèmont et à André de Palma qui ont accepté d'être les rapporteurs de cette thèse. Ma reconnaissance va également à Michel Bierlaire, Jacques-François Thisse, Véronique Simonnet et Catherine Viot pour leur participation au jury.

J'associe également à mes remerciements la Région Rhône-Alpes, pour le financement de la thèse et de l'enquête.

J'aimerais aussi remercier les membres de mes deux laboratoires d'accueil, le LAET et le GAEL. Naviguant entre Lyon et Grenoble, j'ai toujours été bien accueillie. Merci au personnel administratif et informatique qui a su gérer la double localisation et ses contraintes (Marie, Marianne, Gwenaëlle, Janine, Brigitte et Valérie). La thèse a également été l'occasion de nouvelles rencontres et d'amitiés qui ont rendu ce parcours plus sympathique. J'ai une pensée particulière pour Béatrice, Adélaïde, Modou et Rania qui ont été les réceptacles de mes interrogations métaphysiques et qui ont toujours été de bons conseils. Merci également aux relecteurs des différents chapitres de la thèse qui ont fortement contribué à les améliorer, Paolo, Sabrina, Béatrice, Adélaïde, Carole, Anne, Tamsin et mon père.

Cette thèse n'aurait pas pu être finalisée sans toutes les personnes, famille et amies, qui ont donné quelques heures, jours ou semaines de leur temps pour garder mes enfants. Je pense en particulier à ma mère, mes beaux-parents, Muriel, Raphaelle, Stéphanie, Anne, Anaïs et Célanie.

Je terminerai en remerciant mes enfants, Nathanaël et Gabriel, et mon mari, Olivier, dont la gentillesse et la patience, quel que soit mon état de stress ou de fatigue, ont été indispensables à la finalisation de la thèse.

Table of contents

Introduction générale 3

I Enquête de préférences déclarées - revue de littérature, mise en œuvre et analyse 27

Introduction 29

1 Enquêtes de préférences déclarées - état de l'art 35

1.1 Introduction . . . 35

1.2 Protocole d'enquête . . . 38

1.3 Elaboration du questionnaire . . . 43

1.4 Design d'une enquête de préférences déclarées . . . 53

1.5 Définition des attributs . . . 62

2 Mise en œuvre de l'enquête de préférences déclarées en Région Rhône-Alpes 71

2.1 Analyse qualitative . . . 72

2.2 Questionnaire de l'enquête . . . 73

2.3 Constitution et personnalisation des questions de choix . . . 80

2.4 Recrutement des usagers . . . 82

3 Première analyse des résultats 85

3.1 Bases de données brutes . . . 85

3.2 Statistiques descriptives . . . 86

3.3 Modèles de choix discret . . . 93

II Modeling mode choice (habits) with latent variables - contributions to an economic and behavioral perspective 105

Introduction 107

4 Integrated choice and latent variable models: a literature review on mode choice 121

4.1 Introduction . . . 122

4.2 Methodology . . . 123
4.3 Survey dimension . . . 124
4.4 Econometric dimension . . . 127
4.5 Psychological dimension . . . 133
4.6 Economic dimension . . . 143
4.7 Conclusion . . . 146

5 Comfort and heterogeneity in values of travel time savings in the Rhône-Alpes Region 147
5.1 Introduction . . . 147
5.2 Data . . . 151
5.3 Integrated choice and latent variable modeling . . . 154
5.4 Results . . . 160
5.5 Conclusion . . . 173

6 How does environmental concern influence mode choice habits? A mediation analysis 175
6.1 Introduction . . . 176
6.2 Literature review . . . 177
6.3 Hypotheses and theoretical model . . . 182
6.4 Data and method . . . 184
6.5 Results . . . 191
6.6 Conclusion . . . 201

III Rank-dependent utility theory and reference models - contributions to the modeling of mode choice 203

Introduction 205

7 A rank-dependent utility approach to model intra- and inter-individual heterogeneity in risky choices 213
7.1 Introduction . . . 213
7.2 Theories of choice under conditions of risk . . . 217
7.3 Model framework . . . 218
7.4 Data . . . 222
7.5 Results . . . 224

7.6 Conclusion 232

8 Reference models: an economic and behavioral perspective 233

8.1 An introduction to reference models 233

8.2 A new family of regression models for categorical data: an application to travel mode choice 238

8.3 Work in progress and perspectives 272

Conclusion Générale 281

Appendices 293

A Théories de psychologie environnementale (Introduction) 295

B Analyse qualitative : mise en œuvre, supports et résultats (Ch.2) 298

C Déroulement et questionnaires de l'enquête de préférences déclarées (Ch.2) 307

D Literature review on ICLV models: identification, estimation and structural model (Ch.4) 347

E Determinants of value of time (Ch.5) 352

F Goodness-of-fit indicators for the ICLV models (Ch.5) 354

G Estimation results and goodness-of-fit indicators for the M3 models (Ch.6) 356

H Les modèles de choix discret : un cas particulier de modèles linéaires généralisés (Ch.8) 359

I Value of time, time equivalent and marginal effects in reference models (Ch.8) 368

J Synthèse des indicateurs économiques (Conclusion) 372

Bibliography 377

List of Figures 397

List of Tables 399

Avant-propos

La thèse est rédigée en français (introduction, partie I et conclusion) et en anglais pour les parties qui contiennent des articles soumis à des revues ou conférences internationales (parties II et III). L'orthographe américaine est adoptée sauf quand la revue demande explicitement d'utiliser l'orthographe britannique (chapitre 4).

Chaque chapitre des parties II et III est présenté tel qu'il a été soumis, avec des objectifs qui lui sont propres. Les notations varient donc de chapitre en chapitre et la section sur les données se répète puisqu'elle reprend des informations de la partie I qui porte sur la réalisation d'une enquête de préférences déclarées.

Lorsque les sources des figures et tableaux ne sont pas précisées, cela signifie que je les ai réalisés dans le cadre de la thèse. Sauf mention contraire, les tableaux et figures s'appuient sur les données de l'enquête de préférences déclarées conduite pendant la thèse.

Introduction générale

En France, 360 milliards d'euros sont dépensés chaque année pour le transport, soit 16,7% du produit intérieur brut (SOeS, 2017). Les ménages y contribuent pour moitié, le transport représentant 13,2 % de leur dépenses de consommation. Cette consommation n'est pas sans conséquence puisque 79,5% des kilomètres voyageurs parcourus annuellement sont réalisés en véhicule particulier (voiture ou deux-roues motorisés). L'usage massif de la voiture a l'originalité d'engendrer à la fois des externalités [1] et des internalités [2]. Des externalités d'abord puisque la voiture émet des gaz à effet de serre [3], génère de la pollution atmosphérique [4], consomme de l'espace [5], génère du bruit [6] et cause des accidents [7]. L'usage de la voiture génère également des internalités. C'est en effet un mode sédentaire, alors que le vélo, la marche à pied ou même les transports en commun favorisent la pratique d'une activité physique qui permet de réduire l'obésité et les maladies cardio-vasculaires (Sallis, 2016 ; Rojas-Rueda et al., 2016 ; Sallis et al., 2004).

Comprendre pourquoi les usagers choisissent un mode de transport plutôt qu'un autre représente donc à la fois un enjeu de santé publique et un enjeu environnemental [8], en France, mais également dans l'ensemble du monde, même si les enjeux se posent différemment dans les pays en développement. Être capable de comprendre et simuler ce choix est la condition *sine qua non* au développement d'outils de politique publique socialement acceptables et efficaces, c'est-à-dire atteignant, à moindre coût, les objectifs recherchés : la diminution de l'usage de la voiture particulière et le développement des modes alternatifs (transports en commun, modes actifs, co-voiturage et auto-partage).

Cette thèse apporte une pierre à l'édifice de l'ensemble des contributions académiques et opérationnelles cherchant à expliquer le choix modal, c'est-à-dire le choix entre plusieurs alternatives

1. Voir, par exemple, de Palma et Zaouali (2007) pour une revue de littérature sur les externalités liées au transport, y compris la congestion.
2. Ce terme a été introduit pour la première fois par Herrnstein et al. (1993) et peut se définir comme le coût qu'une personne s'impose lorsqu'elle sous-estime ou ignore les conséquences que son comportement peut avoir sur elle-même. Le concept d'internalité a notamment été utilisé pour analyser la consommation de cigarettes (Gruber et Kőszegi, 2004), la taxation de produits non sains (tels que les produits gras) (O'Donoghue et Rabin, 2006) ou encore la consommation d'énergie (Allcott et al., 2014).
3. Depuis 1990, les émissions de gaz à effet de serre ont diminué dans toutes les activités économiques sauf dans les transports, responsables de 29 % des émissions.
4. La pollution atmosphérique a un effet sanitaire majeur sur la morbidité et la mortalité. 48 000 décès chaque année seraient attribuables à la pollution par les particules fines à l'échelle nationale (Pascal et al., 2016). Le coût économique résultant de ces effets sanitaires est colossal et a été estimé à environ 100 milliards d'euros par an pour la France en 2015(Aïchi, 2015)
5. Un déplacement domicile-travail réalisé en voiture consommerait 90 fois plus d'espace-temps (en $m^2 \times$heure) que le même déplacement réalisé en bus ou en tram (Mezghani, 2001).
6. En utilisant la méthode des prix hédoniques, différentes études montrent l'impact du bruit généré par le trafic routier sur le prix des biens immobiliers (Szczepańska et al., 2015 ; Łowicki et Piotrowska, 2015, par exemple,). En particulier, Andersson et al. (2010) trouve que le bruit généré par la route a un effet négatif beaucoup plus important que celui généré par une voie de chemin de fer.
7. 3 461 personnes sont mortes dans un accident de la route en France en 2015 (SOeS, 2017).
8. Mais également un enjeu économique et d'aménagement du territoire.

de transport. L'originalité de la présente contribution tient à la **volonté de mobiliser des éléments de psychologie et d'économie comportementale pour expliquer le choix modal, notamment dans des modèles de choix discret**. En effet, les modèles économiques faisant l'hypothèse de préférences rationnelles, invariantes au contexte et ne dépendant que de caractéristiques observables ne sont pas suffisants pour expliquer pourquoi un mode de transport est choisi. L'intégration de variables latentes, inobservables, dans les modèles de choix (Ben-Akiva et al., 2002b), le développement de nouvelles structures de modélisation (voir, par exemple, Chorus et al., 2008), la prise en compte de biais comportementaux (voir, par exemple, Rasouli et Timmermans, 2015) sont autant de pistes qui permettent de mieux expliquer le choix modal. Le champ de recherche qui s'ouvre est donc large et dépasse largement le cadre de la thèse. La thèse se concentre donc plus précisément sur certains aspects comportementaux du **choix modal** lorsque celui-ci est réalisé à l'**échelle régionale**. L'un des objectifs de la thèse était en effet de réaliser une enquête de préférences déclarées sur l'ensemble de la Région Rhône-Alpes, avec des distances allant de quelques kilomètres à quelques centaines de kilomètres, afin d'affiner la modélisation du choix modal, et plus précisément du choix du train, dans un modèle de prévision de la demande de transport[9] permettant de penser et d'adapter les futures politiques régionales de transport.

Hormis l'arbitrage entre temps et coût de déplacement, commun à toutes les échelles d'analyse, la question du choix modal se pose en des termes différents selon l'échelle géographique à laquelle il est étudié. L'échelle régionale a été relativement peu abordée dans la littérature (Bhat, 1995 ; FitzRoy et Smith, 1995 ; Ben-Akiva et Morikawa, 2002 ; Daziano et Rizzi, 2015) et comporte, même si elles ne sont pas tout à fait exclusives de cette échelle, un certain nombre d'originalités en raison des modes de transport impliqués : le train, le car et la voiture. La thèse se concentre plus particulièrement sur trois d'entre elles : le confort, le retard et la conscience environnementale.

Tout d'abord, se pose la question du **confort**. En raison de temps de déplacement plus longs qu'à l'échelle d'une agglomération et de la possibilité de pratiquer des activités annexes comme travailler, lire ou se reposer, le confort dans les trains et les cars est un enjeu important (Richter et Keuchel, 2012 ; Wardman, 2014), d'autant plus qu'il y a encore encore relativement peu de contributions dans la littérature. En particulier, la question de la perception du confort a été abordée à l'aide de méthodes sociologiques (par exemple, Kaufmann, 2002), mais son influence sur la valeur du temps n'a pas encore été approchée avec des méthodes quantitatives, pourtant nécessaires à l'analyse économique.

9. Ces modèles cherchent à estimer le nombre de véhicules ou de personnes qui vont utiliser une infrastructure dans le futur. Ils se divisent essentiellement en trois catégories : les modèles à quatre étapes, les modèles basés sur les activités et les modèles LUTI (*Land Use and Transport Integrated*) qui combinent modèle d'occupation du sol et modèle de transport.

Comprendre l'hétérogénéité de la valeur du temps dans les transports en commun peut aider à adapter les infrastructures et à identifier des segments de la population à cibler en priorité.

Ensuite, se pose la question des **retards**, souvent évoqués pour justifier d'utiliser la voiture plutôt que le train. La fiabilité des temps de déplacement et leur valorisation a fait l'objet de nombreuses publications (Li et al., 2010). Cependant, ce n'est que depuis récemment que des modèles issus de l'économie comportementale (*prospect theory* et modèle d'utilité dépendante au rang) sont utilisés en transport. Les applications portent principalement sur le choix d'itinéraire (Li et Hensher, 2011) et non sur le choix modal. Par ailleurs, la question de l'hétérogénéité des attitudes vis-à-vis du risque sur le temps de déplacement a été peu étudiée. Et, à ma connaissance, il n'existe pas de travaux qui analysent dans quelle mesure la perception du retard varie entre les modes de transport.

Enfin, se pose la question de l'influence de la **conscience environnementale** sur les habitudes de choix modal, les résultats de la littérature étant relativement contradictoires (voir, par exemple, Garvill, 1999 ; De Groot et Steg, 2007). La conscience environnementale peut agir directement sur les habitudes de choix modal, en favorisant le choix du car ou du train, mais peut également exercer une influence indirecte au travers des attitudes vis-à-vis des différents modes de transport. Intuitivement, une personne ayant une conscience environnementale élevée aura vraisemblablement une meilleure perception des transports en commun et utilisera la voiture plus pour des raisons instrumentales que pour des raisons affectives ou symboliques. De telles relations n'ont pas encore été analysées. Pourtant, les comprendre est utile pour aider à développer des politiques publiques environnementales qui influencent le processus décisionnel de choix modal.

A ces trois questions thématiques s'ajoutent deux questions méthodologiques. Premièrement, depuis le développement des **modèles de choix hybrides** qui combinent modèle de choix discret et modèle d'équations structurelles (Ben-Akiva et al., 2002b), l'incorporation de variables latentes dans les modèles de choix appliqués au transport a connu un essor important. En identifiant des déterminants du choix modal qui n'avaient jusqu'à présent pas été pris en compte, du moins pas dans une approche économique, les variables latentes permettent de mieux rendre compte de l'hétérogénéité des comportements. Face à la diversité des applications, une revue de littérature est nécessaire afin de structurer les apports des modèles de choix hybrides à la modélisation du choix modal.

Deuxièmement, une nouvelle famille de choix discret, **la famille des modèles de référence**, a été introduite par Peyhardi et al. (2015). Leur formulation originale offre *a priori* une alternative intéressante aux modèles de choix discret utilisés jusqu'à présent, tant d'un point de vue économique que d'un

point de vue comportemental. Cette famille de modèles, qui inclut le modèle logit multinomial, permet d'enrichir la modélisation des choix discret par deux aspects, à étudier plus en détails. D'une part, les probabilités de choix calculées avec les modèles de référence sont dépendantes de l'option choisie comme référence, ce qui leur donne une assise comportementale prometteuse. En effet, des données expérimentales montrent que les choix sont dépendants de la référence prise en compte par l'individu (Tversky et Kahneman, 1991). D'autre part, chacun des modèles de cette famille propose une écriture différente de la fonction de probabilité, ce qui étoffe la capacité de modélisation des comportements et peut permettre une meilleure adéquation aux données.

A la lumière de ces différents besoins de recherche, la problématique de la thèse peut se formuler sous la forme de la question suivante :

> *Comment l'intégration de théories de psychologie et d'économie comportementale dans les modèles de choix discret permet-elle de mieux appréhender l'aspect comportemental du choix modal, et plus particulièrement l'hétérogénéité des comportements ?*

Des éléments de réponse à cette question sont apportés en s'intéressant plus particulièrement à certains déterminants du choix modal (confort, retards, conscience environnementale) et en s'appuyant sur les modèles d'équations structurelles, les modèles de choix hybrides et les modèles de référence.

Après ce préambule, l'introduction générale est structurée en trois parties. La première partie approfondit la question de l'intégration entre théorie économique et comportementale pour expliquer le choix modal et montre en quoi cette intégration est nécessaire pour comprendre les déterminants du choix modal et donc l'hétérogénéité des comportements. La deuxième partie pose les fondements théoriques des modèles de choix discret et souligne les avantages théoriques des modèles qui sont utilisés dans la suite de la thèse. La troisième partie passe en revue les différents chapitres de la thèse et les principales conclusions.

Une thématique : les déterminants du choix modal

> *"[...] economic and behavioral science provides insights and models that are useful at the many points of human interaction in transportation systems. The behavioral science useful for transportation research depends on the grain of transportation policies studied, and needs to be updated regularly with the progress of economic and behavioral science. Consumers are likely to resist market solutions for transportation problems, and careful education and framing may be needed to overcome agoraphobia. Finally, transportation research can benefit from continuing close*

connections to economic and behavioral science, and active pursuit of new ideas and findings from the behavioral sciences that can improve your ability to understand and predict the effects of human behavior in transportation systems." McFadden (2007, p.273)

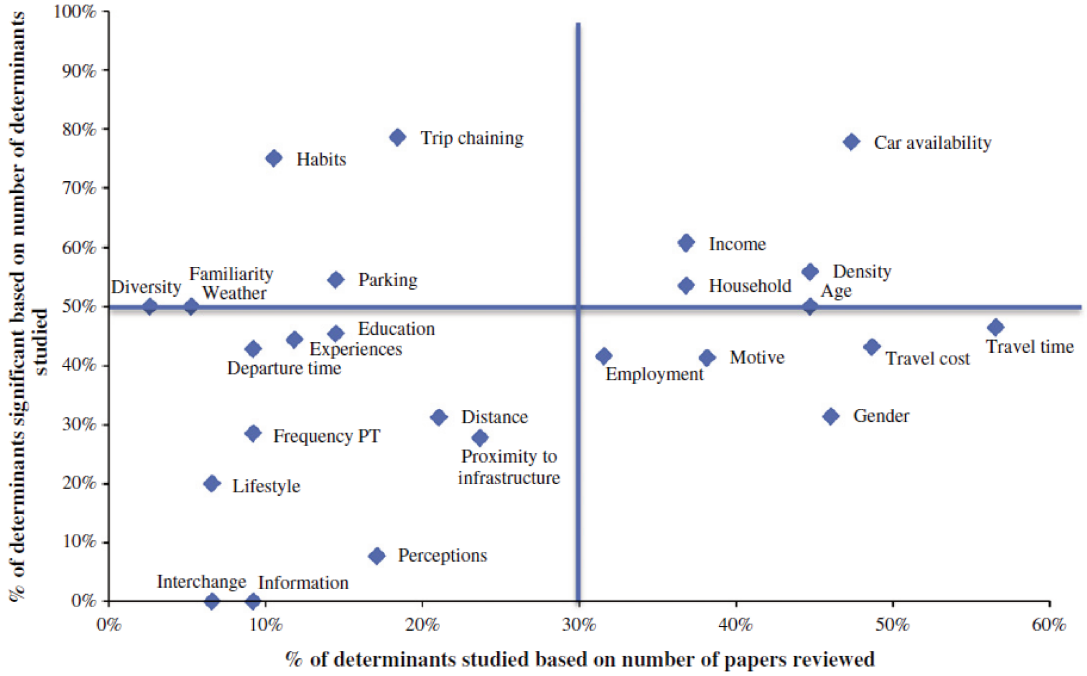
Cette citation de McFadden (2007) souligne la nécessité de s'appuyer à la fois sur des outils économiques et sur des outils issus des théories comportementales pour analyser les décisions de transport, et plus particulièrement le choix modal. Pour comprendre pourquoi cette intégration est utile, il est nécessaire de dresser un panorama des déterminants du choix modal. Dans un premier temps, une entrée thématique est adoptée pour montrer l'importance des déterminants psychologiques. Dans un deuxième temps, une entrée offre / demande est proposée afin de mettre en valeur la diversité de ces déterminants et comment ils peuvent être traités.

Le choix modal peut se définir comme « le processus décisionnel qui conduit à choisir entre plusieurs alternatives de transport et qui est déterminé par une combinaison de facteurs individuels, socio-démographiques et spatiaux, et influencé par des facteurs socio-psychologiques » (De Witte et al., 2013, p. 339). Cette définition montre l'importance de s'inspirer d'une littérature inter-disciplinaire (économie, géographie, sociologie et psychologie) pour comprendre comment se forme le choix modal. Une typologie des déterminants du choix modal est proposée par De Witte et al. (2013). Ils distinguent

1. les caractéristiques socio-économiques : âge, genre, éducation, occupation, revenu, composition du ménage, motorisation ;
2. les caractéristiques spatiales : densité, diversité, proximité aux infrastructures et services, fréquence de passage des transports en commun, parking ;
3. les caractéristiques du déplacement : motif, distance, temps, coût, heure de départ, l'inscription du déplacement dans une boucle, les conditions climatiques, l'information disponible, les changements entre modes ;
4. les caractéristiques psychologiques : l'expérience, la familiarité avec le système de transport, le style de vie, les habitudes, la perception des différents modes.

La figure 1 met en regard ces différents déterminants du choix modal en calculant le pourcentage d'études qui montrent qu'un déterminant est significatif en fonction du pourcentage d'études qui considèrent ce déterminant. Même si ce constat est moins vrai aujourd'hui, cette figure montre que les caractéristiques psychologiques restent relativement peu étudiées comparativement aux caractéristiques socio-économiques et aux caractéristiques des alternatives alors même que certaines

d'entre elles (par exemple, les habitudes) expliquent significativement le choix modal lorsqu'elles sont introduites.



Source : De Witte et al. (2013)

FIGURE 1 – Classification des déterminants du choix modal basée sur la revue de De Witte et al. (2013).
Notes : La figure se lit de la façon suivante. Par exemple, pour les habitudes (*Habits*), 10% des études analysées les prennent en compte comme déterminant du choix modal et 80% de celles-ci les trouvent significatives.

Pour mieux caractériser la diversité des déterminants psychologiques et comprendre pourquoi il est nécessaire de mobiliser des théories comportementales, les déterminants du choix modal, et en particulier les déterminants psychologiques, peuvent également être divisés entre « offre » et « demande ». Premièrement, le choix modal dépend des caractéristiques des alternatives de transport, c'est l'offre déterminée par les infrastructures et les décisions publiques. Dans cette catégorie entrent les caractéristiques du déplacement et les caractéristiques spatiales décrites ci-dessus : le temps de déplacement, le coût de déplacement ou encore le niveau de service (fréquence de passage des transports en commun, parking, qualité des prestations...). Cependant, cette offre « objective » peut être déformée par le prisme des perceptions individuelles (Bhattacharya et al., 2014). Au moment de faire le choix d'un mode de transport plutôt qu'un autre, l'individu prend en compte l'offre telle qu'il la perçoit. Cette perception de l'offre est une première catégorie de variables qui nécessite de mobiliser des théories de psychologie et d'économie comportementale pour pouvoir les intégrer dans les modèles. Elle peut prendre deux formes : des variables latentes qui traduisent la perception

individuelle de l'offre (par exemple, l'évaluation du confort dans les transports en commun) ou une transformation de l'offre « objective » au travers d'hypothèses structurelles permettant de mieux représenter la façon dont l'offre est considérée par les individus (par exemple, la transformation des probabilités dans les modèles de choix en situation de risque).

Deuxièmement, le choix modal dépend des caractéristiques individuelles. Dans ces caractéristiques individuelles entrent en compte les caractéristiques socio-démographiques, des caractéristiques spatiales (par exemple, le lieu de résidence), des caractéristiques du déplacement (par exemple, le motif) et des caractéristiques psychologiques (valeurs, normes, conscience environnementale...). Cette deuxième catégorie de variables psychologiques fait référence à l'individu en tant que tel, au même titre que les variables socio-démographiques. Pour être complet, soulignons que parmi les caractéristiques propres à l'individu, certaines relèvent de son intégration dans un groupe social (ménage, groupe d'appartenance... ; voir Hayakawa, 2000 ; Pollak, 1976). Cette intégration soumet les choix individuels à l'influence sociale du groupe sur l'individu (normes, culture...). Elle génère également une interdépendance des choix entre membres du groupe et donc un processus de négociation, par exemple au sein du couple (de Palma et al., 2011 ; Picard et al., 2013). La décision peut également ne plus s'effectuer au niveau individuel mais s'effectuer au niveau collectif, par exemple à l'échelle d'un ménage (de Palma et al., 2014).

La figure 2 illustre comment ces différents éléments s'articulent. L'offre de transport détermine l'utilité soit directement, soit indirectement si les perceptions individuelles modifient l'offre prise en compte dans l'utilité. La demande, c'est-à-dire l'individu et les contraintes auxquelles il fait face, expliquent à la fois la perception de l'offre et l'utilité. Parmi les variables individuelles, certaines sont observables et d'autres sont latentes, ce sont les variables psychologiques.

La perception de l'offre et les variables latentes individuelles sont inobservables par nature. Pour les inclure dans les modèles, il va donc falloir formuler un certain nombre d'hypothèses en s'appuyant sur les enseignements de la psychologie et de l'économie comportementale. Dans un premier temps, je m'intéresse aux théories de psychologie environnementale, encore peu utilisées en économie comportementale (Croson et Treich, 2014). Ces théories permettent d'identifier les variables latentes qui entrent dans le processus de choix modal et comment ces variables sont liées entre elles. Dans un deuxième temps, je m'intéresse aux enseignements de l'économie comportementale mobilisés dans la thèse, notamment pour expliquer le choix modal face à un temps de déplacement incertain et pour mettre en regard la famille statistique des modèles de référence avec les modèles économiques de dépendance à la référence.

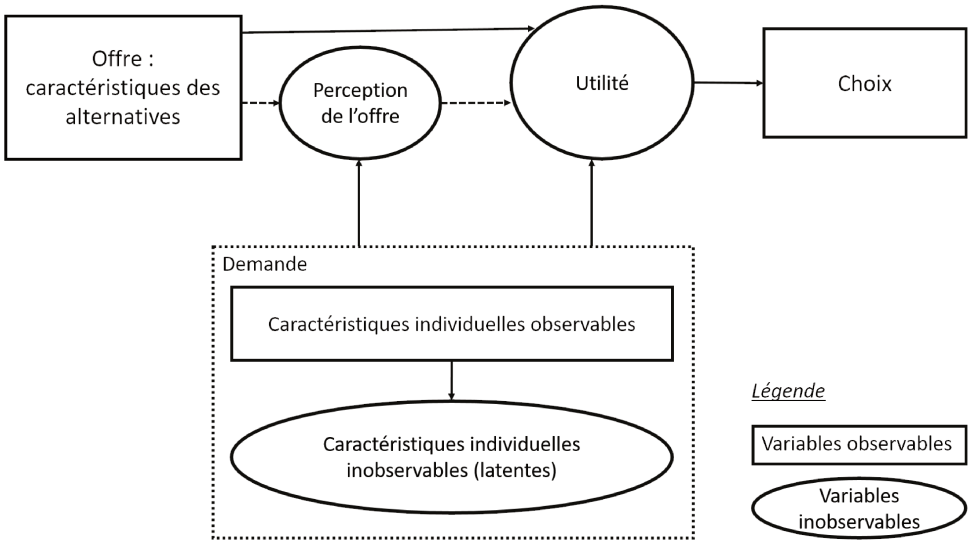


FIGURE 2 – Processus de choix

Apports de la psychologie environnementale

La motivation d'intégrer des variables psychologiques pour expliquer le choix modal s'explique notamment par trois composantes comportementales qui font l'originalité du choix modal.

Comme cela a déjà été évoqué, **le choix modal a des impacts environnementaux**, à la fois au niveau local (pollution atmosphérique, bruit...) et au niveau mondial (changement climatique dû aux gaz à effet de serre). La psychologie environnementale s'intéresse aux relations entre l'Homme et son environnement physique (naturel ou construit) et social (Gifford et al., 2011). L'un de ses domaines d'application est l'analyse des comportements pro-environnementaux et des leviers permettant de les encourager. Les modèles comportementaux développés à cet effet mettent en lien des concepts psychologiques (normes, attitudes, habitudes...) dans un modèle structurel afin de rendre compte du processus de formation des préférences conduisant à un comportement donné. Les théories de psychologie environnementale, décrites en annexe A, sont donc pertinentes pour expliquer le choix du mode de transport (Bamberg et al., 2003a ; Bamberg et Schmidt, 2003 ; De Groot et Steg, 2007 ; Donald et al., 2014). Elles permettent d'identifier les variables psychologiques qui entrent en compte dans le choix modal, et de comprendre comment elles sont reliées entre elles afin d'améliorer le réalisme de la formalisation de l'utilité. En particulier, Hoffmann et al. (2017) montrent à l'aide d'une méta-analyse que les quatre variables psychologiques les plus liées à l'usage de la voiture sont les intentions, la perception de contrôle, les attitudes et les habitudes. Les trois premiers éléments se retrouvent notamment dans la théorie du comportement planifié (*Theory of planned behavior* ; Ajzen, 1991) qui

explique le comportement à partir de l'intention de réaliser ce comportement, elle-même fonction de la perception de contrôle, des attitudes et des normes subjectives de l'individu. L'une des autres principales théories de psychologie environnementale utilisée pour expliquer le choix d'un mode de transport est la théorie *Value-Belief-Norm*, spécifiquement développée pour expliquer des comportements pro-environnementaux (Stern et al., 1999) en s'appuyant notamment sur les normes personnelles, les croyances environnementales, la prise de conscience de l'existence d'un problème environnemental et le sentiment de responsabilité vis-à-vis de ce problème.

Cependant, les facteurs mentaux internes décrits dans la théorie du comportement planifié ou la théorie *Value-Belief-Norm* ne sont pas suffisants pour expliquer le choix modal. Des facteurs externes, observables, traduisant les contraintes auxquelles l'individu fait face doivent également être prises en compte. Facteurs internes et externes dessinent un espace à deux dimensions au sein duquel le choix modal peut être expliqué. Les modèles de choix hybrides qui sont en capacité de prendre en compte ces deux dimensions sont donc prometteurs. Cependant la prise en compte des théories de psychologie environnementale dans les modèles de choix hybrides en est encore à ses débuts. Ce point fait notamment l'objet du chapitre 4. Par ailleurs, la notion de perception de contrôle fait partie des variables latentes mesurant la perception de confort et d'utilité des transports en commun dans les chapitres 5 et 6. Le chapitre 6 mobilise également les éléments de la théorie *Value-Belief-Norm* mesurant la conscience environnementale pour expliquer non pas un choix modal ponctuel mais les habitudes de choix modal.

En effet, le choix modal **se distingue par son caractère répétitif, les habitudes jouent donc un rôle important**. Un même déplacement, souvent le déplacement domicile-travail, est effectué, jour après jour, dans les mêmes conditions. L'influence des habitudes, analysée notamment par Verplanken et al. (1994), Bamberg et al. (2003b) ou Eriksson et al. (2008), se vérifie à la fois sur les déplacements répétés mais également sur le comportement de mobilité en général. Plus une personne utilise un mode de transport, plus elle tendra à utiliser ce mode, quel que soit le type de déplacement considéré. Le changement d'habitudes est un processus complexe, qui se joue sur le long terme et est favorisé par l'intervention d'un élément déclencheur, par exemple financier (Charness et Gneezy, 2009) ou informatif (Allcott et Rogers, 2014) [10]. La prise de conscience environnementale en fait partie comme le

10. Ces deux exemples, issus de l'économie comportementale, portent respectivement sur la pratique d'une activité sportive et sur la consommation d'énergie.

montre Vincent (2008) [11].

Enfin, le choix modal revêt un **caractère de consommation qui ne remplit pas seulement des fonctionnalités instrumentales mais aussi des fonctionnalités affectives et symboliques**, en particulier en ce qui concerne la voiture (Dittmar, 1992 ; Steg, 2005). Steg et al. (2001) montrent que les modèles de comportement qui se concentrent sur les motivations instrumentales ne suffisent pas à expliquer le choix prédominant de la voiture. La prise en compte des motivations affectives et symboliques permet d'approfondir la compréhension du comportement de mobilité. Les motivations affectives font référence au plaisir procuré par la voiture, non pas au travers de sa finalité (arriver à destination) mais au travers de l'objet (aimer avoir une belle voiture par exemple) et de son utilisation (plaisir de la vitesse, de la conduite...). Les motivations symboliques portent sur la représentation de soi. Le comportement de mobilité dépend en effet de la pression sociale, du regard des autres et de ce qu'on en perçoit (normes subjectives). Ainsi, la voiture peut être envisagée comme un symbole qui montre qui on est (Schlenker, 1980), qui permet de se comparer aux autres (Festinger, 1954) et d'être dans les normes (Cialdini et Reno, 1991). Le chapitre 6 analyse ces motivations affectives et symboliques et les mets en regard de la sensibilité environnementales et des habitudes de choix modal.

La mesure des variables latentes et le test des relations structurelles qui les unissent, s'appuient sur des modèles d'équations structurelles (Hoyle, 2012). Ces modèles font notamment l'hypothèse que les variables latentes peuvent être mesurées à l'aide d'indicateurs observables évalués sur des échelles de Likert [12].

Apports de l'économie comportementale

Alors que l'économie des transports s'intéresse fortement aux composantes comportementales des choix et pourrait donc bénéficier des apports de l'économie comportementale qui permet de traduire les phénomènes psychologiques en termes économiques, les transferts entre les deux disciplines sont encore limités (Gaker et al., 2010) - vraisemblablement du fait d'un relatif cloisonnement académique - mais tendent toutefois à augmenter [13].

Selon Rabin (2002), trois catégories de phénomènes psychologiques sont prises en compte par

11. Selon Vincent (2008), le changement d'habitudes de choix modal résulte de la conjonction de trois facteurs : un contexte biographique favorable au changement, une disposition à adopter un mode alternatif et enfin, un élément déclencheur tel qu'un engagement (en faveur d'actions écologiques par exemple), une contrainte (voiture en panne par exemple) ou une opportunité (prise de conscience de l'existence d'une alternative au déplacement automobile, par exemple dans le cadre d'un plan de déplacements entreprise).

12. Une échelle de Likert est une échelle psychométrique sur laquelle le répondant exprime son degré d'accord ou de désaccord vis-à-vis d'une affirmation.

13. Pour plus de détails, voir les travaux de différents auteurs (Rasouli et Timmermans, 2015 ; Chorus, 2012a ; Metcalfe et Dolan, 2012 ; Avineri, 2012 ; Hensher et al., 2013 ; Cherchi, 2009) ayant identifié les (potentielles) applications de l'économie comportementale au domaine des transports.

l'économie comportementale et constituent son apport par rapport à l'économie "classique" pour reprendre ses propres termes. Ces phénomènes permettent de mettre en évidence :

- "De nouvelles hypothèses sur les préférences - à quoi ressemble la fonction d'utilité ?
- Des heuristiques et biais dans les jugements - comment est-ce que les individus forment leurs croyances sur les probabilités d'occurrence des états de la nature ?
- Un manque de stabilité dans la maximisation de l'utilité - est-ce que les individus maximisent réellement leur utilité ?" Rabin (2002, p.7)

Chacun de ces trois apports couvre de nombreuses dimensions dont seulement une petite partie est abordée dans la thèse pour mieux comprendre comment le choix modal est réalisé. Premièrement, une **formalisation plus réaliste de l'utilité est permise par l'intégration de variables latentes**, déjà abordée dans la section précédente.

Deuxièmement, **l'hypothèse que les individus maximisent leur utilité espérée a largement été remise en cause dans la littérature** (pour une discussion, voir Denant-Boèmont et L'Haridon, 2013). Par exemple, Allais (1953) montre par l'expérience que l'axiome d'indépendance, axiome clé de la théorie de l'utilité espérée, impliquant notamment la linéarité en probabilités, n'est pas respectée. Au-delà des observations empiriques, la théorie de l'utilité espérée a également l'inconvénient de modéliser deux phénomènes différents au travers de la fonction d'utilité : l'attitude vis-à-vis du risque et l'attitude par rapport à la richesse (ou par rapport au temps de déplacement dans le cas du choix modal). Avec la théorie de l'utilité aléatoire, un individu ayant une utilité marginale décroissante (utilité concave) est nécessairement averse au risque (pour une discussion, voir Cohen, 2015). Ces observations empiriques et théoriques ont justifié le développement de modèles alternatifs à la théorie de l'utilité espérée, et notamment la **théorie de l'utilité dépendante au rang** (Quiggin, 1982 ; Yaari, 1987) et la *(cumulative) prospect theory* (Kahneman et Tversky, 1979 ; Tversky et Kahneman, 1992). La théorie de l'utilité dépendante au rang introduit l'idée selon laquelle le poids dans la décision associé à chaque évènement dépend à la fois de la probabilité objective qui y est associé et du rang de la conséquence, c'est-à-dire de la probabilité d'obtenir une conséquence strictement meilleure. Les probabilités d'occurrence des évènements sont donc remplacées par une formulation non linéaire des poids de décision. Les théories de choix en situation de risque ont été adaptées à la problématique des retards, les probabilités de gains monétaires étant remplacés par des probabilités de retard et les montants par des temps de déplacement. Li et Hensher (2011) réalisent une revue de littérature de l'utilisation de la *(cumulative) prospect theory* en transport. Ils soulignent notamment que son utilisation a permis de mieux comprendre les conséquences de gains ou de pertes de temps et de mieux expliquer

les choix relativement à la théorie de l'utilité espérée. Par ailleurs, des modèles de dépendance au rang ont été utilisés pour analyser le rôle des probabilités de retard dans le choix d'itinéraire (Michea et Polak, 2006 ; Hensher et Li, 2012) et ponctuellement, le choix modal (de Lapparent et Ben-Akiva, 2014). Ces applications restent tout de même rares et ont peu abordé la question de l'hétérogénéité des comportements face au risque sur le temps de déplacement. Cette question est abordée dans le chapitre 7 qui intègre un modèle d'utilité dépendante au rang dans un modèle de choix discret. C'est un exemple de comment l'offre perçue (perception du retard) peut être intégrée pour expliquer le choix modal en s'appuyant sur les enseignements de l'économie comportementale.

Troisièmement, la formalisation de l'utilité peut être améliorée en prenant en compte le fait que les choix sont dépendants du contexte dans lequel ils sont effectués (Tversky et Simonson, 1993 ; Ben-Akiva et al., 2012). La notion de contexte est large et la thèse s'intéresse à deux situations de dépendance contextuelle en particulier. La première repose sur des observations qualitatives qui me permettent de formuler l'hypothèse que la perception des probabilités n'est pas la même selon le mode de transport emprunté. Il existerait donc une hétérogénéité intra-individuelle. Pour en rendre compte, le modèle d'utilité dépendante au rang du chapitre 7 intègre des coefficients différenciés selon le mode considéré (train ou voiture). La dépendance des choix à une situation de référence est un deuxième exemple dans lequel les choix sont soumis à un effet de contexte. La théorie économique "classique" fait l'hypothèse que les choix sont indépendants de la situation de référence considérée par l'individu. Or, les travaux de psychologie et d'économie comportementale montrent que les préférences sont dépendantes des références prises en considération. D. Kahneman explique intuitivement la dépendance des préférences à une référence en ces termes :

> « La théorie dominante considère que celui qui possède 3 millions d'euros comme fortune aujourd'hui a la même espérance d'utilité, qu'il ait eu hier 5 millions d'euros ou 2 millions. Or, c'est absurde : celui qui avait 5 millions hier est malheureux, ce qui n'est pas le cas de l'autre, au contraire. Pour définir l'espérance d'utilité, il faut donc ajouter une valeur de référence. » D. Kahneman dans une interview au magazine La Recherche en 2003.

La dépendance des préférences à une référence a notamment été analysée dans des situations avec risque (Kahneman et Tversky, 1979 ; Tversky et Kahneman, 1992) et des situations sans risque (Tversky et Kahneman, 1991). L'une des caractéristiques les plus importantes de ce cadre théorique est l'aversion au perte : la perception des pertes, relativement à la référence, apparaît plus importante que la perception des gains. Il en résulte que les consentements à payer sont généralement plus faibles que les consentements à recevoir. Une autre incidence de l'aversion aux pertes est le biais de statu quo

(*statu quo bias*) qui traduit le fait que, face à un choix, les individus ont tendance à rester ancrés à leur référence (Samuelson et Zeckhauser, 1988) [14]. Afin de rendre compte de la dépendance à la référence, Tversky et Kahneman (1991) proposent de définir une relation de préférence et donc une fonction d'utilité qui soit dépendante de la référence prise en compte. D'autres auteurs utilisent également la notion d'utilité dépendante à la référence (Köszegi et Rabin, 2006) ou d'utilité relative (Zhang et al., 2004 ; Pan et al., 2016). La traduction économique des modèles de référence, dans le chapitre 8, peut être réalisée en définissant également une utilité dépendante à la référence. Cet appui sur la notion de dépendance à la référence permet de justifier économiquement et comportementalement l'utilisation des modèles de référence.

Un outil : les modèles de choix discret

Le terme « modèle de choix discret » désigne les modèles qui permettent d'expliquer et de prévoir les choix réalisés entre des alternatives exclusives non ordonnées, telles que les alternatives de transport [15]. Ces modèles peuvent s'appliquer à de nombreuses autres situations telles que le choix d'un produit alimentaire ou le choix d'un contrat énergétique. Les outils et enseignements de la thèse pourraient donc s'étendre à d'autres thématiques que celle des transports. Les modèles de choix discret sont notamment utiles en économie industrielle pour analyser la différenciation de produits (Anderson et al., 1992 ; Thisse et Billot, 1995).

La plupart des modèles de choix discret s'appuient sur la théorie de l'utilité aléatoire (*Random Utility Theory*). Deux exceptions sont les modèles probabilistes basés sur le modèle « Elimination par attributs » de Tversky (1972a,b) et analysés par Laurent (2007), ainsi que le modèle de minimisation du regret aléatoire développé par Chorus et al. (2008) et Chorus (2010). Le développement de ces deux types de modèle a notamment été motivé par des considérations comportementales mais ils ne sont pas abordés dans la thèse. Une troisième voie sont les modèles de référence présentés dans le chapitre 8.

Les modèles basés sur la théorie de l'utilité aléatoire étant encore aujourd'hui les plus fréquents et les mieux fondés économiquement puisque s'appuyant sur la notion d'utilité [16], ce sont ces modèles qui sont utilisés dans les autres chapitres de la thèse.

14. Notons cependant que l'aversion aux pertes n'est qu'une des causes possibles du biais de de *statu quo*.
15. L'introduction du chapitre 8 brosse un tableau plus général des modèles expliquant les variables qualitatives.
16. D'autres processus décisionnels ont été décrits et analysés dans la littérature, voir par exemple Hensher (2010) ou Chorus (2014)

Fondements micro-économiques

La théorie microéconomique fait en général l'hypothèse de biens continus indivisibles. Le comportement du consommateur y est décrit comme un problème d'optimisation dans lequel le consommateur sélectionne le panier de biens de façon à maximiser son utilité sous contrainte budgétaire. Dans la théorie du consommateur classique, c'est la quantité consommée pour chaque bien qui fait l'objet du choix. Les biens sont généralement considérés homogènes : une voiture est une voiture, quelle que soit sa couleur ou sa marque. Pourtant, de nombreux problèmes de consommation font intervenir des biens discrets, le choix modal en est un exemple, ce qui nécessite d'élargir le cadre théorique. En présence de biens discrets, une hypothèse complémentaire doit être ajoutée en mobilisant la théorie de la consommation de Lancaster qui propose de considérer les attributs des biens comme déterminants de l'utilité :

> *"1. The good, per se, does not give utility to the consumer; it possesses characteristics, and these characteristics give rise to utility. 2. In general, a good will possess more than one characteristic, and many characteristics will be shared by more than one good." (Lancaster, 1966, p.134)*

Le raisonnement suivi par la suite est celui décrit par McFadden (1981) et permet de montrer d'où viennent les fonctions d'utilité mobilisées dans la théorie de l'utilité aléatoire et les modèles de choix discret. Soit un consommateur n ($n = 1, \cdots, N$), qui doit à la fois choisir les quantités qu'il consomme sur L biens continus $\mathbf{z^T} = (z_1, \ldots, z_L)$ et sélectionner un bien discret j ($j = 1, \cdots, J$), parmi l'ensemble des alternatives mutuellement exclusives disponibles $\mathcal{C}_n$. $\mathbf{z}$ représente l'ensemble des biens (continus) consommés par l'individu et le bien discret représente une combinaison de l'ensemble des choix discret effectués par le consommateur. Chaque alternative j est décrite par ses K attributs inclus dans le vecteur $\tilde{\mathbf{y}}_\mathbf{j}$. La matrice $\tilde{\mathbf{Y}} = (\tilde{\mathbf{y}}_\mathbf{1}^T, \cdots, \tilde{\mathbf{y}}_\mathbf{J}^T)$, de dimension $(J \times K)$, contient l'ensemble des attributs. Le choix de l'individu est décrit par le vecteur $\mathbf{c} = (c_1, \cdots, c_J)^T$ dans lequel la variable c_j vaut 1 si l'alternative j est choisie et 0 sinon. Par ailleurs, l'individu dispose d'un revenu R et fait face à une contrainte budgétaire. $\mathbf{p^T} = (p_1, \ldots, p_L)$ est le vecteur des prix des biens continus et $\mathbf{s^T} = (s_1, \ldots, s_J)$ est le vecteur du coût des alternatives discrètes. En notant $\boldsymbol{\theta}$ l'ensemble des paramètres à estimer, le programme du consommateur s'écrit

$$\begin{aligned} &\max_{\mathbf{z},\mathbf{c}} \tilde{U}(\mathbf{z}, \tilde{\mathbf{Y}}^T\mathbf{c}; \boldsymbol{\theta}), \\ &\text{sous contraintes } \mathbf{p^T z} + \mathbf{s^T c} \leq R \\ &\text{et } \sum_{j=1}^{J} c_j = 1, c_j \in \{0, 1\}, \forall j \in \mathcal{C}_n. \end{aligned} \tag{1}$$

Pour résoudre ce problème d'optimisation, McFadden (1981) propose d'utiliser une procédure en deux

étapes. La première étape est de déterminer la quantité optimale de biens continus consommée conditionnellement à la sélection d'une alternative j

$$z_j^\star = z(\mathbf{p}, R - s_j, \tilde{\mathbf{y}}_\mathbf{j}; \boldsymbol{\theta}), \forall j \in \mathcal{C}_n. \tag{2}$$

Cette quantité dépend du vecteur de prix $\mathbf{p}$ mais également du choix sur le bien discret au travers du revenu restant[17] $R - s_j$ et de ses caractéristiques $\tilde{\mathbf{y}}_\mathbf{j}$. En remplaçant la fonction de demande conditionnelle dans la fonction d'utilité permet de calculer l'utilité indirecte

$$\tilde{U}_j(z_j^\star, \tilde{\mathbf{y}}_\mathbf{j}; \boldsymbol{\theta}) = V_j(\mathbf{p}, R - s_j, \tilde{\mathbf{y}}_\mathbf{j}; \boldsymbol{\theta}), \tag{3}$$

elle-même conditionnelle au choix sur le bien discret. Elle peut donc être évaluée pour chaque alternative j. La comparaison des utilités indirectes entre elles permet de déterminer l'alternative $j^\star$ qui maximise l'utilité et qui est donc choisie

$$\max_{j \in \mathcal{C}_n} V_j(\mathbf{p}, R - s_j, \tilde{\mathbf{y}}_\mathbf{j}; \boldsymbol{\theta}) \equiv j^\star \left\{ V_1(\mathbf{p}, R - s_j, \tilde{\mathbf{y}}_\mathbf{j}; \theta), \cdots, V_J(\mathbf{p}, R - s_j, \tilde{\mathbf{y}}_\mathbf{j}; \boldsymbol{\theta}) \right\}. \tag{4}$$

En faisant l'hypothèse que le vecteur de prix $\mathbf{p}$ et le revenu R entrent linéairement dans l'utilité indirecte, alors ils peuvent être omis puisqu'ils sont communs à l'ensemble des alternatives. Le choix de l'alternative ne dépend alors que des caractéristiques et du coût de l'alternative j. Par la suite, l'utilité indirecte tronquée $V_j(s_j, \tilde{\mathbf{y}}_\mathbf{j})$ est nommée « utilité » associée à l'alternative j. Pour introduire de l'hétérogénéité entre individus, il est également d'usage de prendre en compte le vecteur des caractéristiques individuelles $\mathbf{x_n}$ de l'individu n ($n = 1, \cdots, N$). Par ailleurs, le coût de l'alternative j est généralement considéré comme faisant partie de ses attributs et donc compris dans $\mathbf{y_{n,j}} = (\tilde{\mathbf{y}}_\mathbf{n,j}, s_j)^T$, l'indice n étant introduit pour refléter le fait que tous les individus ne font pas face aux mêmes caractéristiques des alternatives. L'utilité retirée par l'individu n face à l'alternative j s'écrit alors $V_{n,j}(\mathbf{x_n}, \mathbf{y_{n,j}}; \boldsymbol{\theta})$.

La formalisation de l'utilité peut encore être affinée selon les éléments comportementaux pris en compte et présentés précédemment. Premièrement, les modèles de choix hybrides (chapitres 4 et 5) nécessitent l'adjonction d'un vecteur de variables latentes $\boldsymbol{\xi_n}$. L'utilité s'écrit alors $V_{n,j}(\mathbf{x_n}, \mathbf{y_{n,j}}, \boldsymbol{\xi_n}; \boldsymbol{\theta}), \ \forall j \in \mathcal{C}_n$. Les variables latentes sont elles-mêmes déterminées par un modèle d'équations structurelles décrit dans les chapitres correspondants (chapitres 4, 5 et 6).

17. Si le coût de l'alternative j dépasse le revenu, alors cette alternative n'est pas disponible.

Deuxièmement, le chapitre 7 prend en compte le risque dans un modèle d'utilité dépendante au rang. Chaque état de la nature, ou évènement, $E \in \mathcal{E}$, est associé à une conséquence (un temps de déplacement) et une probabilité, p^E, qui est connue (situation de risque). En situation d'incertitude, les probabilités sont inférées par l'individu, selon une procédure bayésienne, pour représenter ses croyances. Le poids associé à chaque probabilité est une transformation non-linéaire de cette probabilités $W(p^E)$. Avec ces hypothèses, l'utilité individuelle s'écrit $\sum_{E \in \mathfrak{E}} W(p^E) V_{n,j}(\mathbf{x_n}, \mathbf{y_{n,j}}|E; \boldsymbol{\theta}),\ \forall j \in \mathcal{C}_n$.

Troisièmement, les modèles de référence du chapitre 8 nécessitent de définir l'utilité comme dépendante des caractéristiques de l'alternative de référence J. Par ailleurs, l'utilité n'est pas une fonction linéaire des variables explicatives et peut s'écrire $\tilde{V}_{n,j}(\mathbf{x_n}, \mathbf{y_{n,j}} - \mathbf{y_{n,J}}; \boldsymbol{\theta}),\ \forall j \neq J, j \in \mathcal{C}_n$.

Théorie de l'utilité aléatoire

Une hypothèse supplémentaire est nécessaire pour pouvoir modéliser les choix discrets. Alors que la théorie du consommateur suppose des choix déterministes, les modèles de choix discret reposent sur la théorie de l'utilité aléatoire qui envisage les choix comme étant probabilistes. Dans la théorie de l'utilité aléatoire, l'utilité $U_{n,j}$ est modélisée comme une variable aléatoire qui est la somme d'une composante déterministe $V_{n,j}$ et d'une composante aléatoire $\varepsilon_{n,j}$:

$$U_{n,j} = V_{n,j} + \varepsilon_{n,j}. \tag{5}$$

La théorie de l'utilité aléatoire trouve ses origines dans les travaux de psychologues (Thurstone, 1927 ; Luce, 1959) puis d'économistes (Marschak et al., 1960 ; McFadden, 1973) ; voir l'historique retracé par McFadden et al. (2000) dans le discours de remise de son Prix de la banque de Suède.

L'adjonction d'un terme aléatoire peut s'expliquer par différentes raisons (Block et Marschak, 1960 ; de Palma et Thisse, 1989). La raison la plus souvent retenue par les économètres ne relâche pas l'hypothèse de rationalité. Le processus décisionnel reste déterministe, mais la modélisation de l'utilité est stochastique [18]. En effet, un grand nombre de variables interviennent dans la prise de décision et, du fait d'un manque d'informations, le modélisateur ne parvient à en observer qu'un sous-ensemble, les autres variables étant prises en compte dans le terme aléatoire qui traduit alors l'hétérogénéité inobservée. Manski (1977) identifie quatre sources d'incertitude : les caractéristiques des alternatives

18. Une hypothèse alternative est que l'utilité est déterministe mais la règle de décision aléatoire . L'erreur reflète l'« état d'esprit », le contexte du choix ou une erreur de jugement de l'individu en raison par exemple de capacités cognitives restreintes. Les différentes interprétations du terme d'erreur ne sont pas nécessairement antagonistes mais relèvent de raisonnements différents.

non observées, les caractéristiques individuelles (ou les variations) non observées, les erreurs de mesure et la nécessité d'utiliser des variables instrumentales pour estimer l'utilité, qui n'est pas observable avec certitude.

Modèles de choix discret

Selon les hypothèses formulées sur le terme aléatoire $\varepsilon_{n,j}$, différents modèles de choix discret peuvent être envisagés. Ils reposent sur l'hypothèse que, pour chaque individu, la probabilité de choisir l'option j parmi l'ensemble des options possibles est la probabilité $\pi_{n,j}$ que cette option maximise son utilité. Parmi les modèles de choix discret, le modèle logit multinomial, initialement développé par McFadden (1973), a l'avantage d'exprimer les probabilités de choix de façon simple :

$$\pi_{n,j} = \frac{\exp(V_{n,j})}{\sum\limits_{i \in \mathcal{C}_n} \exp(V_{n,i})}, \forall j \in \mathcal{C}_n. \tag{6}$$

Cette simplicité se fait au prix d'un certain nombre d'hypothèses, notamment sur les résidus (chaque $\varepsilon_{n,j}$ est distribué indépendamment et identiquement selon la loi des valeurs extrêmes de type I), qui limitent sa flexibilité et la plausibilité des comportements simulés. Par exemple, le modèle logit multinomial possède la propriété d'indépendance par rapport aux alternatives non pertinentes (*Independance from Irrelevant Alternatives* ; IIA). Dans les faits, les alternatives peuvent varier dans leur degré de similarité du fait de caractéristiques partagées par certaines d'entre elles. De nombreux livres (par exemple, Train, 2009) et articles (par exemple, Baltas et Doyle, 2001) traitent des modèles de choix discret. Seuls les modèles utilisés dans la thèse sont donc décrits ci-dessous avant d'être abordés en détail dans chacune des parties correspondantes.

Pour répondre aux objectifs de la thèse, quatre extensions du modèle logit multinomial sont utilisées. Un premier objectif est d'introduire des variables psychologiques dans les modèles de choix discret par des **modèles de choix hybrides**. Une telle intégration a été réalisée dès les années 70 (Koppelman et Hauser, 1978 ; Prashker, 1979 ; Koppelman et Pas, 1980). Mc Fadden souligne en 1986 la nécessité de prendre en compte des facteurs psychologiques, sociologiques et biologiques dans la modélisation des choix. Cependant, de telles variables sont généralement inobservables (latentes) rendant leur intégration difficile. C'est seulement depuis les années 2000 que, grâce au développement d'un cadre conceptuel flexible combinant modèle de choix discret et modèle d'équations structurelles (Ben-Akiva et al., 2002a,b), le nombre d'applications intégrant des variables psychologiques dans des modèles de choix modal s'est multiplié. Les modèles de choix hybrides sont composés de deux

sous-modèles. Premièrement, un modèle d'équations structurelles traite les variables latentes. Ce modèle est lui-même décomposé en un modèle de mesure qui mesure les variables latentes à l'aide d'indicateurs observés et un modèle structurel qui relie les variables latentes entre elles ou avec d'autres variables observées telles que des variables socio-économiques. Les variables latentes sont ensuite intégrées dans un modèle de choix discret, qui peut prendre la forme d'un logit multinomial, ou tout autre modèle (nested logit, mixed logit...). Malgré la disponibilité des modèles de choix hybrides depuis une quinzaine d'années, la pratique de modéliser les variables latentes au travers d'un modèles d'équations structurelles n'est toujours pas systématique parmi les économistes, ce qui conduit à des erreurs de mesure et des estimateurs non convergents. Les modèles de choix hybrides sont utilisés dans les chapitres 4 et 5. Un modèle d'équations structurelles seul est estimé dans le chapitre 6.

Un deuxième objectif est d'analyser l'hétérogénéité individuelle dans l'attitude face au risque. Cette hétérogénéité peut être modélisée au travers d'un modèle à coefficients aléatoires (*mixed logit* ; McFadden et al., 2000). C'est par exemple l'option retenue par Conte et al. (2011). Une autre option, qui est celle retenue dans la thèse, est d'utiliser un **modèle à classes latentes** (Greene et Hensher, 2003). Ce dernier a l'avantage d'éviter d'avoir à faire des hypothèses sur la distribution des coefficients et est plus intuitif à interpréter. En effet, différentes classes sont définies puis chaque personne est assignée de façon probabiliste à chacune des classes en fonction de ses caractéristiques. Dans le chapitre 7, deux classes sont définies : une classe de personnes neutres au risque et une classe de personnes répondant au modèle d'utilité dépendante au rang. Ce chapitre estime également des modèles **nested logit** qui permettent de prendre en compte une substitution plus forte entre les alternatives train qu'entre le train et la voiture et s'affranchissent donc de la propriété IIA du modèle logit multinomial.

Un troisième objectif est d'analyser les potentialités des **modèles de référence**. Initialement développés en génomique des plantes, ce sont des modèles linéaires généralisés qui peuvent s'adapter en économie par leur capacité à expliquer des variables qualitatives non ordonnées. Ils généralisent le modèle logit multinomial en modifiant la fonction de lien, c'est-à-dire en remplaçant la fonction de répartition de la loi logistique par une autre fonction de répartition (normal, Laplace, Student...). Du fait de leur dépendance à l'alternative de référence, l'une des spécificités de ces modèles est de partiellement s'affranchir de la propriété IIA. Ces modèles font l'objet du chapitre 8.

Plan et synthèse de la thèse

La thèse est structurée en trois parties, chacune composée de deux ou trois chapitres. La partie I porte sur la construction de l'enquête de préférences déclarées et une première analyse des résultats. La partie II porte sur l'intégration de variables latentes pour expliquer le(s habitudes de) choix modal à l'aide de modèles de choix hybrides et de modèles d'équations structurelles. La partie III s'intéresse au choix modal lorsque les temps de déplacement ne sont pas connus, ainsi qu'aux potentialités économiques et comportementales des modèles de référence.

Partie I

L'objet général de la thèse est d'étudier le choix modal en s'appuyant sur la théorie économique mais aussi sur des théories comportementales. Afin de répondre à cet objectif, tout en disposant d'une base de données adaptée et permettant des analyses économétriques robustes, l'ensemble des analyses économétriques repose sur une enquête de préférences déclarées à l'échelle de la Région Rhône-Alpes, entièrement réalisée dans le cadre de la thèse, ce qui a permis d'adapter les choix méthodologiques aux besoins de la thèse et de maîtriser du mieux possible la qualité des données. A cette échelle d'analyse, le choix du mode de déplacement se fait principalement entre le train, le car et la voiture [19]. La finalité de l'enquête est notamment d'analyser les déterminants de l'usage du train.

Outre le temps et le coût de trajet, Ben-Akiva et Morikawa (2002) identifient les variables susceptibles d'expliquer un choix entre rail (train et métro) et bus (local et express) : fiabilité, information disponible, confort, sécurité par rapport aux accidents, sécurité par rapport à des agressions et accessibilité. Ils montrent qu'une fois que le temps et le coût de déplacement sont contrôlés, les préférences pour le rail s'expliquent par une meilleure qualité de service. Ils concluent que les variables inobservables, subjectives telles que le confort, doivent être analysées avec des méthodes adaptées, les modèles de choix hybrides.

Par ailleurs, pour comprendre les déterminants de l'usage du train, il semble important d'analyser les comportements des usagers du train mais aussi celui des usagers des modes alternatifs. Pourtant, les modèles de choix modal incluant le train ont été analysés soit sur un corridor ferroviaire spécifique (Kottenhoff et Lindh, 1995 ; Ben-Akiva et Morikawa, 2002 ; Börjesson et Eliasson, 2011), soit en interrogeant uniquement des usagers du train (Wardman et al., 2004 ; Richter et Keuchel, 2012). Une exception est le travail réalisé par Arentze et Molin (2013) qui ont mené quatre enquêtes de préférences

19. La Région Rhône-Alpes couvre 44 000 km^2 et les distances peuvent dépasser les 300 km.

déclarées aux Pays-Bas en interrogeant tous types d'usage. Cependant, seule l'enquête qui porte sur des distances intermédiaires (20 km) inclue une alternative voiture et celle-ci n'est pas personnalisée.

Au final, la réalisation d'une enquête à l'échelle régionale qui couvre à la fois des variables observables et des variables latentes, qui interrogent des usagers du train, du car et de la voiture à une échelle régionale tout en personnalisant les questions de choix afin d'éviter les biais induits par des situations hypothétiques, constitue une question méthodologique en soi. Afin d'y parvenir, le **chapitre 1** propose une revue de littérature sur les enquêtes de préférences déclarées, et plus particulièrement la méthode de *choice experiment* qui est utilisée. Cette revue de littérature met en particulier l'accent sur le design de l'enquête et examine la façon dont les précédentes enquêtes de préférences déclarées ont présenté et défini les attributs présentés dans les questions de choix.

Le **chapitre 2** décrit la mise en œuvre opérationnelle de l'enquête. Les attributs choisis pour décrire le choix modal sont le temps de déplacement, le coût de déplacement, les probabilités et temps de retard, la fréquence de passage des trains et des cars, le cadencement des horaires de train et de car et le confort dans les trains (place assise garantie ou non). Afin d'éviter une surcharge cognitive, ces attributs sont divisés dans trois exercices de choix séparés. L'échantillon de l'enquête provient de deux sources, certaines personnes sont interrogées par internet et d'autres en face-à-face dans les trains de la Région Rhône-Alpes. La personnalisation des questions de choix à l'échelle de l'ensemble de la région a été permise par la constitution d'une base de données comprenant les temps et coûts de déplacement de l'ensemble des couples origine-destination à moins de 10 km d'une gare de la région.

Le **chapitre 3** propose une première analyse des résultats. Les estimations, les consentements à payer (y compris la valeur du temps), les équivalents temps et les élasticités sont dans l'ensemble cohérents avec les attentes et la littérature. Par exemple, les valeurs du temps sont comprises entre 10 et 14 €/heure, ce qui est dans la fourchette de valeurs du rapport Quinet et al. (2014) qui constitue la référence officielle française.

Partie II

La partie II est composée de trois chapitres qui s'intéressent à l'utilisation de variables latentes pour expliquer le choix modal. Le chapitre 4 fait une revue de littérature sur les modèles de choix hybrides [20], le chapitre 5 analyse l'hétérogénéité de la valeur du temps dans les transports en commun en fonction de différentes mesures du confort et le chapitre 6 analyse la relation de médiation entre conscience environnementale, perceptions de l'utilité des transports en commun, motivations

20. Les modèles de choix hybrides sont nommés "*Integrated Choice Latent Variable models*" dans les chapitres en anglais.

affectives et symboliques d'utilisation de la voiture et habitudes de choix modal.

Le **chapitre 4** réalise une revue de littérature sur l'utilisation des modèles de choix hybrides sur la base de 35 applications au choix modal. Celles-ci sont très riches et montrent la diversité des variables latentes susceptibles d'expliquer le choix modal. Pourtant, depuis l'introduction des modèles de choix hybrides, il n'existe pas, à ma connaissance, de revue de littérature sur les quatre dimensions qui interviennent dans ces modèles : recueil des données, économétrie, psychologie et économie. Un positionnement clair sur chacune de ces dimensions est pourtant nécessaire pour estimer un modèle de choix hybrides. L'objectif de ce chapitre est donc de dresser un bilan des expériences passées afin d'orienter les recherches futures. La première dimension est liée aux données et à leur recueil. En particulier, la mesure des variables psychologiques à l'aide d'indicateurs est réalisée à l'aide de méthodes développées en sciences de gestion et en psychologie. La deuxième dimension concerne la méthode économétrique proprement dite. L'examen des modèles d'équations structurelles compris dans les modèles de choix hybrides révèle que les variables latentes sont souvent intégrées telles qu'elles dans l'utilité sans que ces variables ne soient reliées entre elles dans le modèle structurel. La richesse des modèles d'équations structurelles, qui permettent de prendre en compte des relations de médiation ou de modération, n'est donc souvent pas prise en compte. Par conséquent, les théories de psychologie environnementale sont encore rarement intégrées, ou de façon incomplète, dans les modèles de choix hybrides. Malgré tout, les variables latentes introduites dans les modèles de choix hybrides couvrent une grande variété de thématiques permettant ainsi de mieux comprendre les déterminants du choix modal et de mettre à jour des mécanismes comportementaux qui peuvent être utilisés pour adapter les politiques publiques. En tant que modèles de choix discret basés sur la théorie de l'utilité aléatoire, les modèles de choix hybrides ont une assise économique qui permet de les utiliser pour calculer des outputs économiques (consentements à payer, élasticités). C'est l'une de leur plus-value comparativement aux modèles d'équations structurelles. Et, comparativement aux autres modèles de choix discret, leur valeur ajoutée repose sur la possibilité d'analyser l'hétérogénéité de ces outputs économiques en fonction de composantes comportementales. C'est ce qui est réalisé dans le chapitre suivant.

A l'aide d'un modèle de choix hybride, le **chapitre 5** analyse l'hétérogénéité de la valeur du temps, c'est-à-dire le consentement à payer pour gagner une heure de déplacement, selon des mesures objectives et latentes du confort dans les transports en commun. Le confort est un déterminant important du choix modal, notamment lorsque celui-ci est réalisé à l'échelle régionale et donc sur des distances importantes. Depuis les travaux de Mokhtarian et Salomon (2001), le temps de déplacement

n'est plus forcément perçu comme une perte de temps mais comme un temps que le voyageur peut s'approprier. Ce serait alors davantage la qualité comparée du temps de déplacement que sa quantité comparée qui détermine le choix modal Kaufmann (2002). La qualité du temps de déplacement dépend du sentiment de confort des usagers, qui est mesuré à l'aide de trois variables latentes : les sentiments, positifs ou négatifs, éprouvés durant les déplacements en train ou en car, la perception que le temps de déplacement est un temps (in)utile, et la perception de contrôle que l'individu a sur l'usage des transports en commun. La qualité du temps de déplacement dépend également de facteurs objectifs qui sont approchés par la garantie ou non d'avoir une place assise dans le train. Ce travail est l'un des premiers à expliquer l'hétérogénéité de la valeur du temps en fonction de composantes comportementales. L'estimation du modèle montre que la garantie d'une place assise diminue la valeur du temps, de même que le fait d'éprouver des sentiments positifs dans les transports en commun et de percevoir le temps de déplacement en train ou en car comme un temps utile.

Le **chapitre 6** estime un modèle d'équations structurelles afin d'expliquer les habitudes de choix modal. Ce modèle se base sur l'intuition que les personnes ayant une forte conscience environnementale ont une meilleure perception de l'utilité des transports en commun et donc une propension plus importante à les utiliser. A l'inverse, les personnes ayant une faible conscience environnementale utiliseraient davantage la voiture pour des motivations affectives et symboliques et auraient donc une propension plus élevée à utiliser ce mode de transport. L'utilité indirecte du déplacement, mesurée par les motivations affectives et symboliques pour la voiture et par la perception des transports en commun et notamment la perception du temps de déplacement, servirait donc de médiateur entre la conscience environnementale et les habitudes de choix modal. Dans ce modèle, la conscience environnementale est mesurée par trois dimensions qui sont issues de la théorie *Value-Belief-Norm* : une dimension cognitive qui mesure les croyances environnementale au travers de l'échelle *New Environmental Paradigm*, une dimension affective qui mesure le sentiment de responsabilité vis-à-vis du problème environnemental et une dimension conative qui mesure la prise de conscience du problème environnemental. Différents modèles sont estimés afin de mesurer si la relation de médiation est complète, partielle ou inexistante. Les résultats montrent qu'une perception favorable des transports en commun joue un rôle de médiation partielle entre la conscience environnementale et les habitudes d'usage des transports en commun. Les motivations affectives et symboliques d'utilisation de la voiture sont bien expliquées par la conscience environnementale. En revanche la relation entre motivations affectives et symboliques et habitudes d'usage de la voiture n'est pas robuste.

Partie III

La partie III répond à deux objectifs distincts : analyser l'hétérogénéité intra- et inter-individuelle face à un temps de déplacement incertain (chapitre 7) et analyser les potentialités économiques et comportementales des modèles de référence (chapitre 8).

Les théories alternatives à la théorie de l'utilité espérée, en particulier, l'utilité dépendante au rang, émergent peu à peu en transport pour expliquer les choix lorsque le temps de déplacement est incertain. Cependant, ces théories sont principalement utilisées pour expliquer des choix d'itinéraires. Peu d'applications s'intéressent au choix modal, c'est donc l'objet du **chapitre 7**. La contribution de ce chapitre est d'analyser différentes perceptions des probabilités de retard selon le mode de transport considéré et différentes attitudes par rapport au risque. Dans la vie réelle, les temps de déplacement sont incertains, les individus ne connaissent pas précisément leur distribution et formulent des croyances sur les probabilités de retard. Dans l'enquête de préférences déclarées en revanche, les temps de déplacement et les probabilités de retard sont connus (situation de risque). L'hypothèse est faite que les perceptions des probabilités (sur- ou sous-évaluation) dans la vie réelle se transfèrent dans l'enquête de préférences déclarées. L'estimation de modèles intégrant une utilité dépendante au rang montre que l'évaluation des probabilités de retard en train sont sur-évaluées et l'évaluation des probabilités de retard en voiture sont sous-évaluées, que le modèle prenne en compte une transformation puissance des probabilités ou une transformation à la Tversky et Kahneman (1992). L'estimation du modèle à classes latentes montre que les usagers du train ont une probabilité plus élevée d'appartenir à la classe décrivant un comportement neutre vis-à-vis du risque, alors que les usagers de la voiture ont une probabilité plus élevée d'appartenir à la classe dans laquelle le comportement vis-à-vis du risque est décrite par le modèle d'utilité dépendante au rang.

Le **chapitre 8** porte sur les modèles de référence. Cette nouvelle famille de modèles repose sur une écriture des modèles de choix discret comme des modèles linéaires généralisés. Au lieu d'utiliser la fonction de lien canonique comme dans le modèle logit multinomial, d'autres fonctions de lien sont envisagées. Peyhardi et al. (2015) a montré que la fonction de lien des modèles de choix dédiés aux variables discrètes non-ordonnées peut s'écrire comme la composée d'un ratio et d'une fonction de distribution. Lorsque la distribution est la distribution logistique, alors le modèle est le logit multinomial. Le principe des modèles de référence, qui peuvent être estimés aussi facilement que le modèle logit multinomial, est d'utiliser d'autres distributions que la distribution logistique. Ils n'ont encore jamais été utilisés en économétrie, le travail présenté dans ce chapitre présente donc une avancée méthodologique, réalisée en trois temps. Premièrement, les formules théoriques des outputs

économiques (effets marginaux, élasticités, consentements à payer) qui généralisent ceux des modèles logit multinomiaux sont présentées. Deuxièmement, les modèles de référence sont appliqués aux données de l'enquête de préférences déclarées. Pour cela, six distributions sont testées : logistic, normal, Laplace, Cauchy, Gumbel et Student, avec trois jeux de données variant par l'hétérogénéité inobservée qu'ils contiennent. Les résultats des estimations montrent qu'il existe toujours un modèle de référence (avec nos données, ce sont ceux générés avec les distributions de Laplace, Student et Gumbel) qui surpasse le modèle logit multinomial en termes d'adéquation aux données, ce qui a notamment des incidences en termes de sélection des variables. Par ailleurs, la distribution de Student est la seule qui génère des outputs économiques cohérents, en particulier quand l'hétérogénéité inobservée est élevée. Ce résultat peut s'expliquer par le fait que, pour la distribution de Student, le nombre de degrés de liberté peut être adapté, ce qui permet d'ajuster l'épaisseur des queues de distribution et donc l'adéquation aux données. Avec des queues de distribution épaisses, les probabilités sont relativement insensibles au changement dans le prédicteur, excepté autour d'un seuil autour duquel la pente de la fonction de distribution est élevée. Troisièmement, afin de ne pas réduire les modèles de référence au remplacement d'une hypothèse paramétrique par d'autres hypothèses paramétriques, une interprétation économique et comportementale est proposée. Les probabilités estimées à l'aide des modèles de référence présentent l'originalité d'être dépendant de l'alternative choisie comme référence. Cette dépendance est justifiée d'un point de vue comportemental puisqu'elle a été montrée par l'expérience et a donné lieu à la théorie de dépendance à la référence de Tversky et Kahneman (1991). La traduction économique de cette dépendance passe par la définition d'une utilité qui dépend des caractéristiques de l'alternative envisagée mais également des caractéristiques de l'alternative de référence. L'utilité fait alors intervenir une différence entre attributs qui peut être interprétée comme un gains ou une perte par rapport à l'alternative de référence. Lorsque la distribution choisie est celle de la loi de Laplace ou Student, alors la forme de l'« utilité de référence » est concave dans les gains et convexe dans les pertes et traduit donc une utilité marginale décroissante, cohérente avec la théorie économique. Les distributions qui offrent les meilleurs résultats avec nos données semblent donc aussi être celles qui ont les meilleurs fondements économiques et comportementaux.

Partie I

Enquête de préférences déclarées - revue de littérature, mise en œuvre et analyse

Introduction

Objectifs

L'objectif de cette partie est de présenter l'Enquête de Préférences Déclarées (EPD) réalisée au début de la thèse et sur laquelle l'ensemble des analyses de choix modal de la thèse sont réalisées. D'un point de vue opérationnel, l'EPD complète une enquête de préférences révélées, l'Enquête Déplacements Régionale (EDR). La combinaison des deux enquêtes permet à la Région Auvergne-Rhône-Alpes[21] d'avoir un état des lieux complet des pratiques de déplacement des Rhônealpins sur la base duquel elle construit notamment un modèle multi-modal[22] à l'échelle régionale, afin de prévoir la demande de transport selon différents scénarios et d'adapter les politiques publiques en conséquence. Concernant spécifiquement l'EPD, les objectifs auxquels cette partie répond sont les suivants :

- réaliser une revue de littérature afin de construire une enquête à jour des enseignements des précédentes expériences ;
- présenter la mise en œuvre de l'EPD et les réponses apportées aux différents enjeux posés par la réalisation d'une enquête à l'échelle régionale ;
- présenter l'analyse des données de l'enquête à l'aide d'un modèle de choix discret simple.

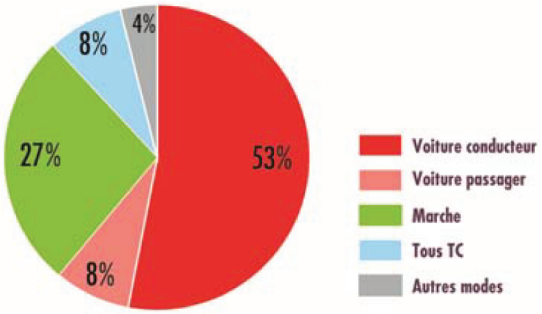
Motivations

La Région Rhône-Alpes a mis en place un dispositif d'observation et de prospective de la mobilité à l'échelle régionale, notamment pour asseoir ses choix de politique de transport régional de voyageurs (train et car), les adapter sur le territoire et favoriser les modes d'écomobilité. L'un des objectifs de ce dispositif est la création d'un modèle multi-modal à l'échelle régionale, modèle qui nécessite la modélisation du choix modal. La question du choix modal se pose en des termes différents selon l'échelle d'analyse. Si intégrer le train dans la réflexion n'est pas pertinent à l'échelle d'agglomérations (par exemple, Grenoble ou Lyon) où l'usage du train est marginal, la modélisation explicite du train devient indispensable à l'échelle régionale. Certaines bases de données existantes, telles que les

21. Au moment du lancement des deux enquêtes, la fusion entre les régions Auvergne et Rhône-Alpes n'avait pas encore eu lieu. Elles ont donc été effectuées à l'échelle de la région Rhône-Alpes.
22. Le modèle multi-modal régional est un modèle de type « quatre étapes » dont l'objectif est de simuler la demande pour chacun des modes de transport selon différents scénarios. Les quatre étapes sont génération de la demande de déplacements, distribution des déplacements, répartition modale et affectation (Bonnel, 2002).

enquêtes photos[23], les données de comptage de la Région ou les bases de données de la SNCF, permettent de décrire la demande mais pas de positionner le train relativement aux modes concurrents à l'échelle régionale : les cars (régionaux ou départementaux) et la voiture.

Un dispositif d'enquête, l'EDR, visant à connaître les pratiques des déplacements des Rhônalpins a été mis en place par la Région Rhône-Alpes. Cette enquête permet d'approfondir la connaissance de la mobilité des rhône-alpins âgés de 11 ans et plus. Une première analyse de l'EDR montre la prédominance de la voiture relativement aux autres modes de transport. Elle représente 53 % des déplacements (voir figure 3), alors que les transports en commun ne représentent que 8% de la part modale, et, en particulier, les trains et cars régionaux représentent 1% de la part modale. Par ailleurs, 2,7% des déplacements sont des déplacements intermodaux, 80% d'entre eux incluant les transports en commun urbains, trains ou cars régionaux.



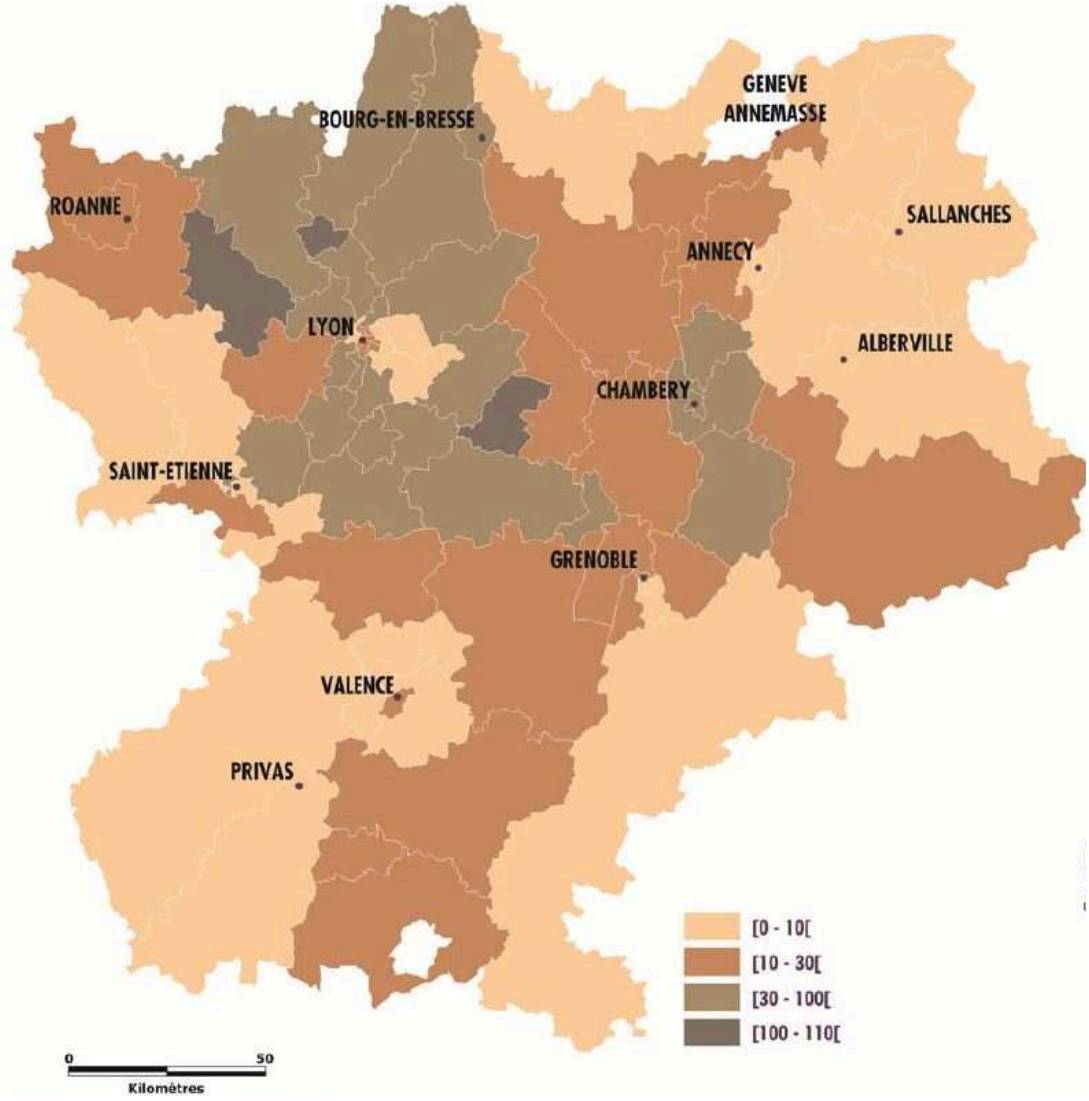
Source : Région Rhône-Alpes EDR (2016)

FIGURE 3 – Répartition modale des déplacements des rhônalpins (de 11 ans et plus)

En termes de part modale, de fortes hétérogénéités sont observées. Elles sont tant géographiques - les parts modales les plus élevées pour les transports en commun sont observés dans les villes de Lyon, Villeurbanne, Saint-Etienne et Grenoble - que démographiques - par exemple, la part modale des transports en commun diminue avec la tranche d'âge. Les trains et cars régionaux sont plus particulièrement utilisés pour des motifs obligatoires, c'est-à-dire le travail et les études. Chaque jour, 1 000 rhône-alpins réalisent en moyenne 25 déplacements en train et car régionaux, d'une durée moyenne de 70 minutes avec une forte disparité régionale (voir figure 4). Les trains et cars régionaux sont empruntés au moins deux fois par mois par 6% des rhône-alpins et 77% des rhône-alpins ne les utilisent jamais.

Parmi les personnes ayant pris le train (régional ou non) moins d'un mois avant l'enquête, 63% l'ont

23. Les enquêtes photos permettent d'interroger les passagers de tous les trains et cars express régionaux (TER) de la Région Rhône-Alpes tous les trois ans. Les questions portent notamment sur la gare de montée et de descente, sur le titre de transport et sur les raisons du choix du TER.



Source : Région Rhône-Alpes EDR (2016)

FIGURE 4 – Nombre de déplacements en trains ou cars régionaux réalisés par 1 000 rhônealpins de 11 ans et plus selon la zone de résidence

fait durant un jour de semaine, hors vacances scolaires, ce qui est cohérent avec le fait que le motif soit à 42% un motif obligatoire (travail ou études). L'autre motif dominant est de visiter la famille ou les amis (30 %). La voiture est utilisée par près de la moitié des usagers du train en rabattement mais beaucoup moins en diffusion (24 %) pour laquelle les transports en commun urbain sont très utilisés (39%). Pour les longs déplacements (plus de 20 km), le choix de ne pas avoir utilisé le train est motivé par plusieurs raisons. En premier lieu vient l'insuffisance de l'offre de service ferroviaire, en termes de fréquence, horaires, temps, coût, changements, etc. Une deuxième raison vient des difficultés d'accès aux gares que ce soit en rabattement ou en diffusion. Une troisième raison relève de préférences individuelles

quant à l'usage de la voiture.

Malgré la richesse de l'EDR, un modèle de choix modal basé sur ces seuls données ne peut être complet. Premièrement, l'EDR ne permet pas d'incorporer les composantes comportementales nécessaires à la modélisation du choix modal, et, en particulier des variables psychologiques telles que la conscience environnementale, la perception du confort ou les motivations affectives et symboliques d'utilisation de la voiture. Deuxièmement, l'EDR ne permet pas non plus de modéliser comment les personnes se comportent face à un choix dans le risque du fait d'un temps de déplacement incertain. Troisièmement, la description de l'offre ferroviaire n'y est pas suffisante. Des caractéristiques de l'offre telles que le cadencement, les retards ou la garantie d'avoir accès à une place assise sont des déterminants importants du choix qui, en l'état actuel des données disponibles, ne peuvent pas être décrits précisément pour chaque couple origine-destination à l'échelle régionale. Par ailleurs, même si une description était possible, dans le monde réel, les caractéristiques des alternatives ne varient pas suffisamment (par exemple, la fréquence sur une ligne ferroviaire est toujours la même), sont corrélées (par exemple, le temps et le coût de trajet), ne sont pas toujours observables ou difficiles à exprimer en unités objectives (par exemple, le confort). Il est également difficile d'évaluer la demande dans des conditions qui n'existent pas encore (par exemple, une nouvelle grille horaire introduisant le cadencement). Il est donc nécessaire de compléter l'EDR par une EPD afin de valoriser les caractéristiques des alternatives qui ne peuvent pas être approchées par l'EDR, d'introduire des situations de risque sur le temps de déplacement et de mesurer des variables psychologiques pouvant expliquer le choix modal.

Plan

Les chapitres 1, 2 et 3 reprennent le rapport que j'ai réalisé pour la Région Rhône-Alpes lors du développement de l'EPD[24]. Le chapitre 1 porte sur une revue de littérature sur les EPD. Le chapitre 2 présente la mise en œuvre concrète de l'EPD en Région Rhône-Alpes. Le chapitre 3 porte sur une première analyse des résultats de l'EPD.

24. Ce rapport pose les bases d'une bonne compréhension de l'EPD et de ses enjeux. Quatre questions récurrentes se posent lors du développement d'une EPD : 1) le mode et le protocole d'enquête, ainsi que la qualité des données qui est associé ; 2) le nombre d'attributs qui peuvent être insérés avec un arbitrage entre complexité et pertinence ; 3) les étapes préparatoires à l'enquête ; 4) le design de l'enquête. Ces questions sont donc abordées de façon générale dans le chapitre 1, sans pour autant prétendre donner de réponses définitives puisqu'elles sont à adapter à chaque situation. Seul le corps du rapport, réalisé dans le cadre d'une assistance à maîtrise d'ouvrage, est repris. Les annexes, détaillant toute la préparation et la mise en œuvre de l'EPD ainsi que l'apurement des données, sont résumées en notes de bas de page. Le rapport est disponible sur demande.

Discussion et limites

Grâce à la maîtrise de l'ensemble du processus d'enquête, de l'analyse qualitative à la conception du questionnaire, en passant par la définition du protocole d'enquête, la qualité et le contenu des données issues de l'EPD sont contrôlés. Cette maîtrise permet de réaliser une analyse statistique adaptée aux objectifs, tout en restant bien conscient des forces et limites des données. Par ailleurs, l'ensemble du processus d'enquête a été réalisé sur la base d'une revue de littérature détaillée et donc au plus proche de l'état de l'art. Cinq enseignements en particulier peuvent être retenus de la revue de littérature :

1. **L'impact du mode d'enquête sur la qualité des données**. Deux modes d'enquête (enquête web et enquête en face-à-face) ont été utilisés afin de tirer partie des avantages de chacun d'entre eux[25]. En mobilisant les répondants à l'EDR, l'enquête web permet de réaliser un gros volume d'enquête pour un coût limité et de contrôler (partiellement) la représentativité de l'échantillon. L'enquête en face-à-face permet de sur-échantillonner les usagers du train ;
2. **L'intérêt d'une analyse qualitative**, outil indispensable pour ancrer les choix méthodologiques, la formulation des questions et l'interprétation des résultats dans la réalité de terrain. Par exemple, les entretiens et les focus groups réalisés dans le cadre de l'EPD ont permis de préciser le choix des attributs à inclure dans l'enquête. Ils ont aussi permis d'affiner la description du cadencement. L'analyse qualitative a également éclairé les perceptions que les usagers ont des retards et l'ancrage des automobilistes à la voiture ;
3. **L'importance d'ancrer les choix dans la réalité vécue par les usagers afin de minimiser le biais hypothétique.** A cet effet, les questions de choix sont personnalisées sur les variables temps et coût de trajet, définies à partir d'un déplacement effectivement réalisé par les personnes interrogées. L'échelle d'analyse portant sur l'ensemble de la Région Rhône-Alpes (et non une unique ligne ferroviaire comme cela se rencontre en général dans la littérature), la personnalisation a nécessité un travail préparatoire pour calculer les temps et coûts de trajet sur l'ensemble des couples origine-destination de la Région, tant en train qu'en voiture ;
4. L'amélioration de l'efficacité des estimateurs grâce à un **design efficient bayésien**, avec pivot, lorsque la littérature est suffisamment fournie pour que les ordres de grandeur des coefficients associés aux attributs puissent être approchés. C'est donc la méthode qui a été utilisée sur la base d'une analyse de la littérature portant sur l'ensemble des attributs insérés dans l'enquête ;

25. Les questionnaires ont été développés afin de pouvoir comparer les résultats produits par les deux modes d'enquête. Cependant, ce travail est laissé pour de futures recherches.

5. La façon la plus adaptée (pour notre problématique) d'intégrer la **fiabilité** dans l'enquête, au travers de temps et probabilités de retard variant selon les questions de choix et non d'un retard par rapport à l'horaire (de type *scheduling*).

Les résultats issus de l'analyse de l'EPD (ceux présentés en chapitre 3 comme ceux du reste de la thèse) sont conformes à la littérature et aux intuitions. Nous identifions néanmoins deux sources de faiblesse des données. La première est de ne pas permettre l'intégration pleine et entière des théories de psychologie comportementale identifiées dans l'introduction générale de la thèse. En effet, la nécessité de réduire la longueur du questionnaire, et une connaissance limitée de ces théories au moment de l'élaboration du questionnaire, n'ont pas permis d'intégrer l'ensemble des éléments nécessaires au test d'une théorie. Par ailleurs, les échelles utilisées sont issues de la littérature et ont fait l'objet de tests et de contrôle de cohérence préalable mais il aurait pu être plus pertinent d'utiliser telle quelle une échelle existante, notamment pour favoriser la comparabilité des résultats. La deuxième faiblesse vient du faible report modal des automobilistes vers le train ou le car, ce qui n'est pas spécifique à cette enquête. Cet ancrage avait été identifié lors des tests préalables et le design adapté en conséquence. Une question spécifique a été posée aux personnes choisissant systématiquement la même option afin de mieux comprendre les raisons de ce choix. Une analyse détaillée de cette question est nécessaire, ainsi qu'une analyse portant spécifiquement sur le (non) report modal des usagers de la voiture vers le train. Cela permettrait de distinguer les personnes n'ayant pas voulu « jouer le jeu » de l'enquête de celles ne pouvant objectivement pas prendre un autre mode de transport que leur mode de référence en raison de diverses contraintes (géographiques, familiales, niveau d'offre de service...) puis d'intégrer cette distinction dans un modèle de choix discret.

CHAPITRE 1

Enquêtes de préférences déclarées - état de l'art

1.1 Introduction

De nombreux articles et livres traitent des enquêtes de préférences déclarées (EPD), avec principalement des applications en transport : Ben-Akiva et Lerman (1985) ; Pearmain et al. (1991) ; Hensher (1994) ; Louviere et al. (2000) ; Ortuzar et al. (2011). Après avoir justifié l'emploi d'une telle méthode dans cette introduction, le chapitre reprend les principaux éléments de la littérature dédiée aux EPD pour apporter des éléments de réponses aux questions qui sont posées lors de l'élaboration de l'enquête en Région Rhône-Alpes (mode d'enquête, nombre d'attributs à insérer, étapes préparatoires, design).

La mise en œuvre d'une EPD repose en effet sur la réalisation d'une enquête qui nécessite d'avoir auparavant déterminé le mode et le protocole d'enquête (voir section 1.2), construit le questionnaire (voir section 1.3) et en particulier d'avoir construit les questions de choix au travers d'un plan factoriel ou design (voir section 1.4). La construction du questionnaire et du design nécessitent d'avoir choisi les attributs et les niveaux d'attributs décrivant l'offre. La section 1.5 fait une revue de littérature sur les attributs insérés dans l'enquête.

Les méthodes existantes

Pour quantifier l'effet des déterminants individuels et des caractéristiques de l'offre sur le choix modal, deux familles de méthodes existent. Une première famille de méthode, appelées méthodes de préférences révélées, s'appuie sur l'observation des choix effectivement réalisés par les individus et à en déduire la valeur (monétaire ou temporelle) de chacun des déterminants évalués. Les méthodes de préférences révélées peuvent s'appuyer sur des données telles que celles collectées dans l'EDR. Elles

ont l'avantage de se baser sur des choix qui ont effectivement été réalisés et ne sont donc pas entachées d'un biais hypothétique. En revanche, seule la valorisation de caractéristiques existantes est possible. D'autres méthodes de préférences révélées existent telles que la méthode des prix hédoniques ou la méthode des coûts de transport. Elles sont principalement utilisées pour valoriser des biens et services environnementaux.

Une deuxième famille de méthodes, appelées méthodes de préférences déclarées, s'appuie sur des choix hypothétiques proposés spécifiquement pour l'enquête. Des options de transport n'existant pas (du moins pas aujourd'hui) peuvent donc être présentées permettant ainsi d'anticiper l'évolution de la demande en cas d'évolution de l'offre. Cette approche est présenté plus en détails ci-dessous.

Les méthodes fondées sur les préférences déclarées s'appuient sur des enquêtes afin d'estimer la valeur accordée à certaines variations de l'offre de biens (Pearce et al., 2006). Elles sont utilisées pour évaluer des biens non-marchands (valeur environnementale), des situations hypothétiques ou encore des biens marchands dont l'évaluation est délicate à travers des enquêtes de préférences révélées. Deux approches sont couramment utilisées :

- L'**évaluation contingente** : après une mise en situation, le questionnaire comporte une question demandant directement à la personne interrogée quel est son consentement à payer (ou consentement à recevoir) pour l'amélioration (ou la dégradation) qui lui a été décrite, par exemple, l'amélioration de la qualité de l'eau d'une nappe phréatique (Aulong et al., 2006) ou l'atténuation du bruit d'un aéroport (Faburel, 2002).
- Un ensemble de méthodes se basant sur la **modélisation de choix hypothétiques** (voir tableau 1.1) : issues du marketing, ces méthodes demandent à la personne interrogée de choisir entre plusieurs alternatives[1] décomposées en attributs, établissant ainsi des priorités parmi différentes composantes du bien ou du service évalué. Le coût correspondant étant l'une des composantes du choix, l'arbitrage effectué par les personnes permet de déduire la valeur de chacune des composantes et donc la valeur totale de l'alternative (Lancaster, 1966). Cette approche regroupe en fait quatre méthodes distinctes, toutes basées sur des choix hypothétiques entre plusieurs alternatives dont la plus courante est la méthode des choix expérimentaux (ou expérimentation des choix).

Comparativement à l'ensemble des méthodes de classement, comparaison ou notation, la méthode des choix expérimentaux est à privilégier car c'est la seule qui se base sur une théorie comportementale clairement établie et étayée, la théorie de l'utilité aléatoire, permettant d'expliquer le

1. Le terme alternative est équivalent à celui de scénario ou option utilisé dans certaines citations.

TABLEAU 1.1 – Variantes des méthodes de modélisation des choix

Approche (français)	Approche (anglais)	Tâche à accomplir	Estimation cohérente du bien-être
Expérimentation des choix	Choice experiment	Choisir entre deux options ou davantage (dont une correspondant au statu quo)	Oui
Classement contingent	Contingent ranking	Classer une série d'options	Cela dépend[1]
Notation contingente	Contingent rating	Noter de 1 à 10 les différentes alternatives	Ce n'est pas sûr
Comparaison par paires	Pairwise comparison	Noter des paires d'alternatives sur une échelle de 1 à 10	Ce n'est pas sûr

[1] Le classement de l'ensemble des alternatives peut théoriquement révéler les relations de préférences entre les alternatives. Cependant, le classement de l'ensemble des alternatives est un exercice difficile. Selon Pearce et al. (2006, p. 148), « les choix paraissent peu fiables et cohérents du point de vue du l'ordre de classement des différentes options ». Voir Beaumais et al. (2016) pour une discussion plus approfondie.

Source : Adapté de Pearce et al. (2006)

comportement et le choix des individus et donc d'aboutir à une estimation cohérente du bien-être. Comme les méthodes basées sur une notation se basent non pas sur une théorie économique mais sur une théorie mathématique[2], le lien avec l'utilité est créé artificiellement par le biais de règles, ce qui rend plus difficile une estimation cohérente du bien-être. Par exemple, Pearce et al. (2006) notent à propos de la notation contingente que « Ses applications dans le domaine du marketing ont jusqu'à présent généralement formulé une hypothèse hardie concernant la cardinalité de l'échelle de notation (une notation de 8 étant ainsi présumée représenter deux fois plus d'utilité que si elle avait été de 4). »

Justification du choix

La méthode mise en œuvre pour répondre aux objectifs de la thèse est la méthode d'expérimentation des choix. Ce choix s'explique par deux raisons. Premièrement, l'objectif étant de modéliser le choix modal, l'emploi d'une méthode décomposant le choix du mode de transport en attributs et permettant de valoriser des attributs tels que les retards ou le confort est donc nécessaire. La méthode de l'évaluation contingente est donc exclue.

Deuxièmement, parmi les méthodes basées sur des choix hypothétiques, la méthode des choix expérimentaux est celle qui permet la meilleure estimation du bien-être. C'est la méthode la mieux fondée micro-économiquement et celle qui est couramment utilisée en économie des transports et qui bénéfice donc d'un retour d'expérience permettant d'assurer la fiabilité des résultats. Par la suite, l'appellation « Enquête de Préférences Déclarées » (EPD) fait systématiquement référence à la

2. Pour une discussion, voir Louviere et al. (2010).

méthode des choix expérimentaux.

1.2 Protocole d'enquête

Le choix du protocole d'enquête est un véritable enjeu puisqu'il influence à la fois l'échantillonnage (qui répond ?) et la formulation des réponses (comment la personne répond-elle ?). Il est notamment conditionné par le mode d'enquête. La section 1.2.1 présente les différents modes d'enquêtes envisageables et la section 1.2.2 s'interroge sur la qualité des données produites par les différents modes d'enquête. Notons toutefois qu'il est difficile de conclure sur la pertinence de l'un ou l'autre des modes d'enquête, dans la mesure où ils se complètent, permettant d'accéder à et de faire répondre des populations différentes [3].

1.2.1 Les modes d'enquête : des pratiques qui évoluent en faveur des enquêtes web

Le mode d'enquête désigne la façon dont les personnes interrogées prennent connaissance du questionnaire. Le questionnaire d'une EPD peut être auto-administré ou administré par un enquêteur.

Dans les enquêtes auto-administrées, les personnes interrogées remplissent elles-mêmes le questionnaire, sans intervention extérieure. Le support peut être papier, envoyé par courrier ou distribué, par exemple dans le train (méthode utilisée par exemple par Börjesson et Eliasson, 2011). Par ailleurs, les enquêtes web sont en plein développement. Ils permettent d'automatiser la saisie des données et donc de limiter les erreurs. En revanche, les enquêtes web supposent un accès internet, conduisant à un biais d'échantillonnage (Lindhjem et Navrud, 2011 ; Bayart et Bonnel, 2010 ; voir section 1.2.2). Les enquêtes auto-administrées ont un faible coût et permettent d'utiliser un support visuel.

Les enquêtes administrées se font soit en face-à-face (*Computer Assisted Personal Interview* ; CAPI), soit par téléphone (*Computer Assisted Telephonic Interview* ; CATI). Elles permettent de guider les interviewés dans leur compréhension du questionnaire. Les enquêtes en face-à-face permettent d'utiliser un support visuel, comme les enquêtes auto-administrées. Cependant, l'investissement (temporel et financier) est plus important, ce qui, à coût égal réduit la taille de l'échantillon. Par ailleurs, les enquêtes en face-à-face nécessitent généralement de se rendre sur le terrain (dans les gares ou les trains), ce qui impose l'utilisation de la méthode des quotas. Les enquêtes téléphoniques sont

3. Pour des exemples d'évaluation et de comparaison de différents modes d'enquêtes et de combinaison de ces modes, voir Bowling (2005) ; Jäckle et al. (2010) ; Bayart et Bonnel (2010) ; Lindhjem et Navrud (2011).

moins coûteuses que les enquêtes en face-à-face. Il est cependant difficile de faire passer des éléments d'explications complexes par téléphone, d'autant plus qu'il n'y a pas d'appui visuel possible (à moins de l'envoi d'un questionnaire au préalable, méthode utilisée par exemple par Pons, 2011).

Jusqu'à la fin des années 90, il existait un consensus selon lequel les réponses sont de meilleure qualité lorsqu'un enquêteur interagit avec l'enquêté. Plusieurs auteurs se prononcent alors clairement en faveur des enquêtes en face-à-face dans le cadre d'évaluations contingentes. Le NOAA Panel, interrogé sur la validité des évaluations contingentes pour rendre une décision de justice concernant la marée noire causée par l'Exxon Valdez, déclarait qu'« il est peu probable que des estimations fiables puissent être produites par un questionnaire courrier. Les enquêtes en face-à-face sont généralement préférables, même si les entretiens téléphoniques ont certains avantages en termes de coûts et de supervision centralisée » (Arrow et al., 1993, p.30).

Cependant, depuis une quinzaine d'années, les enquêtes web se multiplient, l'accès à internet s'est largement démocratisé et le rapport au courrier et au téléphone a évolué. La question du mode d'enquête se pose donc en d'autres termes que dans les années 80 et 90. Plusieurs études comparatives permettent d'étayer ce constat (par exemple, Lindhjem et Navrud, 2011 ; Bowling, 2005).

1.2.2 Qualité des données et types d'erreurs

L'objectif de la définition du protocole d'enquête est d'obtenir des données de la meilleure qualité possible. Cependant, définir ce que sont des données de bonnes qualité est complexe (pour une discussion, voir Bergdahl et al., 2007). D'après l'institut Eurostat, la qualité des données peut s'évaluer au regard de six critères (Dussaix, 2009) :

- La pertinence (*relevance*) traduit dans quelle mesure les résultats de l'enquête répondent aux besoins courants et potentiels des utilisateurs ;
- L'exactitude (*accuracy*) mesure la proximité des résultats ou des estimations avec la valeur vraie inconnue de la variable étudiée ;
- Le laps de temps entre l'enquête, l'obtention des résultats et leur diffusion (*timeliness and punctuality*) ;
- L'accessibilité et l'intelligibilité (*accessibility and clarity*) des données statistiques ;
- La comparabilité (*comparability*) mesure si les différences de résultats entre zones géographiques ou périodes peuvent être attribuées à des différences réelles et non pas à l'évolution des méthodologies ;
- La cohérence (*coherence*) traduit l'aptitude des données à être combinées de différentes façons et

pour différentes utilisations.

Différents types d'erreurs, qui peuvent être organisés autour de quatre dimensions (Dillman et Bowker, 2001), jouent sur l'exactitude des données :

- L'**erreur d'échantillonnage** due au processus de sélection de l'échantillon. Elle se mesure notamment à partir du nombre de répondants et correspond à la différence entre la valeur du paramètre étudié sur la population cible et la valeur du paramètre estimé sur l'échantillon sélectionné ;
- L'**erreur de couverture** qui apparaît lorsque certaines personnes de la population cible ont une probabilité nulle d'être interrogées, soit parce que la population définie par la base de sondage ne coïncide pas avec la population que l'on désire réellement étudier (par exemple, du fait de l'utilisation de l'annuaire téléphonique), soit parce que la base de sondage comporte des erreurs (par exemple, à cause d'oublis ou d'adresses fausses) ;
- L'**erreur de non-réponse** du fait de personnes faisant partie des personnes à enquêter mais refusant de répondre (non-réponse intégrale) ou répondant seulement à une partie du questionnaire (non-réponse partielle). Elle entraine un biais de sélection, les personnes répondant n'ayant pas les mêmes caractéristiques que celles ne répondant pas (ou que partiellement). Par exemple, les personnes ayant des enfants en bas âge peineront certainement à trouver du temps pour accueillir un enquêteur chez elles (Bayart, 2009) ;
- Les **erreurs de mesure** qui proviennent de l'inexactitude des réponses fournies. Elles peuvent être imputables à l'enquêteur, à l'enquêté ou au questionnaire.

Les trois premières erreurs influencent la constitution même de l'échantillon sur la base duquel sont conduites les analyses économétriques. Elles sont pour partie liées au mode d'enquête, et ce pour trois raisons. Premièrement, à budget fixe, le nombre de répondants ne peut pas être le même avec une enquête web et une enquête en face-à-face ce qui influence l'erreur d'échantillonnage. Deuxièmement, selon le mode d'enquête, certaines personnes pourront avoir une probabilité nulle (erreur de couverture) ou plus faible d'être interrogées, l'accès à certaines catégories de population étant plus difficile. Par exemple, les personnes âgées seront plus difficilement accessibles *via* une enquête web (Bayart, 2009), les personnes jeunes n'auront pas forcément de téléphone fixe, etc. Troisièmement, les non-réponses partielles sont moins importantes dans les enquêtes administrées, les personnes interrogées allant généralement jusqu'au bout du questionnaire.

Concernant plus particulièrement les erreurs de mesure, le mode d'enquête influence la façon dont

la personne interrogée comprend la question, traite les informations dont elle dispose, se fait une idée de la réponse à apporter et formule cette réponse (Roberts, 2007). Un consensus se dégage selon lequel les préférences ne sont pas tant révélées par l'enquête mais plutôt « découvertes » ou construites au cours de l'enquête (Carlsson, 2010). Les données sont donc le produit même de l'enquête, elles ne sont pas recueillies mais émergent de l'interaction entre la personne interrogée, le questionnaire et, éventuellement, l'enquêteur. Or le mode d'enquête exerce une influence sur ces interactions, influence qui peut jouer positivement ou négativement sur la qualité des données au travers de deux « facteurs humains ». Le premier joue au travers des valeurs de la personne interrogée et guide, consciemment ou inconsciemment, la formulation de ses réponses et joue donc sur la validité de l'enquête. Le deuxième repose sur les capacités cognitives de la personne interrogée.

1.2.2.1 Validité de l'enquête

Pour que l'enquête soit valide, les personnes interrogées doivent faire un choix qui correspond au comportement qu'elles auraient effectivement réalisé dans une situation réelle. Pourtant, l'essence même du questionnaire lui confère un caractère hypothétique (Loomis, 2011) et des précautions doivent être prises dans le développement du questionnaire afin de diminuer la portée de ce biais : adapter le questionnaire à chaque individu, rendre les propositions les plus réalistes possibles, vérifier que la personne interrogée les juge crédibles, etc. Ce biais hypothétique peut conduire la personne interrogée à déformer ses réponses, consciemment ou non. Une difficulté consiste alors à obtenir des réponses sincères, c'est-à-dire non influencées par le désir de se conformer à une image idéalisée ou par une volonté stratégique.

Selon que le questionnaire soit auto-administré ou non, certaines personnes ont tendance à formuler des réponses les mettant en valeur ou correspondant à des normes sociales telles que l'altruisme ou la conscience environnementale. Ce biais de désirabilité sociale, connu sous les termes anglo-saxons de *compliance bias*, *social-desirability bias* ou encore *warm-glow effect*, est important à contrôler en économie de l'environnement mais plus réduit en économie des transport, notamment parce que les choix de l'enquête sont proches de ceux faits dans la réalité.

D'autres biais peuvent apparaître, tel que le biais stratégique notamment analysé par Fosgerau et al. (2010). Les répondants peuvent penser que leurs réponses influencent les décisions politiques. Par exemple, les usagers des transports en commun peuvent exprimer un faible consentement à payer en espérant une diminution des prix. Du fait de ces comportements stratégiques, les répondent n'expriment pas leurs vraies préférences.

1.2.2.2 Adéquation entre effort demandé, effort consenti et capacités cognitives

Répondre à un questionnaire d'EPD, est un exercice difficile. Souvent relativement long, avec des questions portant sur des arbitrages entre différents attributs, le questionnaire peut représenter un réel obstacle pour certaines personnes. Le risque en cas d'inadéquation entre effort requis et effort consenti par la personne interrogée est d'obtenir des réponses erronées aux questions (réponses rapides, au hasard) et que la personne interrogée élude certaines questions ou parties du questionnaire. La capacité à répondre au questionnaire résulte donc d'une combinaison entre difficulté de l'exercice, aptitudes cognitives et motivation, ces trois facteurs étant étroitement liés au mode de passation du questionnaire.

En première intuition, plus l'exercice demandé est difficile, plus il y a de chances que les réponses soient de mauvaise qualité, la difficulté s'accroissant avec le nombre d'attributs, le nombre de niveaux et le nombre de choix (voir section 1.3). Cette intuition est corroborée par plusieurs chercheurs (par exemple, Lindhjem et Navrud, 2011). Cependant, d'autres auteurs ont une position plus nuancée. Ainsi, selon Malhotra (2009), cette vision est trop simpliste car les personnes interrogées peuvent être plus motivées par l'accomplissement de tâches complexes. Parallèlement, Louviere et al. (2010) suggère d'approfondir la recherche empirique sur les effets du niveau de complexité sur les réponses et les limites cognitives des répondants. Pour plus de détails, voir la section 1.3.3.

Deuxièmement, les aptitudes cognitives font référence aux capacités de traitement de l'information, d'attention et d'abstraction des personnes interrogées. Elles sont dépendantes du niveau d'étude et les personnes ayant les capacités cognitives les plus faibles sont davantage dépendantes du contexte dans lequel les questions sont posées (Toepoel et al., 2008) et donc du mode d'enquête. Les aptitudes cognitives peuvent également varier selon la facilité de chacun à réagir plutôt à des stimuli auditifs ou visuels. Une double stimulation - auditive par le biais d'un enquêteur et visuelle par le biais de supports imprimés – pourrait donc constituer un avantage important pour s'assurer de la bonne compréhension du questionnaire et de la capacité à y répondre. L'intervention d'un enquêteur permet notamment de répéter les questions non comprises, voire de les reformuler. Il est important que la répétition et/ou la reformulation éventuelles aient été prévues en amont lors de la conception du questionnaire afin de s'assurer que tous les enquêteurs diffusent la même information (Bowling, 2005). Sur ce critère, les enquêtes en face-à-face sont à privilégier.

Troisièmement, à difficulté d'exercice et aptitudes cognitives égales, la motivation peut exercer une influence sur la qualité des réponses. Green et Tunstall (1999) mettent en avant le travail de l'enquêteur qui a un rôle d'entrainement vis-à-vis de la personne interrogée et qui peut la motiver à répondre à l'ensemble des questions en suscitant et en maintenant son intérêt. *A contrario*, les enquêtes en face-

à-face peuvent induire une pression importante sur la personne interrogée, l'incitant à couper court à certaines questions.

En définitive, trois leviers permettent au concepteur du questionnaire d'optimiser l'adéquation entre motivation, capacités cognitives et effort demandé, ces trois leviers étant dépendants du mode de passation des questionnaires :

- fournir des supports visuels ergonomiques (tableau de choix, illustration des résultats, etc.) ;
- simplifier le questionnaire, le rendre le plus fluide possible. Pour les enquêtes web en particulier, Dillman et Bowker (2001) définissent 14 principes à respecter pour assurer une conception et une présentation optimales de l'enquête [4] ;
- s'appuyer sur un enquêteur pour motiver la personne interrogée et lui faciliter la compréhension des questions.

1.3 Elaboration du questionnaire

Le questionnaire détermine le contenu de l'enquête. Son élaboration est donc une étape clé pour répondre aux objectifs de l'enquête, avec des données fiables et de qualité. Chaque choix doit être à la fois justifié par la littérature et permettre de répondre aux objectifs. Cette section discute des étapes d'élaboration du questionnaire, des différentes composantes d'un questionnaire d'EPD, du nécessaire arbitrage entre charge cognitive et pertinence et de l'analyse qualitative à réaliser en amont de l'enquête.

1.3.1 Les étapes d'élaboration

> « [...] Les études [de choix expérimentaux] débutent par l'identification d'attributs ad hoc et spécifiques à chaque chercheur. Les méthodes d'identification vont de questions directement posées aux personnes quant aux attributs qui expliquent leurs préférences (par exemple, Louviere, 1988) à des méthodes sophistiquées, à la limite du théorique, comme la grille d'interprétation de Kelly (par exemple, Timmermans et al., 1982). Il n'y a donc pas de façon standard et acceptée par tous de choisir les attributs. [...] Une fois les attributs identifiés, le chercheur doit y associer des niveaux/des valeurs pour exprimer une étendue de variation (observée ou supposée) de ces attributs dans le contexte étudié. Comme pour l'identification des attributs, il n'y a pas de consensus sur la façon d'y parvenir, ce qui explique l'étendue des pratiques. » Louviere et al. (2010, p.60)

4. Par exemple, choisir en premier une question susceptible d'intéresser la plupart des répondants, à laquelle il est facile de répondre, et complètement visible sur le premier écran du questionnaire.

Ce constat montre la difficulté, voire l'impossibilité, de suivre des étapes fiables et éprouvées pour choisir les attributs et leurs niveaux. Cependant, se poser les bonnes questions aide à guider les choix et à construire le questionnaire avec un contenu adapté aux objectifs, au contexte et aux contraintes de l'enquête tout en tirant parti des enseignements de la littérature. La première série de questions porte sur l'objet même de l'enquête. Que cherche-t-on à mesurer ? A quelles questions veut-on apporter des réponses ? La deuxième série de questions porte sur les perceptions et les connaissances que les usagers ou citoyens ont de l'objet à évaluer. Qu'est-ce qui sembler influencer leurs choix, leur comportement ? Qu'est-ce qui semble important et pourquoi ? Répondre à ces questions permet de réaliser un questionnaire adapté à la réalité de terrain, à la fois sur le fond (choix des attributs, niveaux) et sur la forme (explications, représentations graphiques des attributs). C'est notamment dans cet objectif que des focus groups et entretiens individuels sont proposés (voir section 1.3.4). La troisième série de questions s'intéresse à la bibliographie : que nous apprend la littérature sur le choix et la définition des attributs ? Quels choix ont été effectués sur des EPD aux thématiques similaires ? Ces questions sont abordées dans la section 1.5 pour l'application au choix modal.

1.3.2 Définitions et méthode

Dans le domaine des transports, un questionnaire d'EPD est généralement constitué de trois parties. Une partie consacrée à la description d'un déplacement réalisé par le passé, une partie consacrée aux questions de choix et une partie décrivant la personne. La deuxième partie, spécifique aux EPD, nécessite d'identifier les attributs, c'est-à-dire les variables décrivant l'offre, de les définir et de les mesurer avant de les combiner en questions de choix à l'aide du design. Ce sont les éléments de cette deuxième partie qui sont définis et discutés ci-dessous. La troisième partie recueille des informations sur la personne interrogée (variables socio-économiques et psychologiques).

1.3.2.1 Attributs

On nomme « attributs » les composantes ou caractéristiques du bien, du service ou de la situation évalué(e) dans l'EPD. Les attributs choisis doivent respecter certains critères. Pearmain et al. (1991) en recensent notamment trois :

- Être représentatifs de l'objet évalué : les attributs choisis sont la seule « vitrine » de l'objet évalué, ils doivent donc permettre de le décrire aussi complètement et concisément que possible [5] ;

5. La section 1.3.3 discute de l'arbitrage entre cette condition de représentativité et la complexité du choix demandé aux personnes interrogées afin d'apporter un éclairage sur le nombre d'attributs qui peut être intégré dans une EPD.

- Avoir une réelle influence sur le choix des individus : ils ne doivent pas seulement être importants pour l'étude mais traduire de manière pertinente les arbitrages que feraient les personnes interrogées dans la réalité ;
- Être indépendants entre eux : la présence de l'un ne doit pas systématiquement impliquer celle de l'autre afin d'éviter des problèmes de colinéarité.

Avant de le décomposer en attributs, il est donc nécessaire de bien définir l'objet évalué et de comprendre la perception qu'en ont les personnes qui vont répondre à l'enquête. Sur cette base, le choix des attributs peut s'appuyer sur diverses sources de données et d'informations :

- Des données sur les comportements révélés de la population face à une situation de choix identique ou proche (par exemple, par le biais de l'EDR ou des enquêtes photos) ;
- Les enseignements d'études passées portant sur une situation de choix proche (voir section 1.5) ;
- Des processus de recueil des données qualitatives tels que les focus groups et les entretiens en face-à-face (voir section 1.3.4).

Dans tous les cas, il est essentiel que l'un des attributs choisis soit un attribut monétaire, ou éventuellement temporel, afin de pouvoir traduire les autres attributs dans une unité de mesure commune (euros ou minutes).

1.3.2.2 Niveau et mesure des attributs

Les niveaux des attributs sont les différentes valeurs que les attributs prennent dans les alternatives proposées aux personnes interrogées. Les niveaux peuvent être mesurés sur une échelle continue (temps, coût...) ou discrète (niveau de confort, information, fréquence de passage...). Le choix des niveaux d'attributs est important car il conditionne l'interprétation et les résultats de l'analyse. En effet, l'évaluation d'un attribut dépend de la définition des niveaux choisis pour cet attribut. Pour s'assurer que les réponses au questionnaire correspondent bien à ce que le modélisateur souhaite, et non aux connaissances et a priori sous-jacents du répondant, les niveaux doivent être décrits le plus explicitement et clairement possible. Les niveaux choisis peuvent être fixés en valeur absolue pour tous les répondants ou varier selon le répondant selon la méthode pivot (voir section 1.4). Dans ce deuxième cas, les niveaux sont personnalisés selon une situation réelle rencontrée par le répondant. Le niveau de l'attribut est alors exprimé en variation (en niveau ou en %) relativement au niveau annoncé ou vécu par la personne interrogée. Ainsi, pour le temps de trajet en train, il est possible de pré-déterminer trois temps de trajet en valeur absolue (15 minutes, 30 minutes ou 45 minutes) ou de

demander aux personnes interrogées combien de temps elles ont passé dans le train lors de leur dernier déplacement et de leur présenter des temps de trajet correspondant à +/- 50% ou +/- 10 minutes de ce temps de déplacement.

Pearmain et al. (1991) recensent quatre critères de choix pour le niveau des attributs. Ils doivent apparaître comme plausibles ; être reliés à l'expérience que le répondant a de chaque attribut ; générer un réel arbitrage lorsque le choix est réalisé ; couvrir le champ des valeurs attribuées par chacun des répondants. Afin de s'assurer que les niveaux relatifs des attributs traduisent des taux de substitution réalistes, Fowkes et Wardman (1988) et Fowkes (1998) proposent d'identifier des « valeurs frontières [6] » (*boundary values*) issues de la littérature et qui donnent une première idée des ordres de grandeur que l'on souhaite mesurer, puis de vérifier si les niveaux et arbitrages proposés permettent de couvrir ces valeurs. En plus de la littérature, les phases préparatoires (entretiens, focus-groups, tests) aident également à approcher ces ordres de grandeur.

Cette méthode peut être généralisée en simulant une base de données de réponses à partir des questions de choix générées pour les niveaux d'attributs choisis et en vérifiant que les résultats d'estimation produisent des estimations plausibles. Il est également essentiel de réaliser des tests visant à vérifier que l'ensemble des alternatives est bien choisi et donc que les niveaux et leurs valeurs relatives sont bien réalistes. Les deux méthodes (valeurs frontières et simulation de données) ont utilisées lors de la réalisation de l'EPD à l'échelle de la Région Rhône-Alpes.

1.3.2.3 Alternatives et questions de choix

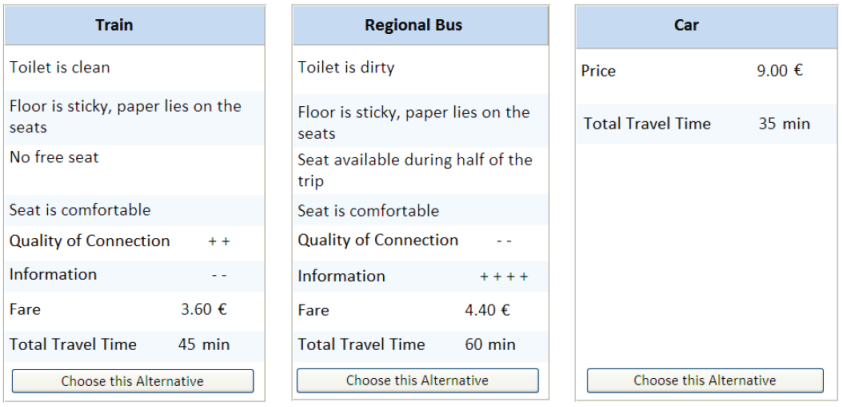
Une alternative (ou option) résulte de la combinaison des attributs et de leurs niveaux. La figure 1.1 représente trois alternatives possibles définies par trois modes de transport. L'ensemble constitué de ces trois alternatives est une question de choix. Les attributs choisis peuvent être communs à toutes les alternatives proposées (exemple du prix et du temps de trajet) ou spécifiques à l'une des alternatives (exemple de la disponibilité d'un siège).

Une question de choix peut comprendre de deux à cinq alternatives. Comme pour le nombre d'attributs, un arbitrage entre complexité et pertinence des alternatives présentées est à réaliser.

1.3.2.4 Exercice de choix, blocs et versions de questionnaire

Pour éviter un questionnaire trop long, il est possible de distribuer les questions de choix en « blocs ». Chaque bloc fait alors l'objet d'une version de questionnaire spécifique.

6. Ce sont également les *a priori* qui pourront être utilisés lors du design (voir section 1.4).



Source : Richter et Keuchel (2012)

FIGURE 1.1 – Exemple de question de choix

Un même questionnaire peut également contenir plusieurs exercices de choix. Si le nombre d'attributs retenus est trop important pour tous les combiner dans une même alternative, alors il est possible d'en retenir certaines pour un premier exercice de choix (série de questions de choix portant sur les mêmes attributs), puis de soumettre les autres dans un deuxième, voire un troisième exercice de choix (voir, par exemple, RFF, 2013). C'est l'option retenue pour l'EPD à l'échelle de la Région Rhône-Alpes.

Une solution alternative est de découper l'enquête en différentes versions qui se concentrent plus spécifiquement sur l'une ou l'autre des composantes de choix (voir, par exemple, Richter et Keuchel, 2012 ; Arentze et Molin, 2013). Quelle que soit la méthode choisie, elle doit prendre en compte les effets d'échelle résultant d'échantillons et de contextes de choix différents.

1.3.3 Arbitrage charge cognitive/pertinence

Le choix des attributs représente une difficulté : ils doivent être judicieusement sélectionnés afin de ne pas dépasser les capacités cognitives des personnes interrogées tout en étant suffisamment exhaustifs pour décrire de manière réaliste et pertinente le bien évalué. Se pose dès lors la question du nombre d'attributs à intégrer dans une EPD[7].

Les sections suivantes proposent d'aborder la question du choix et du nombre d'attributs sous plusieurs angles avant de conclure (section 1.3.3.6).

7. De même, le nombre d'alternatives à présenter dans une même question de choix pose la question de la capacité à analyser simultanément un grand nombre d'alternatives.

1.3.3.1 Impacts de la complexité des questions et de l'omission d'un attribut

Le nécessaire arbitrage entre charge cognitive et pertinence des attributs sélectionnés résulte d'un double constat :

- La complexité des questions de choix induit une charge cognitive qui peut limiter la qualité des réponses : réponses aléatoires, choix systématiques, non-réponses, etc. Les répondants auraient également davantage tendance à développer des heuristiques de choix, par exemple, en fixant leur attention sur un nombre limité d'attributs.
- L'omission d'un attribut important dans le choix des répondants réduit le réalisme de l'enquête et donc sa crédibilité. Par ailleurs, les répondants prendront en compte cet attribut manquant dans leur choix sans qu'il ne soit possible de le contrôler puisque le niveau pris en compte est alors inconnu du modélisateur.

1.3.3.2 La complexité ne se limite pas au nombre d'attributs

Hensher (2006a) propose une interprétation nuancée de la capacité cognitive des répondants et montre que la complexité n'est pas uniquement la résultante du nombre d'attributs mais d'une combinaison entre niveaux, variations de niveaux et nombre d'alternatives présentées. Il teste plusieurs EPD qui diffèrent par le nombre d'alternatives (deux à quatre), le nombre d'attributs (trois à six), le nombre de niveaux par attribut (deux à quatre) et l'étendue des niveaux (base, plus large que la base, moins large). Les six attributs sont le temps sans congestion, le temps avec congestion, le temps au départ et à l'arrivée, la variabilité de temps de trajet, le coût du péage et le coût au kilomètre. Ces attributs sont progressivement regroupés pour aboutir à des versions avec cinq, quatre et trois attributs. Par exemple, dans la version avec trois attributs, les trois temps de trajets sont sommés ainsi que les deux coûts. Il conclue que la probabilité d'omettre certains attributs dans le processus décisionnel augmente lorsque le nombre de niveaux par attribut augmente, lorsque ces niveaux diffèrent peu entre eux (faible étendue) et lorsque le nombre d'alternatives diminue. Hensher se garde pourtant de conclure quant à l'existence d'un design optimal. Il met plutôt en avant l'hétérogénéité avec laquelle les attributs sont considérés dans les processus décisionnels adoptés par les personnes de l'échantillon. Certaines personnes restreignent le nombre d'attributs considérés (soit par un processus d'agrégation, soit par l'exclusion de certains attributs). Mais cela ne doit pas être interprété, selon lui, comme une réponse à la complexité, mais plutôt comme un processus naturel lors d'une prise de décision. Le processus décisionnel étant complexe, et *a priori* inconnu du modélisateur, il est préférable

de donner aux répondants l'information la plus complète et pertinente possible et de les autoriser à établir leurs propres règles de simplification[8] plutôt que des les imposer *a priori* en limitant la dimensionnalité de l'EPD.

1.3.3.3 La complexité est inhérente au processus décisionnel

Suivant la même idée, Louviere et al. (2010) soulignent que la complexité est inhérente au processus décisionnel et qu'omettre des attributs peut avoir des conséquences sur l'analyse.

> « Pour minimiser la complexité de la tâche, les chercheurs omettent souvent des attributs qui ne sont pas importants pour la plupart de la population mais qui peuvent avoir de larges effets pour un segment spécifique ou réduisent le nombre d'options par choix et le nombre d'étapes de choix. La recherche appliquée prend conscience que la simplification des choix n'est pas sans conséquence mais a souvent l'impression qu'il s'agit là d'un moindre mal. Cependant, en plus de 20 ans de recherche, il y a désormais des preuves selon lesquelles les répondants aux questions de choix expérimentaux peuvent effectuer des tâches plus complexes et plus longues avec peu de perte de fiabilité ou de robustesse » (Louviere et al., 2010, p. 5).

La conclusion est similaire à celle de Hensher (2006b) : avant de conclure à une limitation cognitive sur la base de travaux émanant d'autres disciplines, il est nécessaire de faire des tests empiriques dans le contexte même de l'EPD mise en œuvre.

1.3.3.4 Impact de la présence d'attributs peu importants

Relativement aux travaux précédents qui soulignent l'importance d'intégrer tous les attributs pertinents dans le processus décisionnel, Hess (2014) propose de poser la question en sens inverse. Quel est l'impact de la présence d'attributs peu ou pas importants sur l'estimation des autres attributs ? Pour répondre à cette question, l'auteur réalise une EPD en deux étapes. Dans la première étape, chaque répondant répond à huit questions de choix avec chacune trois alternatives de déplacement en train décrites par cinq attributs : le coût et le temps de trajet, la garantie d'une place, l'accès au wifi et la flexibilité de changer le ticket de train. Les répondants devaient ensuite indiquer s'ils avaient ignoré certains attributs lors des questions de choix. Dans la deuxième étape, le répondant effectue à nouveau des choix sur la base uniquement des attributs pertinent pour lui. Hess conclut

8. Voir par exemple les travaux de Gigerenzer et Brighton (2009) qui défendent l'idée que les processus décisionnels basés sur des heuristiques de choix simples permettent de prendre de meilleures décisions que les processus décisionnels plus complexes.

qu'il n'y a pas de preuve que le fait de présenter des attributs non importants affecte indûment le choix que ce soit en termes d'échelle ou de consentement à payer.

1.3.3.5 Une grande hétérogénéité dans le nombre d'attributs

Une grande hétérogénéité est observée dans les EPD quant au nombre d'attributs sélectionnés par question de choix. Dans les EPD recensés pour la thèse, ce nombre va de trois (voir, par exemple, la figure 1.7f, p.69) à treize attributs (par exemple, Arentze et Molin, 2013, voir figure 1.2). Ce deuxième exemple vient d'une EPD avec un choix pour un déplacement multimodal de plus de 65 km.

PT65 experiment			
Main mode	Train	Train	Train
Access mode	Walk, bicycle	Walk, bicycle	Bus, tram, metro
Access time	5, 10, 15 min	Same	Same
Bike parking costs	0, 1 euro	0, 1 euro	
Wait for train	5 min	5 min	5, 10, 15 min
Main travel time	45, 55, 65 min	Same	Same
Transfer time	0, 10, 20 min	Same	Same
Wait for PT egress		5, 10, 15 min	
Egress mode	Walk, PTbike	Bus, tram, metro	Walk, PTbike
Possibility of delay	0, 15, 30 min	Same	Same
Travel costs access			1.50 euro (bus/tram/metro)
Travel costs main	0, 5, 10 euro	Same	Same
Travel costs egress	3,00 euro (PTbike)	1.50 euro (bus/tram/metro)	3,00 euro (PTbike)
Facilities at stations	No, some, many	Same	Same

Source : Arentze et Molin (2013)

FIGURE 1.2 – Choix multimodal pour un déplacement de plus de 65 km avec le train pour mode principal

Green et Srinivasan (1990), ainsi que Pearmain et al. (1991) font référence à un nombre maximum de six à sept attributs par question de choix sans pour autant justifier ce nombre. Sans citer ses sources, Pons (2011) rapporte que des « études réalisées dans la sphère de la psychologie mettent en évidence qu'à partir de quatre dimensions simultanées, le cerveau humain évince naturellement de son choix les éléments auxquels il accorde le moins d'importance ». Une analyse dédiée est donc nécessaire pour mieux comprendre le rôle du nombre d'attributs sur les choix. Louviere et al. (2008) ont comparé 44 enquêtes de choix expérimentaux dont ils ont contrôlé les différentes composantes, dont le nombre d'attributs (4, 8, 12, 16). Leur analyse suggère que les choix deviennent moins cohérents lorsque le nombre d'attributs augmente, ce qui inviterait à choisir les catégories comportant le moins d'attributs. Ils ne précisent pas s'il faut s'arrêter à quatre attributs ou si huit reste convenable. RFF (2013) a testé la complexité des étapes de choix lors de focus groups. Les choix avec cinq attributs ont été jugés difficiles, ce qui a conduit à proposer dans l'enquête finale une introduction progressive des attributs : pour un même exercice, ce sont d'abord quatre puis cinq attributs entre lesquels les personnes interrogées

doivent arbitrer.

1.3.3.6 Conclusion

De l'ensemble de ces travaux, la conclusion suivante peut être retirée. Les attributs choisis doivent décrire le bien ou la situation évalué(e) de manière pertinente et complète, de façon à mimer des décisions réelles, quitte à ce que certains attributs soient ignorés par certains individus. Du fait de l'hétérogénéité des processus de choix, il est préférable de ne pas omettre d'attributs potentiellement importants pour une part de la population. La préoccupation inverse, l'impact de la présence d'attributs non importants sur les choix, ne semble pas être justifiée. Il est donc préférable d'introduire des attributs qui seront potentiellement ignorés plutôt que de ne pas les introduire et prendre le risque que les répondants infèrent une valeur non connue. Cette conclusion est vraie jusqu'à un certain point au-delà duquel la charge cognitive peut effectivement devenir un problème. Cependant ce point (5, 6, 7 attributs ?) est certainement supérieur à ce qui est parfois observé dans la littérature avec des questions de choix ne faisant intervenir que deux ou trois attributs (Hess, 2014). En définitive, il n'y a pas de réponse unique à la question : quel est le bon nombre d'attributs à intégrer dans les questions de choix ? Les recherches empiriques semblent montrer que la réponse est à adapter selon le nombre de niveaux et leur étendue, le nombre d'alternatives et certainement aussi la nature des attributs (plus ou moins complexes à appréhender), le nombre de questions de choix, la longueur globale du questionnaire et le mode d'enquête.

1.3.4 Analyse qualitative

L'analyse qualitative est une étape importante du choix des attributs et de leur niveau, et plus généralement, de la réalisation du questionnaire. Les perceptions et connaissances de la population vis-à-vis des différents modes de transport, ainsi que les déterminants du choix modal, ont largement été abordés dans la littérature. Les entretiens et focus groups complètent donc un support bibliographique déjà existant. Ils permettent d'adapter les enseignements de la littérature :

- Au territoire étudié (Région Rhône-Alpes) et à ses habitants ;
- A la méthodologie employée (EPD) et à la décomposition du système en attributs ;
- Aux objectifs de l'étude, centrés sur le transport ferroviaire et les déplacements à l'échelle régionale.

Deux méthodes peuvent être utilisées, de façon complémentaire. Premièrement, les entretiens

individuels permettent de combiner à la fois une approche libre, inductive, permettant de faire parler librement la personne interrogée, une approche déductive par l'utilisation d'un guide d'entretien et une approche plus quantitative cadrée par des questions précises qui consistent à faire classer ou évaluer des assertions ou des objets par les personnes interrogées. Timmermans et al. (1982) listent en particulier trois techniques utilisées pour choisir des attributs :

- Le positionnement multidimensionnel est un outil d'analyse des préférences qui passe par la représentation graphique d'un ensemble de similarités ou de dissimilarités et par la construction d'échelles communes à un ensemble d'attributs subjectifs. Similarités et dissimilarités sont obtenues en demandant directement aux sujets d'estimer ou de classer des ressemblances ou dissemblances entre paires d'objets ;
- L'utilisation d'échelles de Likert qui sont des échelles de mesure sur lesquelles les répondants peuvent exprimer l'importance relative de chacun des attributs qui leur sont présentés ;
- La grille d'interprétation de Kelly, utilisée par exemple par Timmermans et al., 1982, qui consiste à sélectionner différents items et à les présenter trois par trois à la personne interrogée (voir figure B.1, annexe B, p. 298). Cette technique permet de travailler sur la construction mentale et les perceptions associées à chacun des items sélectionnés sans avoir à définir une liste *a priori*.

Deuxièmement, la méthode des focus groups, ou groupe de discussion, utilisée à l'origine pour des recherches en marketing, s'applique de plus en plus dans le cadre d'études basées sur les préférences déclarées mobilisant une phase d'analyse qualitative préparatoire (Louviere et al., 2010 ; Davies et Laing, 2002). Un focus group est une discussion structurée en plusieurs phases autour d'un script prédéfini au sein d'un petit groupe de citoyens répondants à des critères précis et animée par un modérateur neutre. La discussion interactive du groupe dans un cadre non contraignant et détendu permet d'obtenir de nombreuses informations à moindre coût sur les préférences et valeurs des personnes concernant un sujet défini. Il s'agit là de l'essence même de ce processus participatif (et de sa valeur ajoutée par rapport à des entretiens individuels), où les participants sont amenés à se situer et à réagir par rapport aux opinions et aux affirmations des autres.

La plus-value des focus groups repose sur quatre actions :

- prendre connaissance et faire émerger la diversité des vues et opinions sur un sujet, les confronter et observer leur évolution éventuelle suite à l'ajout d'informations nouvelles ;
- comprendre le raisonnement qui sous-tend une opinion, objectiver les facteurs associés aux pratiques ;

- donner aux participants la possibilité d'exposer et d'expliquer leurs demandes et leurs attentes ;
- déterminer le degré de consensus existant sur un sujet donné.

Les focus groups ne sont donc pas uniquement un préalable aux questionnaires, ils constituent un matériau qualitatif utile à l'interprétation des résultats quantitatifs. Ils permettent de rattacher les faits à leur contexte, de les faire parler, de dégager des impressions, des tendances, des hypothèses.

1.4 Design d'une enquête de préférences déclarées

Le design, également appelé plan factoriel, désigne la façon de combiner les attributs et leurs niveaux dans les questions de choix.

1.4.1 Plan factoriel complet

1.4.1.1 Un plan factoriel idéal ...

Un plan factoriel complet (*full design*) contient l'ensemble des alternatives possibles. Par exemple, si l'on considère quatre attributs à trois niveaux, alors il y a 64 alternatives possibles. Un plan factoriel complet a en particulier la propriété d'être orthogonal : tous les attributs présentés varient indépendamment les uns des autres. Il permet également d'estimer les effets principaux (effets des attributs sur l'utilité), mais aussi les effets secondaires qui correspondent à l'influence simultanée de deux attributs, voire des effets d'ordre supérieur.

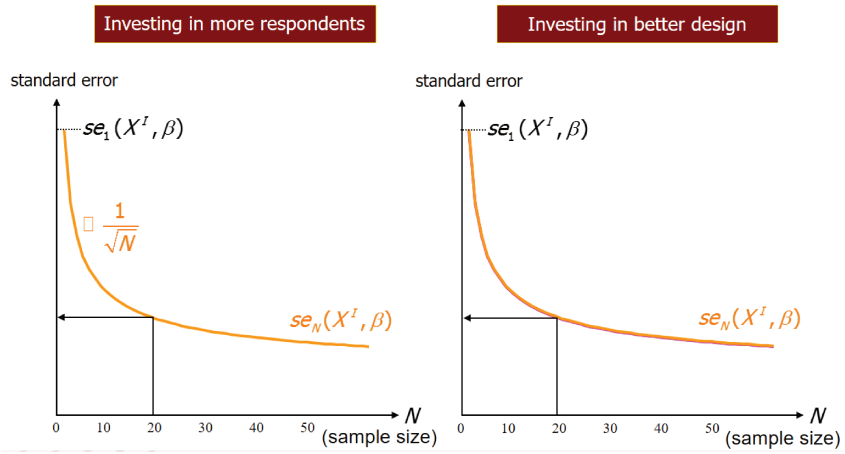
1.4.1.2 ...peu applicable en pratique

Un plan factoriel complet est rarement applicable en pratique. Le nombre de combinaisons possibles s'élevant souvent à une, voire plusieurs, centaine(s). Lorsque le nombre d'alternatives existantes est trop important, alors deux solutions sont envisageables :

- Tirer aléatoirement les alternatives selon le questionnaire et les combiner entre elles de façon également aléatoire. Cette solution est parfois mentionnée dans la littérature (par exemple, Walker et al., 2016) mais rarement mise en œuvre.
- Réaliser un plan factoriel partiel, i.e. une partition des alternatives possibles résultant du plan factoriel complet. Les alternatives présentées aux personnes interrogées suivent alors une structure, répondent à des règles pré-établies.

Le plan factoriel partiel (*fractional design*) doit répondre à certains critères :

- Permettre l'identification des paramètres. Cela signifie en particulier qu'il faut veiller à ce que le design choisi permette d'estimer les effets souhaités. Si l'on sait d'ores et déjà que l'on souhaite intégrer des variables croisées dans le modèle (par exemple, temps de trajet×confort), alors il est nécessaire d'adopter un design qui permette d'estimer des effets d'ordre deux.
- Être le plus efficace possible : augmenter l'efficacité du design permet d'augmenter la précision des estimateurs (diminuer l'écart-type et augmenter la t-value) ou, de façon équivalente, permet de diminuer la taille de l'échantillon tout en gardant la même précision des estimations (voir figure 1.3).



Source : Rose et Bliemer (2013)

FIGURE 1.3 – Sources d'efficacité de l'enquête

Pour réaliser un plan factoriel partiel, deux méthodes sont possibles (Scarpa et Rose, 2008). La première, nommée design orthogonal (*orthogonal design*), se base sur la propriété d'orthogonalité des différentes alternatives alors que la seconde, nommée design efficient (*efficient design*), chercher à tirer partie des informations préalables (*priors*) disponibles pour maximiser l'efficacité du design.

1.4.2 Design orthogonal

Un design orthogonal est un design dans lequel il n'y a pas de corrélation entre les attributs, les attributs variant indépendamment les uns des autres. Une fois les alternatives choisies, il faut les combiner entre elles afin de construire les questions de choix. Différentes méthodes sont possibles et décrites dans Pons (2011), un résumé est présenté dans la liste ci-dessous. Le critère d'orthogonalité est également important lors de cette étape. « Lors de la construction des trade-off, le principe permettant de réduire ou d'éliminer l'orthogonalité consiste, au fil du design, à maintenir un écart régulier

lorsque l'on compare les niveaux de mêmes attributs au sein de différents alternatives composant un même trade-off. » (Pons, 2011, p.41).

- Combiner les alternatives aléatoirement, ce qui permettrait de toutes les représenter (ou du moins le plus grand nombre). Cette méthode est peu ou pas utilisée ;
- Faire un tirage aléatoire contrôle (Louviere et al., 2000). Dénommée *mix and match* ou *foldover*, cette méthode consiste à construire des plans factoriels avec des alternatives différentes puis à les combiner aléatoirement ;
- Présenter chaque alternative conjointement avec une alternative opposée (dans le cas d'attributs à deux niveaux) ou le niveau suivant (dans le cas d'attributs à trois niveaux). Cette méthode, appelée *shifting*, ne permet d'estimer que les effets principaux ;
- Génération simultanée ou $L^M A$ *design*, avec L le nombre de niveaux, M le nombre d'attributs et A le nombre d'alternatives dans chaque question de choix. Il faut alors générer un plan factoriel de $M \times A$ attributs à N niveaux. Les $M \times A$ colonnes sont alors réparties entre A groupes de M colonnes chacun. « Cette méthode permet d'estimer les effets principaux, de dissocier les effets d'interactions s'ils le sont dans le plan factoriel de départ et de maintenir l'orthogonalité d'un même attribut entre les alternatives ce qui n'était pas le cas des trois méthodologies précédentes » (Pons, 2011, p. 45).

Un design orthogonal est optimal en termes d'efficacité pour les modèles linéaires (Rose et al., 2009). Cette propriété ne vaut pas pour les modèles non-linéaires et donc *a fortiori* pour les modèles de choix discret. Dans la pratique, l'application de designs orthogonaux reste fréquente, même pour les modèles de choix discret, du fait de leur facilité de mise en œuvre et des bons résultats qu'ils permettent d'obtenir sans pour autant être optimaux.

1.4.3 Design efficient

1.4.3.1 Principe

Dans l'optique d'utiliser les données dans des modèles de choix discret, la méthode théoriquement la plus adaptée est d'utiliser un design efficient. Un design efficient est un design qui minimise les éléments de la matrice de variance-covariance asymptotique (l'opposé de l'inverse de la matrice d'information de Fisher), permettant ainsi d'obtenir des estimations de paramètres plus fiables pour un même nombre d'observations (Rose et al., 2008). Un design efficient cherche à tirer partie des informations préalablement (*priors*) disponibles pour maximiser l'efficience du design. L'idée

sous-jacente est que chaque alternative ait une probabilité proche d'être choisie, ce qui permet ainsi de maximiser l'information issue de chaque question de choix. Construire un design efficient nécessite donc de se baser sur des valeurs *a priori* pour les paramètres. Ces *a priori* peuvent être le résultat d'une revue de la littérature ou de pilotes du questionnaire. Par ailleurs, il est nécessaire de connaître à l'avance la spécification du modèle. En effet, construire un design sur la base d'un modèle et l'estimation finale d'un autre modèle conduit à une perte d'efficacité (Bliemer et al., 2009).

Il existe de nombreuses mesures de l'efficience d'un design, dont les deux les plus utilisées, la *D-error* et la *C-error*, sont détaillées ci-dessous. Les autres mesures sont brièvement abordées.

1.4.3.2 D-error

La *D-error* est la mesure de l'efficience la plus utilisée dans la littérature. Elle est calculée à partir du déterminant de la matrice de variance-covariance asymptotique (voir équation (1.1)) qui est l'opposé de l'inverse de la matrice d'information de Fisher $\mathbf{I_N}$ (voir équation (1.2)), elle-même égale à la dérivée seconde de la fonction de log-vraisemblance $\mathcal{L}_N$ (Rose et al., 2009 ; Scarpa et Rose, 2008).

Notons $\mathbf{\Omega_N}$ la matrice $(K \times K)$ de variance-covariance asymptotique, avec K le nombre de paramètres du modèle et N le nombre de répondants. Ceux-ci font face à Q questions de choix ayant J alternatives chacune. $\mathbf{\Omega_N}$ dépend de $\mathbf{X} = [x_n]$ $(n = 1, \cdots, N)$, les valeurs prises par les variables explicatives issues du design, $\tilde{\boldsymbol{\beta}}$ les valeurs *a priori* des paramètres $\boldsymbol{\beta}$ associés aux variables explicatives et $\mathbf{C} = [c_{nqj}]$ $(n = 1, \cdots, N;\ q = 1, \cdots, Q;\ j = 1, \cdots, J)$, les variables de choix valant 1 si le répondant n choisit l'alternative j à la question q et 0 sinon, la probabilité π_{nqj} que la personne n choisisse l'alternative j à la question de choix q. L'écriture de cette probabilité dépend du modèle choisi. Un facteur d'échelle $1/K$ est appliqué pour prendre en compte le nombre de paramètres du modèle. La *D-error* est

$$D - error = \det(\mathbf{\Omega_N}(\mathbf{X}, \tilde{\boldsymbol{\beta}}))^{\frac{1}{K}} \tag{1.1}$$

avec

$$\mathbf{\Omega_N}(\mathbf{X}, \tilde{\beta}) = -[\mathbb{E}(\mathbf{I_N}(\mathbf{X}, \mathbf{C}, \tilde{\boldsymbol{\beta}})]^{-1} = -[\frac{\partial^2 \mathcal{L}_N(\mathbf{X}, \mathbf{C}; \tilde{\boldsymbol{\beta}})}{\partial \beta \partial \beta'}] \tag{1.2}$$

et

$$\mathcal{L}_N(\mathbf{X}, \mathbf{C}, \tilde{\boldsymbol{\beta}}) = \sum_{n=1}^{N} \sum_{q=1}^{Q} \sum_{j=1}^{J} c_{nqj} \ln \pi_{nqj}(\mathbf{X}; \tilde{\boldsymbol{\beta}}). \tag{1.3}$$

Le design qui minimise la *D-error* est dit *D-optimal*. Il est cependant difficile d'atteindre un tel design. Un design *D-efficient* est donc souvent choisi : il minimise la *D-error* sur le nombre d'itérations choisi

sans pour autant garantir l'optimalité dans l'absolu [9]. Le calcul de la *D-error* peut varier selon les *a priori* disponibles sur les paramètres. Trois cas de figure se distinguent :

1. La première s'applique lorsqu'aucun a priori n'est disponible sur les paramètres, alors fixés à zéro. Minimiser cette erreur, la $D_z - error$,

$$D_z - error = \det(\mathbf{\Omega_N}(\mathbf{X}, \mathbf{0}))^{\frac{1}{K}}, \tag{1.4}$$

   revient à approcher un design orthogonal.

2. La deuxième, la $D_p - error$,

$$D_p - error = \det(\mathbf{\Omega_N}(\mathbf{X}, \tilde{\boldsymbol{\beta}}))^{\frac{1}{K}}, \tag{1.5}$$

   s'applique lorsque le design est construit pour une valeur donnée des paramètres. L'optimalité se comprend alors uniquement au niveau local puisqu'il n'existe pas de design efficient pour toutes les valeurs des paramètres.

3. La troisième, la $D_b - error$,

$$D_b - error = \int_{\tilde{\boldsymbol{\beta}}} \det(\mathbf{\Omega_N}(\mathbf{X}, \tilde{\boldsymbol{\beta}}))^{\frac{1}{K}} \phi(\tilde{\boldsymbol{\beta}}|\boldsymbol{\theta}) d\tilde{\beta}, \tag{1.6}$$

   est un design efficient bayésien [10] qui permet de prendre en compte l'incertitude des estimations sur les paramètres. Au lieu de se baser sur la valeur des coefficients, le design se base sur la distribution de ces coefficients selon une loi de densité ϕ (par exemple, loi normale ou loi uniforme) dont les paramètres sont compris dans le vecteur $\boldsymbol{\theta}$.

1.4.3.3 C-error

Dans une EPD, le résultat le plus important est la traduction monétaire, ou temporelle, des différents attributs évalués. En particulier, le consentement à payer (*Willingness To Pay* ; WTP) pour obtenir une unité supplémentaire d'un attribut k s'écrit comme le ratio du coefficient β_k correspondant à l'attribut

9. Différents algorithmes peuvent être utilisées (voir Rose et al., 2009). Avec le logiciel NGENE, à chaque itération, un design est créé et évalué. S'il a une *D-error* plus faible que le design précédent, alors il est stocké.

10. C'est ce type de design, avec une distribution uniforme des paramètres, qui a été retenu dans l'EPD à l'échelle de la Région Rhône-Alpes.

évalué et β_{cost} le coefficient de l'attribut coût

$$\text{WTP}_k = -\frac{\beta_k}{\beta_{cost}},\ \forall k = 1, \ldots, K. \tag{1.7}$$

Un consentement à payer est donc une fonction non linéaire de deux coefficients. Le calcul de sa variance $\mathbb{V}$ n'est pas trivial et le développement d'un nouvel indicateur d'efficience est nécessaire. Utilisant la méthode delta, Scarpa et Rose (2008) proposent la *C-error*

$$\mathbb{V}\left(\frac{-\beta_k}{\beta_{cost}}\right) \quad \cong \quad \beta_k^{-2}[\mathbb{V}(\beta_k) - 2\beta_k\beta_{cost}^{-1}Cov(\beta_k, \beta_{cost}) + (\beta_k/\beta_{cost})^2\mathbb{V}(\beta_{cost})],\ \forall k = 1, \ldots, K, \tag{1.8}$$

qui minimise la variance approximée du consentement à payer mais reste encore peu utilisée.

1.4.3.4 Autres mesures

De nombreuses autres mesures peuvent être utilisées. Citons en particulier la *A-error*

$$A - error = \text{tr}(\boldsymbol{\Omega_N}(\mathbf{X}, \tilde{\boldsymbol{\beta}})) \tag{1.9}$$

qui s'appuie sur la trace de la matrice de variance-covariance asymptotique. Contrairement à la *D-error*, elle ne minimise que les variances (et non les variances et les co-variances) des paramètres puisque par définition la trace d'une matrice ne considère que ses éléments diagonaux. La *S-error* permet quant à elle de déterminer quelle taille de l'échantillon est nécessaire pour obtenir des valeurs significatives pour les coefficients.

1.4.3.5 Design efficient et méthode pivot

Dans les EPD appliquées au transport, la personnalisation des niveaux d'attributs est une pratique courante. Elle permet de minimiser le biais hypothétique et d'augmenter la qualité des réponses. Une alternative de référence, celle décrivant la pratique de transport actuelle, est proposée et les attributs des autres alternatives pivotent autour du niveau de référence de chacun des attributs.

Malgré l'intérêt certain d'utiliser des valeurs de référence dans la génération du design, l'utilisation de telles valeurs « est incompatible avec les méthodes actuelles de génération de designs efficients » (Rose et al., 2008, p.396). En effet, selon la situation de référence de chaque personne, les niveaux d'attributs diffèrent et le design doit donc être adapté à cette situation unique. Rose et al. (2008)

proposent donc quatre solutions qu'ils testent sur des données de choix d'itinéraire :

1. Créer le design sur les caractéristiques de la population moyenne ;
2. Créer le design en segmentant la population ;
3. Créer le design une fois l'ensemble des données sur la situation de référence recueillies ;
4. Créer en direct un design adapté à la situation de chaque personne.

Leur comparaison porte sur ces quatre designs ainsi que sur un design orthogonal. Ils concluent que le design orthogonal est significativement moins performant que les quatre designs efficients. Parmi les designs efficients, ceux basés sur des données réelles (3 et 4) semblent meilleurs en termes de précision des résultats. Ils sont cependant longs à mettre en œuvre puisque nécessitant deux étapes (recueil de données puis design) pouvant également générer une perte de l'échantillon [11].

1.4.3.6 Une forte dépendance de l'efficience aux *a priori*

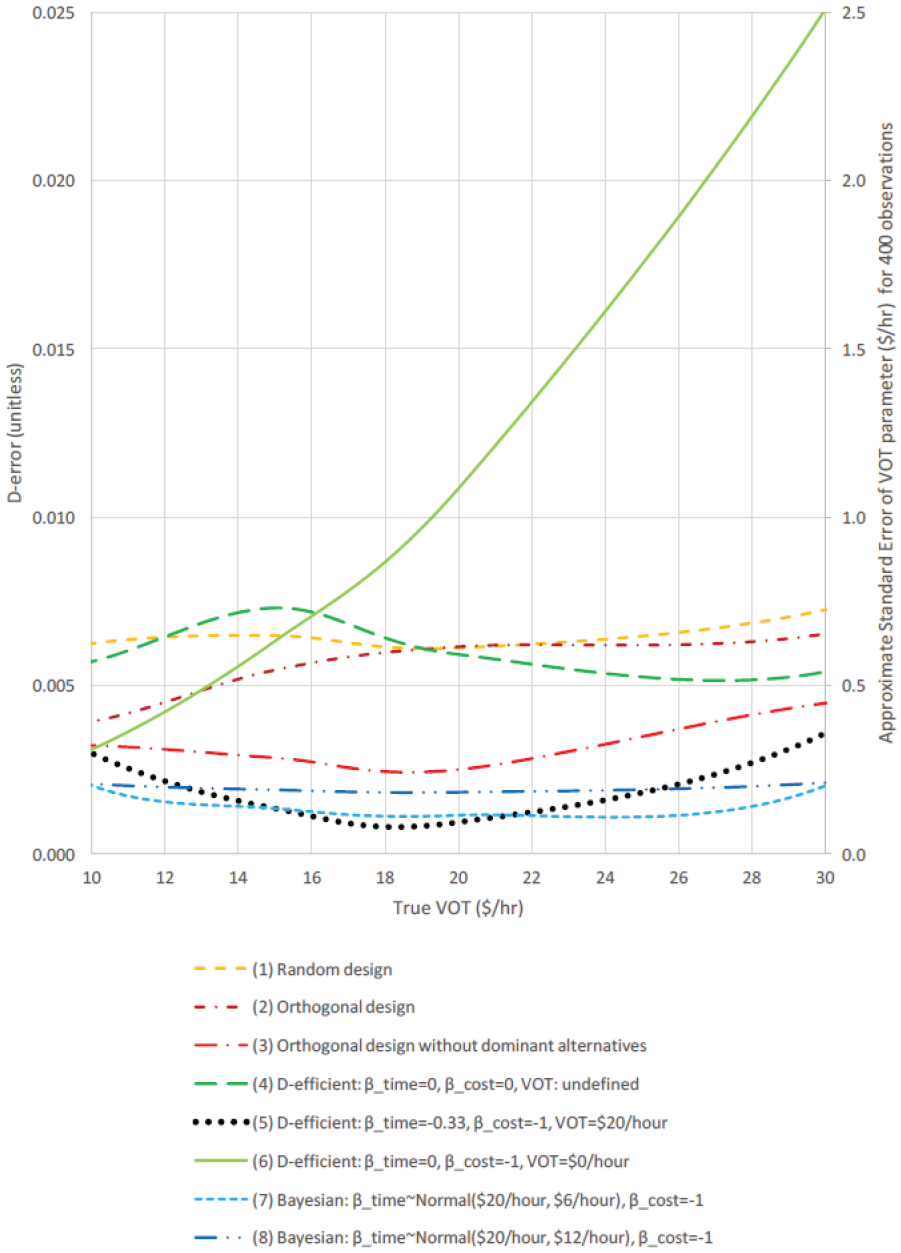
Walker et al. (2016) réalisent une revue de littérature sur les designs utilisés dans les EPD et leurs performances relatives. Plus l'information sur les *a priori* est précise, meilleur est le design. Une information très précise peut être apportée par exemple par l'utilisation d'un design efficient itératif (Bliemer et Rose, 2010), avec mise à jour des *a priori* au fur et à mesure de l'avancement de l'enquête. Le design itératif offre d'aussi bonnes performances que le design efficient basé sur les vraies valeurs des paramètres et les deux sont supérieurs au design orthogonal. L'avantage du design itératif est qu'il n'est pas sensible à une mauvaise spécification des *a priori*. Il nécessite cependant un lourd dispositif d'enquête, coûteux en temps et en calculs, peu opérationnel.

Puisque les designs efficients sont fortement dépendants des valeurs *a priori*, que se passe-t-il si ces valeurs ne reflètent pas le vrai paramètre ? Walker et al. (2016) répondent à cette question à l'aide de simulations. Ils génèrent 200 questions de choix à partir de trois types de design : orthogonal, *D-efficient* avec *a priori* fixes et *D-efficient* bayésien. Pour chaque type de design, différentes variantes sont déclinées. Par exemple, pour le design efficient bayésien, les *a priori* sont distribués selon une loi normale. L'analyse porte plus particulièrement sur la valeur du temps, dont la valeur *a priori* est fixée à 20 $/heure. Les différents designs sont comparés pour 15 « vraies » valeurs du temps, allant de 0 à 100 $/heure. L'efficacité est mesurée au travers du calcul analytique de la *D-error* et d'une microsimulation de l'écart-type.

La figure 1.4 montre les performances des différents designs sur la base du calcul de la *D-error*, avec

11. C'est donc la deuxième solution, avec segmentation de la population, qui a été choisie pour l'EPD à l'échelle de la Région Rhône-Alpes.

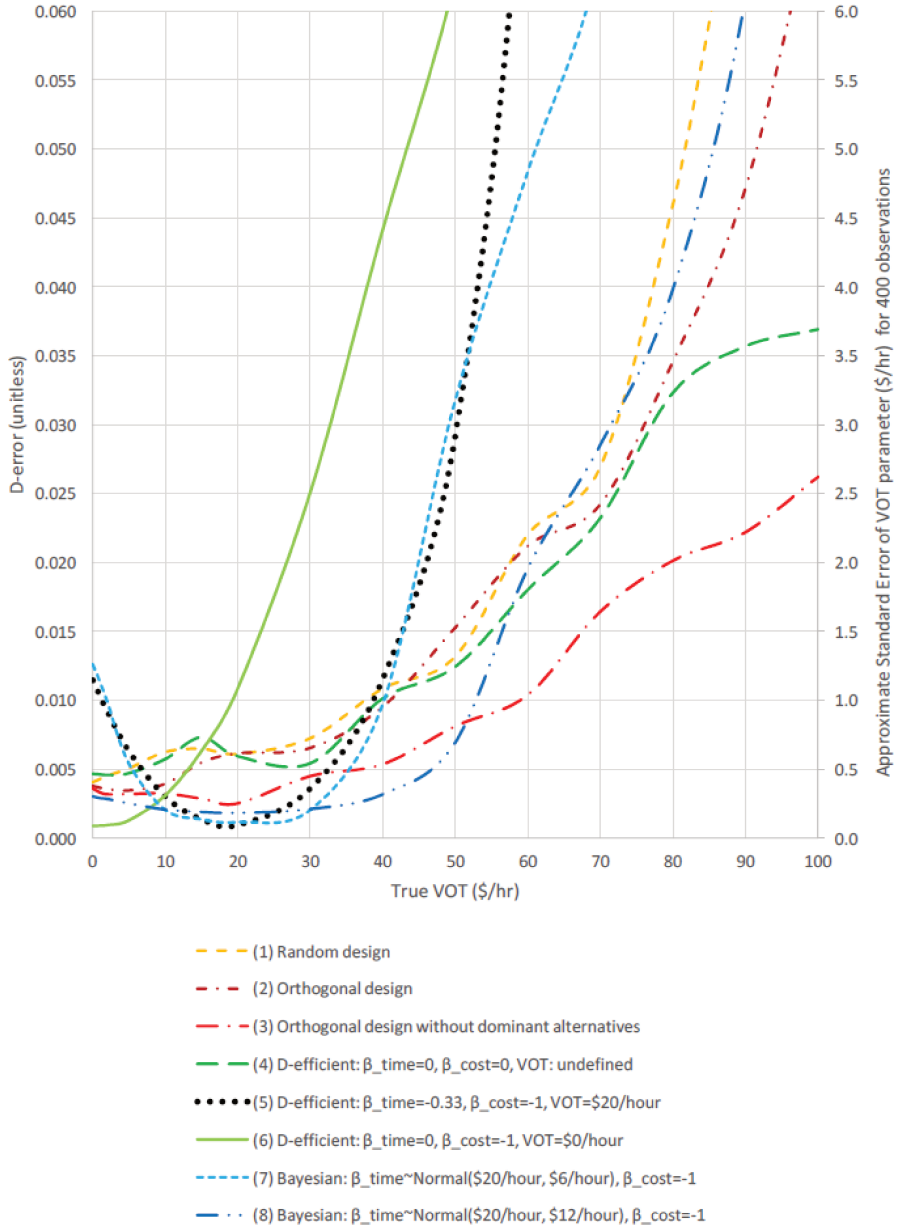
un zoom sur les « vraies » valeurs du temps comprises entre 10 et 30 $/heure.



Source : Walker et al. (2016)

FIGURE 1.4 – Robustesse des résultats avec zoom sur des valeurs du temps proches de l'*a priori*

Les trois designs efficients basés sur l'*a priori* d'une valeur du temps de 20 $/heure sont ceux qui offrent les meilleurs performances. Le design orthogonal avec suppression des alternatives dominantes offre également des bons résultats, ce qui montre que le fait de supprimer ces alternatives permet des gains d'efficacité. Les résultats sont différents si l'ensemble des « vraies » valeurs du temps sont considérées (voir figure 1.5).



Source : Walker et al. (2016)

FIGURE 1.5 – Robustesse des résultats sur l'ensemble des valeurs du temps

En-deçà de 50 $/heure, le design bayésien avec un écart-type élevé offre les meilleurs résultats. Au delà de cette valeur, le design le plus robuste pour de fortes valeurs du temps est le design orthogonal sans alternative dominante. En effet, ce design permet de couvrir une large étendue de valeurs du temps sans gaspiller d'observations sur des choix dominés. Tout type de design gagnerait à bénéficier d'un tel « nettoyage ». Par ailleurs, des designs à deux étapes qui utilisent un petit échantillon pour générer les *a priori* ne sont pas plus robustes.

En résumé, ces travaux montrent que si le modélisateur a confiance dans les *a priori* pour les paramètres du modèle, alors il doit choisir un design efficient en privilégiant un design efficient bayésien (c'est le choix qui a été effectué pour l'EPD de la thèse). Dans le cas contraire, il est préférable d'utiliser un design orthogonal en supprimant les alternatives dominantes.

1.5 Définition des attributs

De nombreuses EPD ont été réalisées pour analyser le choix modal. Certaines d'entre elles intègrent le train comme l'un des modes de transport, aux côtés par exemple de la voiture (Pons, 2011), de la voiture et du bus (Richter et Keuchel, 2012), de la voiture, du bus/tram, du train, du vélo et de la marche à pied (par exemple, Arentze et Molin, 2013). D'autres EPD ne proposent que le train comme mode transport (par exemple, RFF, 2013 ; Börjesson et Eliasson, 2011 ; Börjesson et al., 2012). Le choix est donc laissé entre plusieurs trains se différenciant par leurs caractéristiques.

L'intégration du train dans une EPD nécessite l'inclusion d'attributs spécifiques pour le décrire. Suite à la revue de la littérature, à l'analyse qualitative (voir résultats en section 2.1) et à une définition précise des besoins opérationnels pour le modèle multi-modal régional, six attributs ont été retenus : le temps de déplacement (de l'origine à la destination), le coût de déplacement (de l'origine à la destination), la fiabilité, le cadencement, la fréquence de passage et le confort, c'est-à-dire la possibilité ou non d'avoir une place assise. Le cadencement, combiné à la fréquence, ainsi que la fiabilité posent des questions méthodologiques dans leur définition et leur intégration dans l'enquête. Sans être exhaustives de la littérature, les parties suivantes donnent donc des indications sur la définition du cadencement, de la fréquence et de la fiabilité dans des EPD. La question du confort est plus spécifiquement abordée dans les chapitres 5 et 8.

1.5.1 Cadencement et fréquence

1.5.1.1 Définition du cadencement

Le schéma régional des services de transport définit un réseau de transport cadencé sur les critères suivants :

- Les trains ou cars sont regroupés en « familles » (intercités, périurbains, etc.) aux caractéristiques simples et régulières : mêmes itinéraires, mêmes politiques d'arrêt, mêmes temps de parcours, et ce dans les deux sens.

- Les trains ou les cars circulent à intervalles réguliers, l'intervalle étant appelé « cadence » avec des minutes de départ identiques, un fonctionnement similaire en heures creuses et en heures de pointe, des plages de fonctionnement (amplitude du service) étendues.

En Région Rhône-Alpes, le cadencement a été mis en place en décembre 2007. Il se traduit par une offre plus lisible puisque les trains partent à la même minute à une fréquence régulière, facilitant ainsi la mémorisation des horaires, ainsi qu'une offre plus riche et mieux coordonnée puisque le cadencement a permis d'augmenter le nombre de trains en circulation et de développer les correspondances entre trains, multipliant ainsi les destinations possibles.

1.5.1.2 Prise en compte dans les EPD

Si l'évaluation économique du cadencement et son intégration dans des modèles de demande est en développement, celui-ci reste rarement intégré comme un attribut du choix modal. Deux références donnent cependant des pistes. Toutes deux analysent le cadencement conjointement avec la fréquence de passage des trains.

Premièrement, à l'aide de focus groups, Pons (2011) confirme que le cadencement est un concept particulièrement difficile à appréhender et qu'il convient dès lors de le décomposer en différents attributs à articuler ensuite dans l'évaluation. Les trois composantes qu'il propose sont la fréquence de passage des trains, la régularité et la lisibilité des horaires. La fréquence est exprimée en laps de temps entre deux passages de trains : « les trains passent tous les quart d'heure ». Pons (2011) relève également dans la littérature que la fréquence est l'élément le plus déterminant dans la fréquentation des trains. FitzRoy et Smith (1995) recommandent notamment de systématiquement intégrer la fréquence comme variable explicative du choix modal lorsque le train fait partie des alternatives de choix. La régularité et la lisibilité des horaires sont introduites sous la forme d'une variable binaire : « les trains passent à horaires fixes/les trains passent à horaires variables ». Auparavant, un texte explicatif permet de définir les horaires « fixes » (cadencés) et les horaires « variables » (non cadencés) et de les visualiser avec une grille horaire type (voir figure 1.6a).

Deuxièmement, Wardman et al. (2004) introduisent le cadencement sous la forme d'une variable composite comprenant la fréquence horaire de passage des trains (4, 2 ou 1), la régularité de la durée entre deux passages, le passage à minute fixe et la mémorabilité des horaires (voir figure 1.7a). A partir de ces variables, ils construisent 13 alternatives (plus une qui correspond à la situation actuelle) qu'ils distribuent entre deux versions du questionnaire. Les résultats des estimations sont ensuite utilisés pour prévoir l'évolution de la demande en fonction de ces différents paramètres.

* Un train peut partir d'une gare :
- à horaires fixes (par exemple, 9h20, 9h50, 10h20, 10h50, etc.) ;
- à horaires variables (par exemple, 9h07, 9h36, 10h14, 10h44, etc.).

Vous pouvez voir ci-dessous deux grilles horaires différentes. A gauche, les trains passent à horaires fixes. A droite, ils passent à horaires variables. Dans les deux cas, le nombre de trains par heure ne change pas.

Les trains passent à horaires fixes.			
6h05	6h20	6h35	6h50
7h05	7h20	7h35	7h50
8h05	8h20	8h35	8h50
	9h20		9h50
	10h20		10h50
11h05	11h20	11h35	11h50
12h05	12h20	12h35	12h50

Les trains passent à horaires variables.				
6h03	6h18	6h33	6h48	
7h05	7h10	7h22		7h55
	8h13	8h28	8h39	8h48
9h07		9h36		
	10h14		10h44	
	11h12	11h28	11h39	11h53
12h03	12h18	12h33	12h48	

De plus, lorsque les horaires sont fixes, la grille horaire est identique tous les jours et le temps de parcours entre deux gares est toujours le même.

(a) Pons (2011)

FIGURE 1.6 – Scénarios de grilles horaires utilisées dans les exercices de choix

Timetable scenarios used in SP exercise

Scenario	Timetable	Clock	Mem	Design
TT1	4 per hour, even interval, 00 15 30 45	Yes	Yes	1
TT2	4 per hour, uneven interval, 00 05 30 35	Yes	Maybe	1
TT3	4 per hour, even interval, 08 23 38 53	Yes	No	2
TT4	4 per hour, uneven interval, 08 16 40 51	Yes	No	2
TT5	4 per hour, varies across day	No	No	2
TT6	2 per hour, even interval, 00 30	Yes	Yes	1
TT7	2 per hour, uneven interval, 00 10	Yes	Maybe	2
TT8	2 per hour, even interval, 08 38	Yes	No	1
TT9	2 per hour, uneven interval, 08 23	Yes	No	2
TT10	2 per hour, varies across day	No	No	1
TT11	1 per hour, 08	Yes	No	Both
TT12	1 per hour, 00	Yes	Yes	Both
TT13	1 per hour, varies across day	No	No	Both

(a) Wardman et al. (2004)

Le tableau 1.2 synthétise les valeurs obtenues à partir de ces deux études.

TABLEAU 1.2 – Valeurs estimées des attributs en lien avec le cadencement et la fréquence

Source	**Désignation de l'attribut**	**Equivalent temps ou équivalent coût**
Pons (2011)	Fréquence de passage des trains	2,26 à 2,63 €/train supplémentaire ; 5,75 à 5,60 min de trajet/train/train supplémentaire
	Lisibilité des horaires (régularité des intervalles + passages à minutes fixes)	2,55 à 3,75 € pour des minutes fixes 5,60 à 5,75 min de trajet pour des minutes fixes
Wardman et al. (2004)	Régularité des intervalles	5,62 à 11,85 min de trajet pour des intervalles réguliers (selon le motif et la fréquence horaire)
	Passages à minutes fixes	3,59 à 14,25 min de trajet pour des minutes fixes
	Horaires mémorisables	2,40 à 8,02 min de trajet pour des horaires mémorisables

Notes : Les fourchettes de valeur sont imputables au motif de déplacement (loisir ou professionnel), ainsi qu'à la fréquence de passage des trains : la régularité des intervalles n'a pas la même importance selon qu'il passe deux ou quatre trains par heure. min=minutes.

1.5.2 Fiabilité

1.5.2.1 Définition

La fiabilité, ou la régularité, se définit comme l'absence de variations imprévisibles du service offert. Un service fiable repose donc sur l'adéquation entre les caractéristiques affichées du service (horaire de départ, temps de trajet, horaire d'arrivée) et le service rendu *in fine*. En anglais, les mots employés pour désigner la fiabilité sont *regularity*, *(travel time) reliability*, *unreliability*, *variability* et *punctuality*. Les termes *regularity* et *punctuality* ont un usage limité aux transports en commun, *regularity* étant plus pertinent pour les transports définis par une fréquence et *punctuality* pour ceux définis par un horaire (RFF, 2013).

Deux paramètres permettent de définir la régularité : le pourcentage de trains ne respectant pas les caractéristiques affichées (horaire de départ, temps de trajet, horaire d'arrivée) et l'ampleur du décalage entre affichage et réalité (nombre de minutes d'avance ou de retard à l'arrivée). Le cas le plus fréquent et le plus impactant est l'arrivée tardive du train, soit du fait d'un départ trop tardif, soit du fait d'un temps de parcours plus long que prévu.

1.5.2.2 Cadre théorique

D'un point de vue économique, deux approches de la fiabilité sont possibles, l'une extensive, basée sur l'horaire, et l'autre de forme réduite, basée sur le temps de trajet et sa variation (Li et al., 2010). Si les deux types de modèles sont théoriquement équivalents, Börjesson et al. (2012) montrent que, empiriquement, les résultats sont différents.

1.5.2.2.1 Approche extensive : modélisation basée sur les horaires Dans la forme extensive, appelée modélisation basée sur les horaires (*scheduling approach*), l'utilité intègre explicitement l'heure d'arrivée préférée (*Preferred Arrival Time* ; PAT) et l'heure de départ, fixée à 0. Le temps de déplacement, t, est égal à un temps de déplacement de base, $\underline{t}$, auquel il faut ajouter une probabilité de retard p^L et un temps de retard, d, de sorte que $t = \underline{t} + p^L \times d$ (et/ou une probabilité et un temps d'avance). La désutilité marginale associée au temps de déplacement est égale à α, la valeur du temps. Si la personne arrive en avance, c'est-à-dire si $t < PAT$, alors cela génère une désutilité marginale β. L'ampleur de l'avance est mesurée par une variable SDE (*Schedule Delay Early*) égale à $PAT - t$ [12]. Si la personne arrive en retard, c'est-à-dire si $t > PAT$, alors cela génère une désutilité marginale γ, avec $\gamma > \beta$. L'ampleur du retard est

12. Il est également possible de considérer que la personne prend une marge a par rapport à son heure d'arrivée préférée. Elle n'est alors en avance que si $t < PAT - a$

mesuréee par une variable SDL (*Schedule Delay Late*) égale à $t - PAT$ [13]. Par ailleurs, indépendamment de l'ampleur du retard, le fait même d'arriver en retard génère une désutilité θ.

L'espérance de la fonction d'utilité s'écrit donc

$$\mathbb{E}(U) = -\alpha\mathbb{E}(t) - \beta\mathbb{E}(SDE) - \gamma\mathbb{E}(SDE) - \theta p^L. \tag{1.10}$$

Cette approche repose sur des fondements micro-économiques et prend explicitement en compte le fait que, du fait de la variabilité du temps de trajet, l'usager peut arriver soit en avance, soit en retard par rapport à son heure préférée d'arrivée. Les conséquences d'arriver en retard ou en avance sont clairement intégrées au travers des coefficients β, γ et θ. Une telle modélisation est utile dans une perspective de prévision, par exemple, pour modéliser l'heure de départ. En revanche, elle n'est pas adéquate pour évaluer des projets, par exemple par des analyses coûts-bénéfices.

1.5.2.2.2 Approche par une forme réduite Dans une perspective d'évaluation, par exemple pour réaliser une analyse coûts-bénéfices cherchant à évaluation les coûts de la variabilité du temps de trajet (Kouwenhoven et al., 2014), il est nécessaire de passer par une forme réduite, c'est-à-dire une formulation qui intègre non pas explicitement les horaires de départ et d'arrivée, mais synthétise le risque de retard sous la forme d'un indicateur mesurant le risque de retard, tel que le temps de retard moyen, l'écart-type des retards (Batley et Ibáñez, 2012) ou encore des percentiles de la durée de trajet (Brownstone et Small, 2005). La forme réduite peut facilement être déduite d'une modélisation basée sur les horaires. En effet, elle s'appuie sur des mesures statistiques de la variabilité telles que l'écart-type. Il s'agit cependant d'une boîte noire qui ne prend pas en compte les mécanismes comportementaux sous-jacents puisqu'elle ne rend pas compte de comment la variabilité du temps de trajet impacte l'organisation de l'usager (Xiao et Fukuda, 2015).

D'après Börjesson et Eliasson (2011), la forme réduite la plus utilisée pour les services non basés sur des horaires (voiture, TCU) est basée sur l'écart-type du temps de trajet, $\sigma(t)$, et l'espérance du temps de trajet réel, $\mathbb{E}(t)$. Dans cette approche « mean-variance », l'espérance de l'utilité peut s'écrire :

$$E(U) = -\alpha\mathbb{E}(t) - \delta\sigma(t), \tag{1.11}$$

avec δ un paramètre à estimer. Cette approche permet de calculer un ratio de régularité (*reliability ratio*) défini comme le rapport entre l'utilité d'une minute d'écart-type de temps de trajet et l'utilité d'une minute de temps de trajet moyen. L'approche « mean-variance » n'est guère pertinente pour les

13. Si la personne prend une marge, alors elle est en retard si $t > PAT - a$.

transports avec une grille horaire (train, avion) car la distribution du temps de trajet est fortement asymétrique (plus étalée vers la droite). Une alternative consiste donc à formuler l'espérance de l'utilité en fonction de l'espérance de la durée du retard. L'espérance de l'utilité peut alors s'exprimer de la façon suivante :

$$E(U) = -\alpha\mathbb{E}(t) - \lambda\mathbb{E}(d), \tag{1.12}$$

avec λ un paramètre à estimer.

Börjesson et Eliasson (2011) montrent avec cette approche une sur-estimation de la valeur des retards ayant une faible probabilité de survenir. En effet, la valeur du retard augmente non-linéairement avec la probabilité d'occurrence de ce retard : un retard de 50 minutes ayant une probabilité d'occurrence de 0,02 représente une désutilité plus importante qu'un retard de 10 minutes ayant une probabilité d'occurrence de 0,1. Ces estimations économétriques sont complétées d'un modèle théorique qui montre que lorsque l'usager du train peut choisir la « marge » de temps qu'il se réserve, alors le coût du retard croit linéairement avec le temps de retard mais non linéairement avec la probabilité de retard (forme concave). Ce phénomène peut s'expliquer par le fait que lors de retards prévisibles, car ayant une forte occurrence, l'usager du train peut s'organiser pour réduire les conséquences de ce retard. Une telle organisation ayant un coût, l'usager ne la mettra pas en œuvre pour les faibles probabilités de retard. Ces résultats sont similaires à ceux prédits par les modèles de choix dans le risque, et en particulier la *prospect theory* (Tversky et Kahneman, 1992, voir chapitre 7).

1.5.2.3 Prise en compte dans les EPD

La fiabilité du service est régulièrement prise en compte dans les EPD intégrant le train. La déclinaison de la fiabilité en attributs, son intégration dans les questions de choix et les valeurs estimées sont résumées pour six études représentatives de la littérature : Bates et al. (2001), Pons (2011), Börjesson et Eliasson (2011), Börjesson et al. (2012), Batley et Ibáñez (2012), RFF (2013).

Bates et al. (2001) s'appuient sur un modèle basé sur l'horaire. Ils construisent les questions de choix portant sur la fiabilité par étapes successives. Dans un premier temps, les répondants doivent estimer la proportion de train n'arrivant pas à l'heure pour six durées d'avance (plus ou moins de 5 minutes d'avance) ou de retard (moins de 10 minutes de retard, entre 11 et 30, entre 31 et 60 et plus de 60 minutes de retard). Ces questions servent à construire un cadran représentant dix durées d'avance ou de retard. Sur la figure 1.7a, 7E signifie par exemple sept minutes d'avance et 15L signifie quinze minutes de retard.

Pons (2011) prend en compte la durée du retard à l'arrivée. L'exercice de choix indique « il est possible que vous preniez un retard de X minutes » (figure 1.7b). Les durées de retard sont de 0, 5 ou 15 minutes. La probabilité de retard n'est pas prise en compte, ce qui ne permet donc pas d'évaluer un temps moyen de trajet ou l'impact de la durée de retard sur les choix.

Börjesson et Eliasson (2011) utilisent deux enquêtes. Dans la première, la fiabilité est exprimée sous la forme d'une durée de retard et d'un risque de retard en nombre de trains sur 10 (voir figure 1.7c). La seconde exprime la fiabilité sous la forme d'une durée de retard pour un niveau de risque donné [14] et présente également le niveau d'information et la possibilité d'une compensation en cas de retard. Dans cette seconde enquête, il est également demandé aux personnes interrogées de choisir l'amélioration du service qui leur semble la plus souhaitable parmi un ensemble d'améliorations possibles (figure 1.7d).

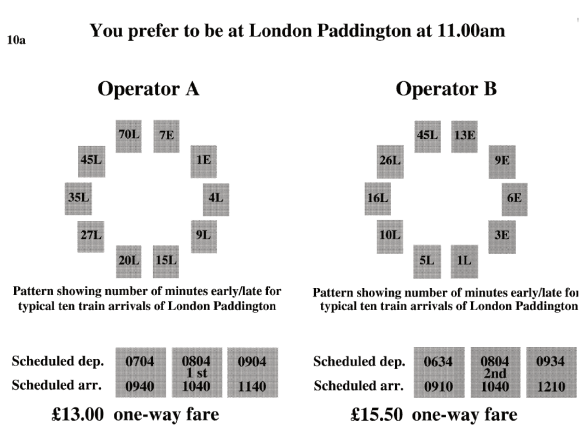
Börjesson et al. (2012) se basent sur deux questionnaires. Le premier est développé dans le cadre d'un modèle basé sur l'horaire et prend en compte l'heure de départ, le temps de déplacement et l'heure d'arrivée qui en découle (figure 1.7e). Le deuxième permet d'estimer une forme réduite et s'appuie sur la durée de retard et le risque de retard exprimé par semaine ou par mois (figure 1.7f).

Batley et Ibáñez (2012) prennent en compte la distribution des retards et avances, et donc du temps de trajet, pour cinq trajets équi-probables (figure 1.7g). Leur modélisation s'inscrit dans une approche basée sur la moyenne et la variance.

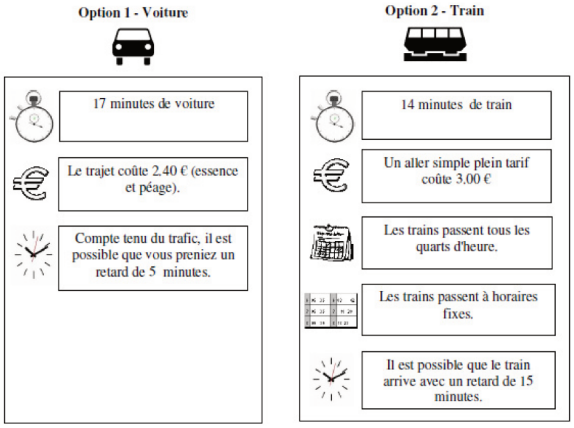
RFF (2013) prend en compte la durée de retard et le risque de retard, exprimé en pourcent et en nombre de trains sur 20 (figure 1.7h). Deux valeurs de retard sont annoncées dans chaque exercice, l'une comprise entre 5 et 25 minutes de retard [15], l'autre entre 30 et 60 minutes de retard [16]. Les durées de retard sont tirées aléatoirement dans les valeurs disponibles[15 et 16]. Les valeurs prises par la probabilité de retard sont 0,05 ; 0,1 ; 0,15. Selon les exercices de choix, l'information aux voyageurs est évaluée simultanément ou non.

Le tableau 1.3 résume les estimations issues des différentes enquêtes et montre une grande hétérogénéité entre les différentes valeurs.

14. Le niveau de risque est le même pour les deux alternatives présentées dans une même question de choix.
15. Retards de 5, 10, 15, 20, 25 minutes.
16. Retards de 30, 40, 45, 50, 60 minutes.



(a) Bates et al. (2001)



(b) Pons (2011)

Choose one!	Departure 1	Departure 2
Fare	**20** SEK more than today	As today
Delays	**1** out of 10 trains are **20** minutes late, the rest are on time	**3** out of 10 trains are **10** minutes late, the rest are on time
Travel time	The scheduled travel time is **2 hours 30 minutes**	The scheduled travel time is **2 hours 45 minutes**
I prefer:	☐ 1	☐ 2

(c) Börjesson et Eliasson (2011) (1)

Which of these improvements would you prefer?	
☐	**30** SEK lower fare than your current fare
☐	The train is delayed on average **once every two weeks** instead of once every week
☐	If a delay occurs, it is on average **15** minutes instead of **30** minutes
☐	You get **information** about the delay length and connections, should a delay occur.

(d) Börjesson et Eliasson (2011) (2)

	Departure 1	Departure 2
Start Time	**25 min later** than today	**5 min later** than today
Travel Time	**15 min longer** than today	**45 min longer** than today
Ticket Price	**€0.70 higher** than today	**€0.40 lower** than today
Arrival Time	*(40 min later than today)*	*(50 min later than today)*
I choose	☐ 1	☐ 2
	Cancel the trip ☐ 3	

(e) Börjesson et al. (2012) (1)

	Departure 1	Departure 2
Delay (if you made this trip every day):	Once **every other month**, the train is **45 min** delayed. All other trips are on-time.	Once **every other week**, the train is **10 min** delayed. All other trips are on-time.
Travel time according to the timetable:	**3 min shorter** than today	**10 min shorter** than today
Ticket price	**€0.20 higher** than today	**€1.00 higher** than today
I choose	☐ 1	☐ 2
	Cancel the trip ☐ 3	

(f) Börjesson et al. (2012) (2)

(g) Batley et Ibáñez (2012)

TER A	TER B
Temps de trajet prévu : … **h** …	Temps de trajet prévu : … **h** …
n_{1A} **trains sur 20** à l'heure *(soit x % de trains en retard)*	n_{1B} **trains sur 20** à l'heure *(soit x % de trains en retard)*
n_{2A} **trains sur 20** en retard de **x min** *(soit x % de trains en retard)*	n_{2B} **trains sur 20** en retard de **x min** *(soit x % de trains en retard)*
n_{3A} **trains sur 20** en retard de **y min** *(soit x % de trains en retard)*	n_{3B} **trains sur 20** en retard de **y min** *(soit x % de trains en retard)*

(h) RFF (2013)

FIGURE 1.7 – Prise en compte de la fiabilité dans les EPD

TABLEAU 1.3 – Valeurs estimées des attributs en lien avec la fiabilité des trains

Source	Désignation de l'attribut	Consentement à payer ou équivalent temps	Ratio de régularité
Bates et al. (2001)	Durée d'avance ou de retard exprimée sur 10 trains	1,8 €/min de retard	
Pons (2011)	Possibilité de prendre de 0 à 15 min de retard	0,4 à 0,46 €/min de retard, selon motif	
Börjesson et Eliasson (2011) (1)	Temps de retard : de 5 à 55 min ; Prob. de retard : 0,05 à 0,15	0,65 à 3,42 €/min de retard, selon motif et prob. de retard	
Börjesson et Eliasson (2011) (2)	Temps de retard : de 15 à 30 min ; Prob. de retard : 0,05 à 0,20	1,48 à 5,26 €/min de retard, selon motif et prob. de retard	
Börjesson et al. (2012)	Temps de retard : de 5 à 70 min ; Prob. de retard : 0,025 à 0,20	0,38 à 0,75 €/min selon modèle utilisé et prob. de retard	
Batley et Ibáñez (2012)	Durée du retard exprimée sur 5 trains		2,1
RFF (2013)	Prob. de retard : 0,05 à 0,15 ; Deux temps de retard présentés : 5 à 25 minutes de retard et 30 à 60 minutes de retard	0,8 à 1,5 min/min de retard (selon le motif et la durée du trajet) ; 4,7 à 11,9 €/min de retard	

Notes : min=minutes, prob.=probabilité. Pour les méthodes de calcul des équivalents temps et consentements à payer, voir section 3.3.2 du chapitre 3. Le ratio de régularité (*reliability ratio*) est défini comme le rapport entre l'utilité d'une minute d'écart-type de temps de trajet et l'utilité d'une minute de temps de trajet moyen.

CHAPITRE 2

Mise en œuvre de l'enquête de préférences déclarées en Région Rhône-Alpes

---

Ce chapitre porte sur la mise en œuvre opérationnelle de l'EPD. Initialement réalisée pour le Conseil Régional Rhône-Alpes, l'EPD permet de répondre aux objectifs de la thèse. L'élaboration de l'EPD s'appuie sur la revue de littérature présentée dans le chapitre 1, dans l'objectif d'obtenir des données conformes à l'état de l'art et dont la qualité est maîtrisée.

L'EPD s'est déroulée en quatre grandes étapes. Dans une première étape, une analyse qualitative a été menée afin d'adapter l'enquête aux spécificités régionales et de pouvoir ensuite donner un sens aux résultats quantitatifs (voir section 2.1). Dans une deuxième étape, les questions de choix (voir section 2.3) ont été développées puis intégrées dans le questionnaire (voir section 2.2). La troisième étape concoure au développement du protocole de recrutement (voir section 2.4). La quatrième étape est la mise en œuvre proprement dite de l'enquête (voir annexe C).

Deux modes d'enquête ont été combinés, ce qui a permis de tirer parti des avantages de chacun d'entre eux. Premièrement, une enquête web a permis d'interroger des personnes identifiées dans l'EDR comme ayant effectué un déplacement en train ou pouvant être fait en train. Deuxièmement, une partie des questionnaires a été administrée en face-à-face afin de sur-échantillonner les usagers du train. En effet, ceux-ci constituent la cible privilégiée de l'EPD mais sont très peu représentés dans l'EDR. Malgré toutes les précautions prises, l'objectif n'a pas été de sélectionner un échantillon représentatif de la population des personnes habitants en Région Rhône-Alpes, mais d'interroger des personnes aux profils diversifiés, notamment en termes de pratiques de déplacement, afin d'évaluer leur sensibilité aux différents déterminants du choix modal.

2.1 Analyse qualitative

La revue de littérature a mis en évidence l'importance d'ancrer l'enquête dans le contexte local, notamment par une analyse qualitative. Onze entretiens individuels et deux focus groups ont donc été réalisés dans le cadre de la préparation de l'EPD. L'annexe B présente les modalités d'application pratiques, le guide d'entretien individuel, le guide d'animation des focus groups ainsi que quelques exemples d'analyses issues de ce travail[1]. L'objectif de l'analyse qualitative était de répondre à trois questions :

- Quelles sont les caractéristiques des modes de transport qui expliquent le choix modal ?
- Quelle est la terminologie employée ?
- Des niveaux de variables se dessinent-ils ?

Les réponses à ces questions ont servi de support pour choisir les attributs. Par exemple, lors des entretiens, il était demandé aux personnes interrogées de classer différentes caractéristiques du train de la plus à la moins importante. Il en ressort que la fiabilité est le critère le plus important. Viennent ensuite en second rang le temps de trajet, la fréquence et le prix. En troisième lieu, l'amplitude horaire est invoquée. Enfin, l'équipement, le cadencement et la propreté sont les éléments jugés comme les moins importants. Les quatre premiers critères (temps, fiabilité, prix et fréquence) sont ceux qui apparaissent dans le premier exercice de choix. Ces attributs sont également ceux qui ressortent le mieux de l'ensemble de l'analyse qualitative (discussion libre, portraits chinois...). Le cadencement apparaît dans le second exercice de choix et le confort dans le troisième exercice de choix, notamment parce qu'il s'agissait d'attributs importants pour la Région Rhône-Alpes en termes opérationnels.

Le choix des niveaux d'attributs s'est également appuyé sur l'analyse qualitative qui a, par exemple, mis en évidence un effet de seuil sur les durées de retard. La plupart des personnes interrogées considèrent qu'en-dessous de 15 minutes, il ne s'agit pas véritablement d'un retard notamment parce qu'ils ont prévu une marge par rapport à l'heure à laquelle ils ont des obligations[2]. Les niveaux du temps de retard (10, 15 et 20 minutes) ont donc été choisis de façon à mettre en évidence ce seuil.

La terminologie employée a quant à elle été utile pour adapter le questionnaire et le rendre accessible au plus grand nombre. Par exemple, les probabilités de retard ont été exprimées en nombre de train sur 10 et en pourcentage. Le nombre de trains sur 10 est parlant pour les personnes utilisant le train tous les

1. Les comptes-rendus bruts et l'analyse complète des réponses sont disponibles sur demande.
2. Ces évaluations dépendent également du temps total de trajet.

jours puisque c'est le nombre d'allers-retours qu'ils font sur une semaine. L'expression en pourcentage permet d'être plus précis. Par ailleurs, le cadencement n'étant pas un terme connu de tous, il a été nécessaire de le définir. Plusieurs formulations et représentations graphiques ont été successivement proposées afin de choisir la meilleure.

L'analyse qualitative a également permis de justifier, illustrer et mettre en perspective les résultats quantitatifs obtenus suite aux estimations économétriques (voir les introductions des parties II et III).

2.2 Questionnaire de l'enquête

Le questionnaire a été construit de façon itérative. Outre les entretiens et focus groups décrits en partie 2.1, des tests de différentes parties du questionnaire, et en particulier les questions de choix, ont été réalisés au fur et à mesure du développement du questionnaire entre mars et juin 2014 [3]. Finalement, des pilotes ont été réalisés dans les trains les 19 et 20 janvier 2015.

Le questionnaire (voir annexe C pour les deux versions du questionnaire) est constitué de cinq parties : recrutement et sélection d'un déplacement (voir section 2.2.1), description du déplacement sélectionné (voir section 2.2.2), questions de choix (voir section 2.2.3), questions socio-économiques (voir section 2.2.4) et questions opinion (voir section 2.2.5)

2.2.1 Recrutement et sélection d'un déplacement

La première partie du questionnaire est consacrée au recrutement de la personne interrogée. L'objectif est de savoir si elle correspond ou non aux critères (voir section 2.4 pour la liste des critères). Pour le CAWI, le recrutement comporte également la sélection d'un déplacement en Région Rhône-Alpes. Pour le CAPI, le déplacement en cours est automatiquement considéré. Pour être sélectionné, un déplacement doit 1) être effectué en Région Rhône-Alpes intégralement ; 2) être effectué en train, car ou voiture ; 3) si le déplacement est fait en voiture, alors il doit pouvoir être reportable vers le train [4]. Pour vérifier cette condition, une base de données contenant l'ensemble des couples origine-destination de la Région Rhône-Alpes réalisables en train a été construite (voir section 2.3.1).

3. Ces tests ont été d'abord réalisés sur la ligne Lyon-Grenoble puis étendus en gare de Lyon et sur la ligne Saint-Etienne-Lyon.
4. Un report est considéré comme possible si un déplacement en train a été observé dans l'EDR pour le même couple origine-destination ou si le temps calculé en train n'excède pas trois fois le temps en voiture.

2.2.2 Description du déplacement sélectionné

La deuxième partie du questionnaire est consacrée à la description du déplacement sélectionné. La description porte en particulier sur le motif, le mode de transport ainsi que les temps et coûts de déplacement. Les temps et coûts de déplacement (déclarés ou calculés) servent à personnaliser les questions de choix à l'aide de la méthode des pivots. Cette personnalisation est nécessaire pour la mise en contexte des questions de choix. Elle est faite à la fois sur les alternatives transports en commun (train et car) et sur les alternatives voiture (voir section 2.3). Les personnes interrogées font donc face à des alternatives réalistes sur l'ensemble des modes.

2.2.3 Questions de choix

Dans la troisième partie, 14 questions de choix sont posées. Elles sont découpées en trois exercices, ce qui évite une surcharge cognitive tout en permettant d'avoir des attributs pertinents dans la prise de décision (voir section 1.3.3 du chapitre 1). Dans les deux premiers exercices, les usagers du train et de la voiture ont le choix entre deux alternatives en train et une en voiture. Les usagers du car ont le choix entre deux alternatives en car et une en voiture. Dans le troisième exercice, les trois modes sont le train, le car et la voiture. Les attributs utilisés sont le temps et le coût du déplacement, le temps de retard et la probabilité de retard des trains, la fréquence de passage des trains, le cadencement et la garantie ou non d'avoir une place assise.

Afin de faciliter la tâche de choix demandée aux enquêtés, chaque exercice est précédé 1) d'une description précise de chacun des attributs (voir figures 2.1, 2.4 et 2.7) ; 2) d'un exemple de choix (voir figures 2.2, 2.5 et 2.8). Pour le CAPI, ces descriptions et exemples sont présentés sur un support papier à part. Les figures 2.3, 2.6 et 2.9 montrent des questions de choix qui auraient pu être réalisés[5]. Par ailleurs, dans l'exercice 1, la question 5 comporte une option en train dominée par l'autre afin de vérifier la cohérence des réponses.

5. Certaines des questions proposent des options de transport en commun peu attractives, ce qui est dû au déplacement saisi pour faire les captures d'écran.

Imaginez que vous ayez 3 options : deux en train et une en voiture pour effectuer le déplacement de Lucenay (69) à Vaulx-en-Velin (69). Merci de nous indiquer laquelle vous choisiriez si ces trois options étaient à votre disposition.

Votre réponse doit être faite en vous plaçant dans la situation du déplacement du 20/02/2015 (même motif de déplacement, même contraintes horaires...). L'ensemble du déplacement est pris en compte, de Lucenay (69) à Vaulx-en-Velin (69) (y compris les trajets pour accéder à la gare de départ et depuis la gare d'arrivée jusqu'à la destination finale).

Les caractéristiques proposées pour ce déplacement varient entre les questions et peuvent différer de la réalité. Nous vous demandons de vous baser sur les valeurs indiquées dans les questions et non sur la situation vécue le 20/02/2015 pour faire vos choix.

Selon quelles caractéristiques faire vos choix ?

Les 5 premiers choix que vous allez faire vont dépendre du retard éventuel, de la fréquence de passage des trains, du temps de déplacement et du coût du déplacement de Lucenay (69) à Vaulx-en-Velin (69) :

TRAIN	VOITURE
Retard : Un retard de 15 minutes 1 fois sur 10 signifie que tous les 10 déplacements, vous aurez 15 minutes de retard à cause d'accidents ou de bouchons en voiture, de problèmes techniques en train ou pour toutes autres raisons. Les autres fois, vous serez à l'heure Le retard est donc défini par **le nombre de retards** : moins d'1 train sur 10 (5%), 1 train sur 10 (10%) ou 2 trains sur 10 (20%) et la **durée de retard** : 10, 15 ou 30 minutes.	
Fréquence : Nombre de trains qui passent chaque heure. Il peut passer un train toutes les deux heures ; 1, 2 ou 4 trains par heure. Exemple avec 4 trains par heure : 4 trains par heure	
Temps de déplacement : temps d'accès à la gare (à pied, en bus, à vélo...), temps passé dans le train et temps pour vous rendre à votre destination finale.	**Temps de déplacement** : temps pour récupérer votre voiture, temps passé dans la voiture, à chercher une place de parking et à vous rendre à votre destination finale.
Coût du déplacement : Coût du ou des titres de transport, y compris le prix du ticket de bus ou de tram ou le coût éventuel de l'essence pour aller ou partir de la gare. Le coût est exprimé par déplacement (aller simple).	**Coût du déplacement** : Coût de l'essence, et éventuellement du parking et du péage. Le coût est exprimé par déplacement (aller simple).

FIGURE 2.1 – Explications pour l'exercice 1

Exemple de réponse

	Option A	Option B	Option C
Moyen de transpor	Si vous choisissez l'option B, vous choisissez de vous déplacer en train.		
Retard	9 fois sur 10 vous êtes à l'heure. Mais 1 fois sur 10, vous êtes en retard de 30 minutes.	Retard de 30 minutes 1 fois sur 10 (10%)	Retard de 15 minutes 1 fois sur 10 (10%)
Fréquence	Il passe 2 trains par heure.	2 trains par heure	Départ quand vous le voulez
Temps de déplaceme	Le déplacement dure 1H21 et vous coûte 18 euros par déplacement.	1H21	1H30
Coût du déplacement		18€ / déplacement	21€ / déplacement
Cochez l'option choisie :	○	◉	○

FIGURE 2.2 – Exemple de choix pour l'exercice 1

	Option A	Option B	Option C
Moyen de transport			
Retard	Retard de 15 minutes 1 fois sur 10 (10%)	Retard de 30 minutes Moins de 1 fois sur 10 (5%)	Retard de 10 minutes 2 fois sur 10 (20%)
Fréquence	 2 trains par heure	 4 trains par heure	Départ quand vous le voulez
Temps de déplacement	1H30	1H07	45 minutes
Coût du déplacement	7€ / déplacement	5€ / déplacement	6€ / déplacement
Cochez l'option choisie :	○	○	○

FIGURE 2.3 – Question de choix pour l'exercice 1

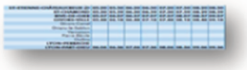
Qu'est-ce que le cadencement ?

Le cadencement signifie que les trains circulent à intervalles de temps réguliers.
Actuellement, la plupart des trains partent à heure fixe d'une gare donnée, par exemple : 5h20, 5h50, 6h20, 6h50...
Le cadencement permet de savoir quand partira le train, quelle que soit l'heure où on rejoint la gare, puisque les minutes de départ sont les mêmes à toutes les heures.

Horaires cadencés

ST-ETIENNE-CHÂTEAUCREUX (D	05.20	05.50	06.20	06.50	07.20	07.50	08.20	08.50
ST-CHAMOND	05.30	05.59	06.30	06.59	07.30	07.59	08.29	08.59
RIVE-DE-GIER	05.37	06.07	06.37	07.07	07.37	08.07	08.37	09.07
GIVORS-VILLE	05.48	06.18	06.48	07.18	07.48	08.18	08.48	09.18
Givors-Canal								
Grigny-le-Sablon								
Vernaison								
Pierre-Bénite								
Oullins								
LYON-PERRACHE								
LYON-PART-DIEU	06.06	06.36	07.06	07.36	08.06	08.36	09.06	09.36

Dans cet exemple de grille horaire, les horaires de train sont cadencés.



Horaires non cadencés

ST-ETIENNE-CHÂTEAUCREUX (D	05.34	05.50	06.13	06.20	06.34		07.04	
ST-CHAMOND	05.43	05.59		06.30	06.43		07.13	
RIVE-DE-GIER	05.51	06.07		06.37	06.50		07.21	
GIVORS-VILLE	06.02	06.18		06.48	07.02	07.08	07.32	07.38
Givors-Canal	06.05					07.12	07.35	07.42
Grigny-le-Sablon	06.09					07.16	07.39	07.46
Vernaison	06.13					07.21	07.43	07.51
Pierre-Bénite	06.20					07.27	07.50	07.57
Oullins	06.23				07.15	07.30	07.53	08.00
LYON-PERRACHE	06.27				07.20	07.35	07.57	08.05
LYON-PART-DIEU		06.36	06.59	07.06				

Dans cet exemple de grille horaire, les trains passent à intervalles de temps irréguliers, les horaires de train ne sont pas cadencés.

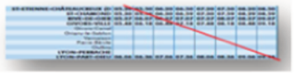


FIGURE 2.4 – Explications pour l'exercice 2

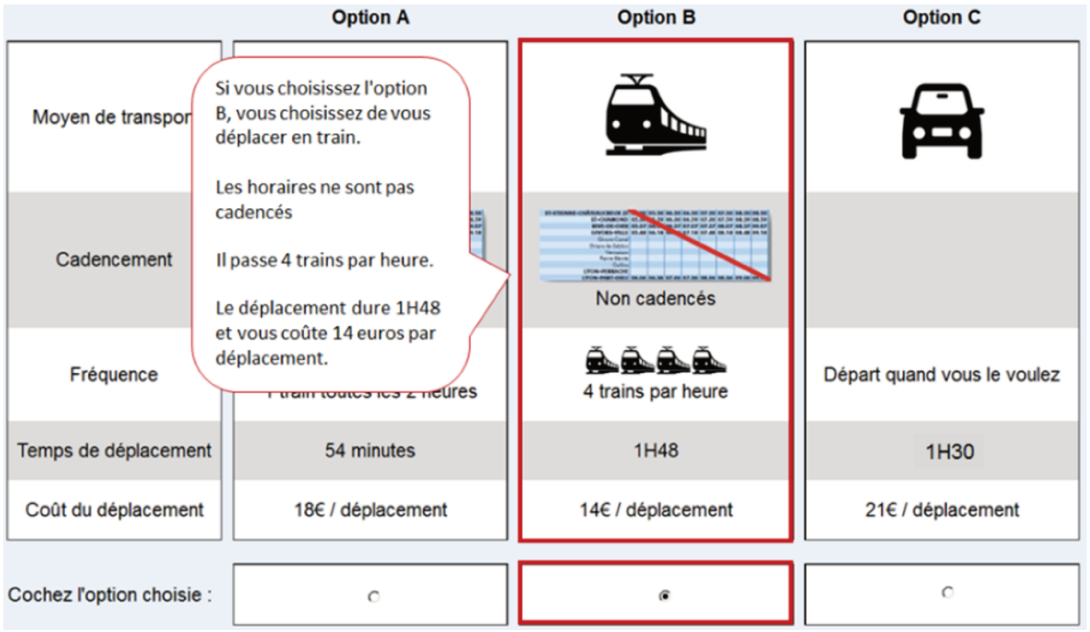


FIGURE 2.5 – Exemple de choix pour l'exercice 2

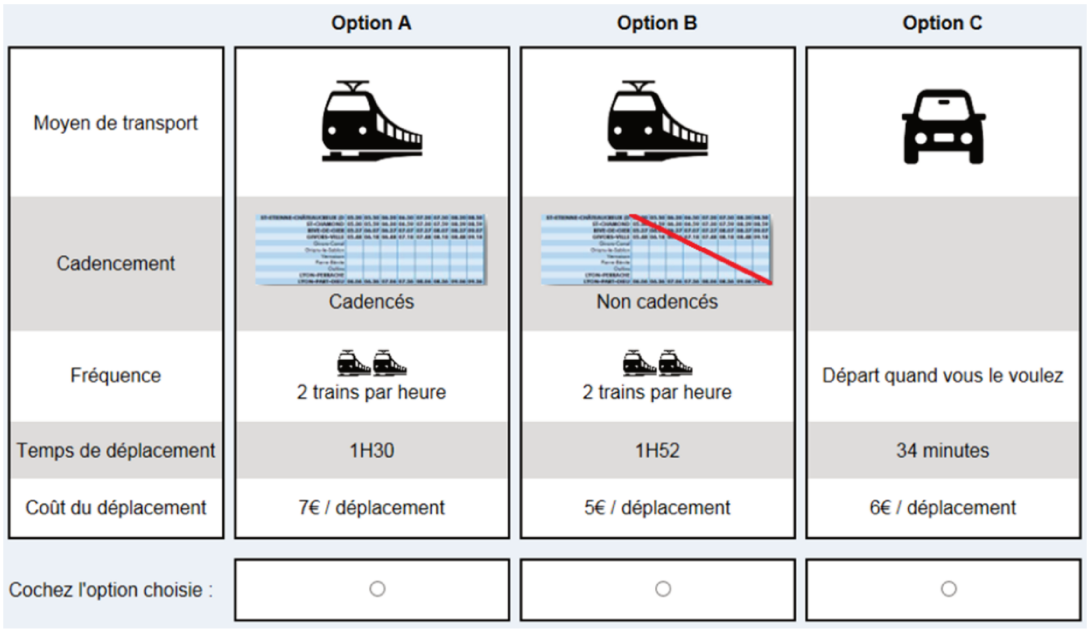


FIGURE 2.6 – Question de choix pour l'exercice 2

Vous allez maintenant faire une dernière série de choix avec une option en train, une option en car et une option en voiture.

Selon quelles caractéristiques faire vos choix ?

Les 4 choix que vous allez faire maintenant vont dépendre de la garantie ou non d'avoir une place assise, du temps de déplacement et du coût du déplacement de Lucenay (69) à Vaulx-en-Velin (69) :

TRAIN	CAR	VOITURE
Place assise : Garantie ou non d'avoir une place assise dans le train. Si la place assise n'est pas garantie, il peut vous arriver de voyager debout sur tout ou une partie du déplacement. La place assise est toujours garantie dans le car. Place assise non garantie : Place assise garantie :		
Temps de déplacement : temps d'accès à la gare ou à l'arrêt de car (à pied, en bus, à vélo...), temps passé dans le train ou le car et temps pour vous rendre à votre destination finale.		Temps de déplacement : temps pour récupérer votre voiture, temps passé dans la voiture, à chercher une place de parking et à vous rendre à votre destination finale.
Coût du déplacement : Coût du ou des titres de transport, y compris le prix du ticket de bus ou de tram ou le coût éventuel de l'essence pour aller ou partir de la gare ou de l'arrêt de car. Le coût est exprimé par déplacement (aller simple).		Coût du déplacement : Coût de l'essence, et éventuellement du parking et du péage. Le coût est exprimé par déplacement (aller simple).

FIGURE 2.7 – Explications pour l'exercice 3

	Option A	Option B	Option C
Moyen de transport			
Place Assise		Place assise garantie	
Temps de déplacement		1H21	1H30
Coût du déplacement	18€ / déplacement	23€ / déplacement	21€ / déplacement
Cochez l'option choisie :	○	◉	○

Si vous choisissez l'option B, vous choisissez de vous déplacer en car.

Vous êtes assuré(e) de disposer d'une place assise

Le déplacement dure 1H21 et vous coûte 23 euros par déplacement.

FIGURE 2.8 – Exemple de choix pour l'exercice 3

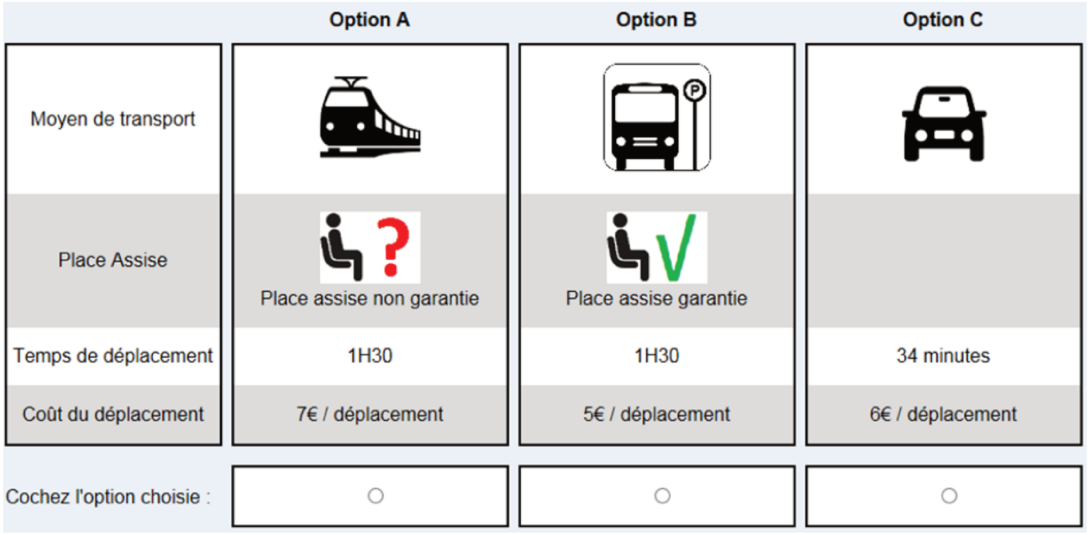


FIGURE 2.9 – Question de choix pour l'exercice 3

L'une des trois alternatives correspond systématiquement au statut quo avec un mode, un temps et un coût identiques au déplacement de référence. Les attributs sont exprimés soit en niveau, soit en variation (voir tableau 2.1). Pour plus de détails sur la personnalisation, voir la partie 2.3. Les temps et les coûts de déplacement varient de -25%, 0% et $+25\%$ par rapport aux temps de référence. Cependant, pour les usagers de la voiture, il a été nécessaire de rendre plus attractif le train et le car en ne faisant pas varier le temps à la hausse mais davantage à la baisse, avec les niveaux suivants : -50%, -25% et 0%.

TABLEAU 2.1 – Variables présentées par exercice

	Niveaux (Unité)	**Exercice 1**	**Exercice 2**	**Exercice 3**
Mode		Train ; train ; voiture	Train ; train ; voiture	Train ; car ; voiture
Temps de déplacement	-50% ; -25% ; 0% ; $+25\%$ (minutes)	X	X	X
Coût de déplacement	-25% ; 0% ; $+25\%$ (€/déplacement ou €/mois)	X	X	X
Fréquence de passage des trains	0,5 ; 1 ; 2 ; 4 (trains/heure)	X	X	-
Temps de retard	10 ; 15 ; 30 (minutes)	X	-	-
Probabilité de retard	5% ; 10% ; 25%	X	-	-
Cadencement	Passage à heures fixes ou non	-	X	-
Confort	Place assise garantie ou non	-	-	X

2.2.4 Questions socio-économiques

Les questions socio-économiques portent sur diverses questions telles que le revenu, l'activité pratiquée à bord ou le fait d'avoir une carte de réduction pour les transports en commun (voir le questionnaire en annexe C).

2.2.5 Questions opinion

La partie opinion du questionnaire comprend des indicateurs mesurant des variables psychologiques sur des échelles de Likert : conscience environnementale, motivations d'utilisation de la voiture et perceptions des transports en commun (voir le questionnaire en annexe C).

Les questions de la partie opinion ont fait l'objet d'un développement particulier. A partir d'une revue de la littérature, l'ensemble des indicateurs candidats a été listé pour refléter les trois facteurs latents (environnement, motivations, perceptions). Un premier épurement a été mené du fait de la redondance de certains indicateurs (formulations différentes pour la même idée) ou de la difficulté à répondre à certaines questions. Ce travail a permis de construire un premier questionnaire visant à évaluer un maximum d'indicateurs, soit 49 indicateurs. Comme dans les enquêtes ménages déplacements, les candidats ont été invités à qualifier trois modes de transport, le train, le car et la voiture à l'aide d'adjectifs (trois au maximum). Parallèlement aux indicateurs socio-psychologiques, des questions visant à refléter le comportement, les habitudes et le contexte socio-économique dans lequel les personnes interrogées évoluent ont été posées.

335 personnes ont répondu à ce questionnaire [6], l'objectif étant de tester la pertinence de l'approche à l'aide d'une analyse factorielle exploratoire et d'un modèle d'équations structurelles qui ont permis 1) de sélectionner les construits *a priori* les plus déterminants pour expliquer le choix modal et 2) de sélectionner dans chacun des construits les indicateurs les plus discriminants à insérer dans le questionnaire de l'EPD.

2.3 Constitution et personnalisation des questions de choix

Cette section est consacrée à la constitution et à la personnalisation des questions de choix pour les attributs du temps et du coût. La personnalisation permet notamment de minimiser le biais

6. Le questionnaire préalable a été distribué en version papier sur la ligne de train Grenoble-Lyon et a été rempli par 50 personnes. Une version internet a également été mise en ligne et 285 personnes ont répondu entre août et septembre 2014. La majorité des réponses émane du personnel de l'Ecole Nationale des Travaux Publics de l'Etat et de l'Université Grenoble Alpes, l'objectif n'étant pas de constituer un échantillon représentatif de la population mais de réaliser des tests.

hypothétique souligné dans la revue de littérature.

2.3.1 Calcul du temps et du coût de référence

Les variations des attributs temps et coûts présentées dans le tableau 2.1 ont été calculées à partir de temps et coûts de référence spécifiques à chaque mode. Cette démarche est originale par rapport à la littérature puisqu'en général les EPD personnalisées portant sur le choix modal se concentrent uniquement sur un corridor.

Pour chaque mode, les temps et les coûts[7] de référence tout compris, de l'origine à la destination (rabattement, en véhicule, diffusion), sont calculés à partir de deux sources d'information : 1) les déclarations faites dans le questionnaire et 2) les informations contenues dans une base spécifiquement créée pour l'EPD (matrice origine-destination)[8]. Celle-ci porte sur les temps et distances afférents aux alternatives train et voiture pour les 8,6 millions de paires de zones fines[9] situées à moins de 10 km d'une gare à vol d'oiseau.

Deux situations ont été différenciées. Le coût est présenté au mois si la personne réalise le déplacement tous les jours ou presque. Sinon, le coût est présenté au trajet. En particulier, pour les automobilistes réalisant le trajet tous les jours ou presque, le coût du train et du car ont été calculés comme s'ils avaient un abonnement.

2.3.2 Constitution des questions de choix : le design

Les questions de choix nécessitent de combiner les attributs et leurs niveaux entre eux pour constituer les options entre lesquelles les usagers font leurs choix. Suite à la revue de littérature (voir section 1.4), il a été décidé de réaliser un design bayésien D-efficient à l'aide du logiciel NGENE. En effet, les *a priori* sur les paramètres sont suffisamment étayés par la littérature pour utiliser un design efficient mais leur diversité mérite d'être prise en compte. Par ailleurs, les designs basés sur la *D-error* sont les plus courants parmi les designs efficients. Quatre designs différents ont été réalisés sur des segments de population : deux pour les automobilistes, selon qu'ils font l'aller-retour moins de trois fois par semaine ou plus de trois fois par semaine et deux pour les usagers du train ou les usagers du

7. Les calculs de coûts diffèrent selon le mode de déplacement, la fréquence de réalisation du déplacement, la disposition ou non d'un abonnement, le mode de rabattement (ticket de transport en commun urbain ou coût de l'essence et du parking en voiture). Pour les déplacements en voiture, le coût prend en compte le coût du péage (pour les déplacements dont la distance est supérieure à 10 km) et de l'essence.

8. Pour chaque couple origine-destination, un calcul d'itinéraires en voiture et en train a été réalisé avec le logiciel Musliw (Cerema). Pour le train, une base horaire simplifiée de la SNCF a été utilisée et les temps de rabattement et de diffusion calculés. Pour chaque itinéraire possible, un coût généralisé a été calculé en se basant sur les travaux de Arentze et Molin (2013) et Wardman (2001) pour les pénalités de correspondance et de marche à pied. Tous les détails de calcul et de mise en œuvre sont dans le rapport remis à la Région Rhône-Alpes.

9. Une zone fine est une zone géographique définie à l'échelle infra-communale.

car selon qu'ils ont ou non un abonnement[10]. Pour les usagers du train et du car, la première alternative correspond au statut quo. Pour les usagers de la voiture, c'est la troisième alternative qui correspond au statu quo.

2.4 Recrutement des usagers

La figure 2.10 représente l'articulation entre recrutement des usagers, mode d'enquête et EDR.

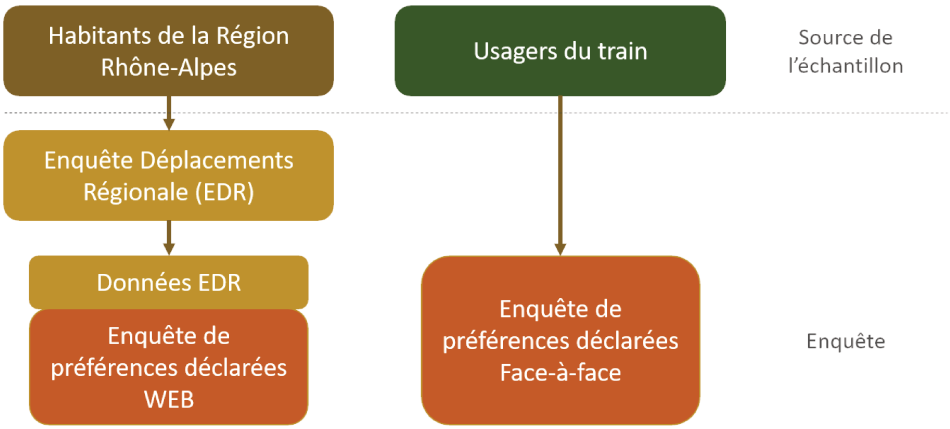


FIGURE 2.10 – Articulation entre source de l'échantillon, mode d'enquête et EDR

2.4.1 Echelle géographique

L'enquête porte sur l'ensemble de la Région Rhône-Alpes (avant reconfiguration des régions françaises). Les déplacements sélectionnés doivent donc avoir leur origine et leur destination dans la Région. Certains déplacements ayant transité par des gares situées en dehors de la Région Rhône-Alpes[11] ont cependant été retenus. En effet, il est possible d'être passé par ces gares à la frontière de la Région, tout en ayant fait un déplacement dont l'origine et la destination sont en Région Rhône-Alpes.

2.4.2 Recrutement CAPI

Les enquêtes CAPI ont été réalisées sur six lignes ferroviaires de la Région Rhônes-Alpes : Lyon/Grenoble, Lyon/Saint-Etienne, Lyon/Bourg-en-Bresse, Valence/Grenoble/Chambéry,

10. Pour chacun des quatre designs, des hypothèses sur les temps et les coûts moyens de déplacement ont été effectuées à partir des données des enquêtes photos et de l'EDR. Tous les détails sur la construction des designs sont dans le rapport remis à la Région Rhône-Alpes.

11. En Saône-et-Loire (Macon TGV, Macon ville, Romaneche Thorins, Pontanevaux, Creche sur Saône), en Suisse (Genève) et dans le Vaucluse (Bollène la croisière)

Grenoble/Chambéry (voir figure 2.11). Ces lignes ont été choisies pour leur représentation du maillage territorial.

608 personnes ont répondu aux questions de choix. Les refus de réponse étaient ponctuels (personnes travaillant ou descendant à l'arrêt suivant) et moins de 6% des personnes ayant commencé le questionnaire ne sont pas allées jusqu'à la fin des questions demandées. Les autres personnes n'ayant pas répondu à l'ensemble des questions ne correspondaient pas aux filtres suivants : réaliser un déplacement en Région Rhône-Alpes, ne pas être ayant-droit SNCF, avoir plus de 18 ans, avoir le permis et avoir au moins une voiture à disposition dans le ménage (pour être en mesure de considérer l'alternative voiture).

Pour l'échantillonnage, la méthode des quotas a été utilisée selon cinq critères : âge $\times$ sexe, motif de déplacement, durée de déplacement et ligne ferroviaire [12].

2.4.3 Recrutement CAWI

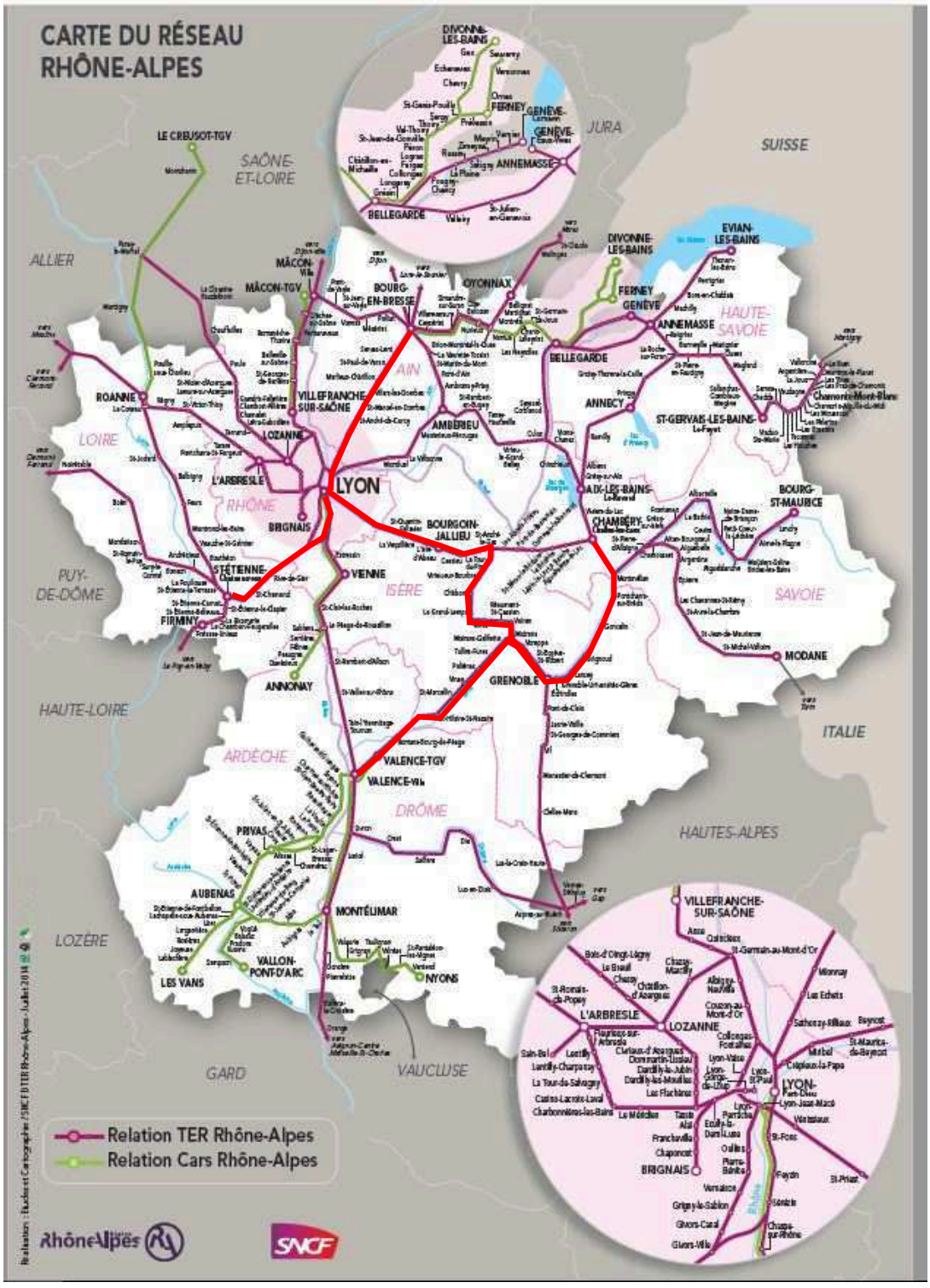
Parmi les 37 479 personnes enquêtées dans l'EDR (échantillonnage stratifié géographiquement), 18 958 personnes ont été sélectionnées et ont reçu un courrier les invitant à répondre au questionnaire de l'EPD en ligne. La sélection s'est faite sur la base des réponses aux questions de l'EDR. Outre la contrainte que la personne ait accepté de répondre à nouveau à une enquête, une première série de filtres permet de garantir que les personnes interrogées aient un accès à la voiture en tant que conducteur :

- appartenir à un ménage ayant au moins une voiture à disposition ;
- avoir au moins 18 ans ;
- avoir le permis.

Après application de ces filtres, la sélection s'est faite selon quatre critères, toujours sur la base des réponses à l'EDR :

- avoir effectué un déplacement en train ou en car la veille du jour enquêté ;
- avoir effectué un déplacement en train dans le mois précédent le jour enquêté ;
- avoir effectué un déplacement en voiture de plus de 20 km la veille du jour enquêté ;
- avoir effectué un déplacement de plus de 20 km dans le mois précédant le jour enquêté.

12. Les détails des quotas demandés et effectués sont disponibles dans le rapport remis à la Région Rhône-Alpes. Ces quotas ont été définis à partir de statistiques descriptives réalisées sur les enquêtes photos et la première vague de l'EDR.



Source : Région Rhône-Alpes

FIGURE 2.11 – Lignes ferroviaires en Région Rhône-Alpes. En rouge, les lignes enquêtées pour l'EPD.

CHAPITRE 3

Première analyse des résultats

Ce chapitre présente les bases de données brutes issues de l'EPD, quelques statistiques descriptives et l'estimation de premiers modèles de choix discret.

3.1 Bases de données brutes

3 455 personnes ont commencé le questionnaire, dont 2 396 CAWI et 1 059 CAPI. Le taux de retour est donc de 12,2% pour le CAWI. Toutes les personnes ayant commencé le questionnaire ne sont pas allées jusqu'au bout soit par ce qu'elles ne répondaient pas aux critères de sélection (ne pas être ayant droit SNCF, effectuer son déplacement intégralement en Région Rhône-Alpes...), soit par choix d'arrêter avant la fin. Dans le CAPI (respectivement, CAWI), 617 personnes (resp. 1 297) ont répondu au moins à la première question de choix et 608 (resp. 1 215) sont allées jusqu'au bout des trois exercices de choix. La proportion de personnes ayant arrêté est plus importante dans le CAWI (6,32%) que dans le CAPI (1,46%) mais reste globalement relativement faible (4,75%). Le tableau 3.1 donne le détail pour chaque exercice de choix. Après apurement [1], 1 799 ont répondu aux questions de choix de l'exercice 1, 1 752 à l'exercice 2 et 1 722 à l'exercice 3. L'apurement engendre une perte de 5%.

TABLEAU 3.1 – Nombre de personnes ayant commencé et fini chaque exercice de choix

	Personnes			**Observations**
	Début	Fin (avant apurement)	Fin (après apurement)	
Exo 1	1 914	1 893	1 799	8 988
Exo 2	1 853	1 841	1 752	8 746
Exo 3	1 827	1 823	1 722	6 853

1. Un travail d'apurement a du être réalisé sur les données. Quatre principaux types d'erreurs ont été relevés : 1) des erreurs dans le dessin d'enregistrement ; 2) quelques personnes qui ont été interrogées et n'auraient pas dû l'être ; 3) des erreurs ou, plus généralement, des incohérences dans les réponses fournies par les personnes enquêtées ; 4) des erreurs dans la structure de la BDD. Sans qu'il ne s'agisse d'erreurs à proprement parler, les réponses systématiques qui ont été faites dans les questions de choix ont systématiquement été relevées. Afin de détecter les erreurs, une vérification « manuelle » de toutes les questions ouvertes (champs « autres ») et recherche automatique de mots clés potentiellement « suspects » (avion, Paris, villes étrangères...) a été réalisée. Par ailleurs, une vérification systématique des niveaux de temps et de coûts, des ratios de temps entre options, des ratios de coûts entre options et des ratios temps/coût pour chaque option a été réalisée. Tous les détails de l'apurement (liste des erreurs, variables et personnes concernées) sont dans le rapport remis à la Région Rhône-Alpes.

3.2 Statistiques descriptives

Afin d'évaluer l'ampleur de la sélection intervenue entre la base initiale de l'EDR et la base finale de l'EPD, et en particulier celle s'appuyant sur l'échantillon de l'EDR (EPD CAWI), le tableau 3.2 présente des statistiques descriptives sur l'âge, le genre et le revenu.

TABLEAU 3.2 – Comparaison des échantillons (ensemble des personnes de l'EDR, personnes sélectionnées et échantillon final de l'EPD, par mode d'enquête) sur trois variables (âge, genre et revenu)

Age (en années)						
Echantillon	**N (personnes)**	**Moyenne**	**Médiane**	**S.d.**	**Min.**	**Max.**
Personnes enquêtées dans l'EDR CATI	37 479	50,86	52,00	20,21	11,00	99,00
Personnes sélectionnées pour l'EPD	18 958	52,30	52,00	15,49	18,00	99,00
Echantillon final EPD CAWI	1 209	53,30	54,00	13,27	8,00	87,00
Echantillon final EPD CAPI	590	36,48	32,00	15,40	19,00	90,00
Echantillon final EPD	1 799	47,78	48,00	16,07	18,00	90,00
Genre (proportion d'hommes)						
Echantillon	**N (personnes)**	**Proportion**				
Personnes enquêtées dans l'EDR CATI	37 479	43,98%				
Personnes sélectionnées pour l'EPD	18 958	46,53%				
Echantillon final EPD CAWI	1 209	49,55%				
Echantillon final EPD CAPI	590	51,86%				
Echantillon final EPD	1 799	50,31%				
Revenu (proportion > 4 000 €/mois)						
Echantillon	**N (personnes)**	**Proportion**				
Personnes enquêtées dans l'EDR CATI	37 479	16,62%				
Personnes sélectionnées pour l'EPD	18 958	19,41%				
Echantillon final EPD CAWI	1 209	28,95%				
Echantillon final EPD CAPI	590	21,05%				
Echantillon final EPD	1 799	26,37%				

Il en ressort que les personnes de l'EDR sélectionnées pour répondre à l'EPD sont légèrement plus âgées que l'ensemble des personnes ayant répondu à l'EDR (52,30 ans contre 50,86 ans), ce qui s'explique facilement par la sélection uniquement de personnes de plus de 18 ans. Les personnes ayant finalement répondu à l'EPD CAWI sont encore un peu plus âgées (53,3 ans). En revanche, l'échantillon final de l'EPD est globalement constitué de personnes un peu plus jeunes (47,78 ans) car celles ayant répondu dans les trains sont en moyenne beaucoup plus jeunes (36,48 ans)[2]. Dans l'échantillon final de l'EPD CAWI, la proportion d'hommes est plus élevée (49,55%) que dans l'échantillon initial de l'EDR (43,98%), ce qui peut s'expliquer par une mobilité de plus courte distance pour les femmes. Elles ont donc pu être moins interrogées que les hommes puisque leurs déplacements se prêtent moins à un report modal vers le train. Cet effet est encore accentué dans l'échantillon final par une proportion supérieure à 50% d'hommes interrogées dans le train. La proportion de personnes habitant dans un

2. Ce qui est conforme à ce qui est observé parmi les usagers du train de l'EDR et des enquêtes photos.

ménage aux revenus supérieurs à 4 000 €/mois est plus faible dans l'échantillon initial de l'EDR (16,62%) que dans l'échantillon de personnes sélectionnées pour l'EPD (19,41%) et que dans l'échantillon final de l'EPD CAWI (28,95%), ce qui peut également s'expliquer par une mobilité de plus courte distance pour les personnes des ménages eu-deçà du seuil de 4 000 €/mois. Cet effet est partiellement compensé par l'échantillon de personnes interrogées dans le train dans lequel la proportion est de 21,05%. En conclusion, la principale différence entre l'échantillon initial de l'EDR et celui de l'EPD provient de différences de revenus (du moins sur ces trois indicateurs).

La suite de la section présente des statistiques descriptives sur les variables prises en compte dans les modèles de choix discret (attributs, variables individuelles, part modale). Du fait de données manquantes pour la taille du ménage, les statistiques descriptives portent sur un échantillon de 24 473 réponses fournies par 1 788 personnes, dont un-tiers a été interrogé dans le train et deux-tiers par internet [3].

Des statistiques descriptives sont également présentées sur les activités pratiquées durant le déplacement et sur l'évaluation du temps de transport, ces deux variables pouvant également expliquer le choix modal et étant examinées dans le chapitre 4.

3.2.1 Attributs

Le tableau 3.3 présente les statistiques descriptives pour l'ensemble des attributs de l'EPD. La figure 3.1 permet de visualiser la relation entre temps et coût de déplacement pour chacune des trois alternatives. La figure 3.2 (respectivement la figure 3.3) trace la distribution du temps (respectivement, du coût) de déplacement en voiture en fonction du temps (respectivement, du coût) de déplacement en train pour les usagers des transports en commun (train, car) et pour les usagers de la voiture.

3. Dans la section 3.3 et les parties suivantes (parties II et III), la taille de l'échantillon peut varier selon les variables explicatives retenues et les besoins d'apurement des données.

TABLEAU 3.3 – Statistiques descriptives sur les attributs

Définition	Label	Ex.	My.	Méd.	S.d.	Min.	Max.
Temps en trainA (min)	Temps (train)	1,2,3	69,13	52,00	52,03	5,00	325,00
Temps en trainB/car (min)	Temps (train/car)	1,2,3	68,70	52,00	53,42	5,00	375,00
Temps en voiture (min)	Temps (voiture)	1,2,3	57,55	45,00	37,80	4,00	330,00
Coût en trainA (€)	Coût (train)	1,2,3	8.66	6.00	7.37	0.63	62.00
Coût en trainB/car (€)	Coût (train/car)	1,2,3	8.02	5.54	7.59	0.04	78.00
Coût en voiture (€)	Coût (voiture)	1,2,3	10.09	7.00	8.89	0.38	71.00
Temps retard en trainA (min)	Tretard (train)	1	15.90	15.00	6.18	10.00	30.00
Temps retard en trainB (min)	Tretard (train)	1	19.20	15.00	8.46	10.00	30.00
Temps retard en voiture (min)	Tretard (voiture)	1	17.55	15.00	6.63	10.00	30.00
Prob. de retard en trainA	Pretard (train)	1	0,10	0,10	0,05	0,05	0,20
Prob. de retard en trainB	Pretard (train)	1	0,11	0,10	0,06	0,05	0,20
Prob. de retard en voiture	Pretard (voiture)	1	0,11	0,10	0,39	0,05	0,20
Fréquence (trainsA/heure)	Freq (train)	1, 2	2.05	2.00	1.06	0.50	4.00
Fréquence (trainsB/heure)	Freq (train)	1, 2	2.10	2.00	1.42	0.50	4.00
Horaires cadencés (1 si oui ; 0 sinon)	Cad. (train)	2	0.74	1.00	0.44	0.00	1.00
Horaires cadencés (1 si oui ; 0 sinon)	Cad. (train)	2	0.49	0.00	0.50	0.00	1.00
Place garantie (1 si oui ; 0 sinon)	Confort	3	0.50	1.00	0.50	0.00	1.00

Notes : Ex. : exercice. My. : moyenne. Méd. : médiane. Min. : minimum. Max. : maximum. Les exercices de choix 1 et 2 contiennent deux alternatives trains (A et B). Pour l'exercice 3, les alternatives sont le train, le car et la voiture.

FIGURE 3.1 – Coût de déplacement en fonction du temps de déplacement

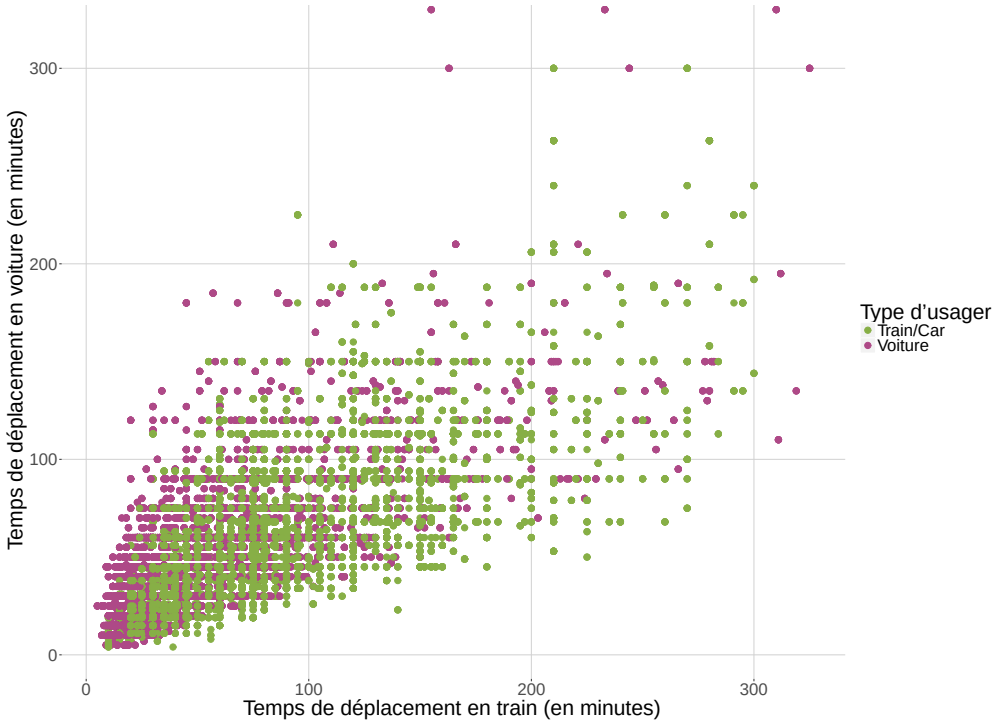


FIGURE 3.2 – Temps de déplacement en voiture en fonction du temps de déplacement en train

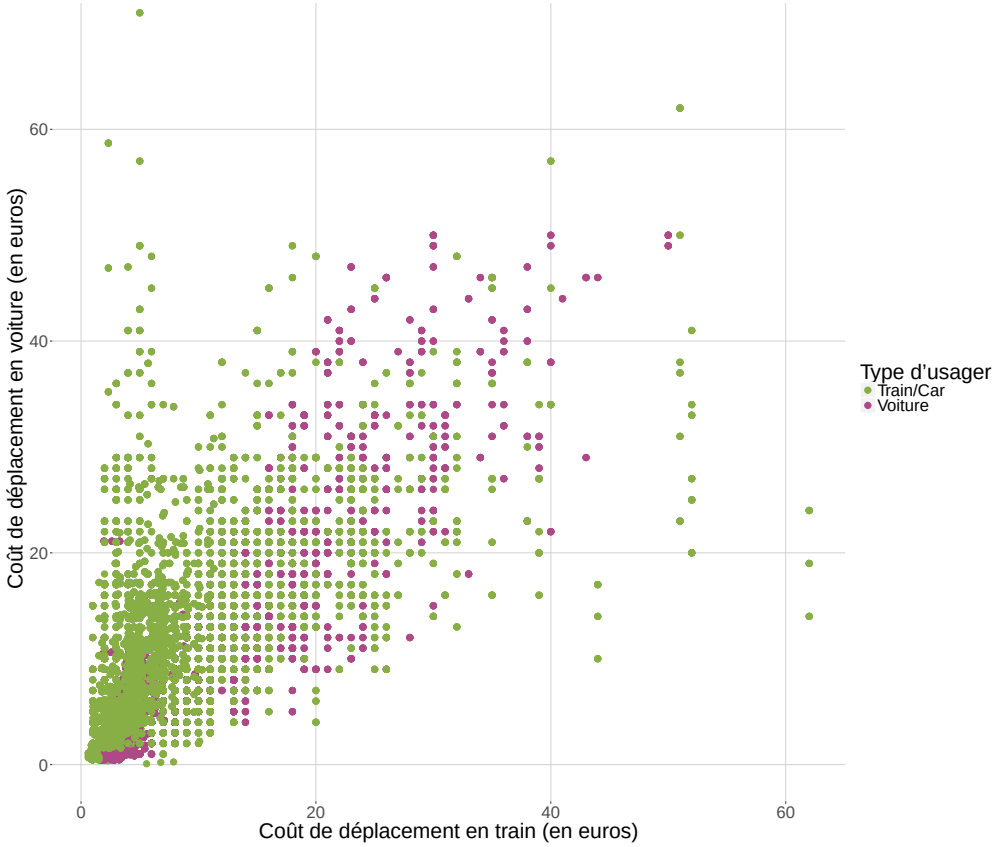


FIGURE 3.3 – Coût de déplacement en voiture en fonction du coût de déplacement en train

3.2.2 Variables individuelles

Parmi les répondants, 54% ont utilisé la voiture pour le déplacement de référence. 43% sont des usagers du train et 3% des usagers du car. La différence d'âge en fonction du mode d'enquête se retrouve par type d'usagers puisque les automobilistes ayant répondu à l'enquête ont en moyenne 53 ans, les usagers du car 57 ans et les usagers du train 40 ans.

TABLEAU 3.4 – Statistiques descriptives sur l'âge (en années)

		Moyenne	**Médiane**	**S.d.**	**Min.**	**Max.**
Ensemble		47,8	48	16,1	18	90
Type d'usager	Usager voiture	53,5	54	13,1	20	86
	Usager train	39,9	37	16,3	18	90
	Usager car	56,6	58	14,1	20	87

En termes de motif, le travail domine, suivi des visites ou activités de loisir (voir tableau 3.5). Au total, les motifs obligatoires (travail, études) représentent 43% des motifs (à l'origine ou à destination).

TABLEAU 3.5 – Motif à l'origine et à la destination (part en %)

	Origine	**Destination**
Domicile	73.72	15.84
Travail	12.96	25.86
Collège	0.00	0.06
Lycée	0.40	0.42
Universités et grandes écoles	2.62	2.74
Démarches personnelles (administratif, santé, recherche d'emploi…)	2.44	6.81
Visite ou activité de loisir	4.05	21.97
Manger hors du domicile	0.34	2.89
Aller chercher/déposer quelqu'un	0.65	3.95
Grands Magasins et leurs galeries marchandes	0.51	3.99
Hypermarchés et Supermarchés	0.13	1.45
Petits et moyens commerces, marchés	0.95	2.70
Autre	1.22	11.31

En termes de motorisation, plusieurs indicateurs peuvent être construits. Afin de représenter la contrainte que l'accès (ou le non accès) à une voiture représente pour certains ménages, la motorisation est ici calculé comme le nombre de voitures à disposition dans le ménage divisé par le nombre d'adultes dans le ménage. Il y a évidemment un choix simultané qui peut se faire entre mode de transport et équipement du ménage, d'où un problème d'endogénéité dans le modèle. En moyenne dans l'échantillon, il y a 2,6 personnes par ménage (dont 2,1 adulte) et ils possèdent 1,7 voiture. Chaque adulte a en moyenne un peu moins d'une voiture (0,9) à disposition, avec une forte hétérogénéité selon les ménages.

Notons enfin que l'échantillon est parfaitement équilibré avec 50% d'hommes et 50% de femmes.

Le tableau 3.6 synthétise les statistiques descriptives pour les variables individuelles utilisées dans le modèle de choix discret.

TABLEAU 3.6 – Statistiques descriptives des variables individuelles

Label	Définition	Moyenne	Médiane	S.d.	Min.	Max.
Age	En années	47.79	48.50	16.10	18.00	90.00
Genre	Genre (1 si homme ; 0 sinon)	0.50	0.00	0.50	0.00	1.00
Motif_obligatoire	Motif à destination (1 si obligatoire ; 0 sinon)	0.42	0.00	0.49	0.00	1.00
Type_car	Type d'usager (1 si car ; 0 sinon)	0.03	0.00	0.18	0.00	1.00
Type_voiture	Type d'usager (1 si voiture ; 0 sinon)	0.54	1.00	0.50	0.00	1.00
Nb_voit_ad	Nombre de voitures/adulte	0.87	1.00	0.36	0.12	4.00

3.2.3 Part modale

La première alternative train (ou car, pour les usagers du car) a été choisie par 30% des répondants, 18% ont choisi la deuxième alternative qui représente soit le train, soit le car selon l'exercice et 52% des répondants ont choisi l'alternative voiture (voir tableau 3.7).

TABLEAU 3.7 – Part modale (en %)

Option	Proportion
A (Train)	0,30
B (Train/Car)	0,18
C (Voiture)	0,52

3.2.4 Evaluation du temps et activités

L'une des spécificités du train relativement à la voiture et, dans une moindre mesure, au car, est la possibilité de pratiquer des activités durant le déplacement. La figure 3.4 illustre la diversité d'activités pratiquées dans le train [4] : 8% des usagers du train ne font rien. Les autres en profitent pour travailler (28%), écouter de la musique ou la radio (28%), observer le paysage ou encore les gens (19%). La proportion de personnes qui ne fait rien dans le car est plus élevée et s'élève à 12% et près de 60% passe du temps à regarder le paysage, les gens... Seuls 4% travaillent durant leur déplacement en car. En voiture, 45% des automobilistes ne font rien d'autre que conduire ou attendre, pour les passagers. Environ la même proportion écoute de la musique ou la radio. Certains d'entre eux, en faible proportion (1,5%), arrivent tout de même à travailler ou étudier.

Du fait de la pratique de ces activités, nous pourrions nous attendre à ce que les usagers du train (et du car) évaluent leur temps de déplacement plus favorablement que les usagers de la voiture.

4. Seules les activités les plus pratiquées sont illustrées.

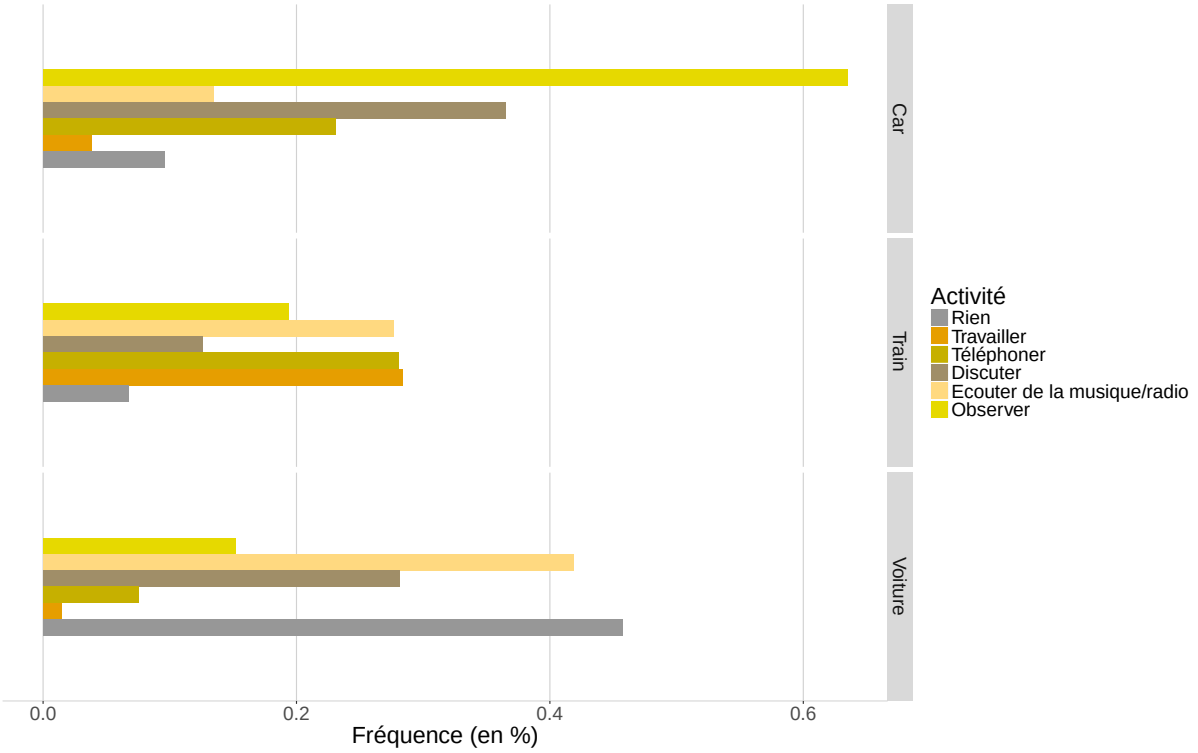


FIGURE 3.4 – Activités pratiquées pendant le déplacement par type d'usager

Pourtant, la figure 3.5 montre que les automobilistes évaluent plus favorablement leur temps de trajet que les usagers des transports en commun. Un test du χ^2 confirme une différence significative, même en prenant en compte les différences de temps de trajet (en moyenne plus courts en voiture) et en regroupant les catégories d'évaluation.

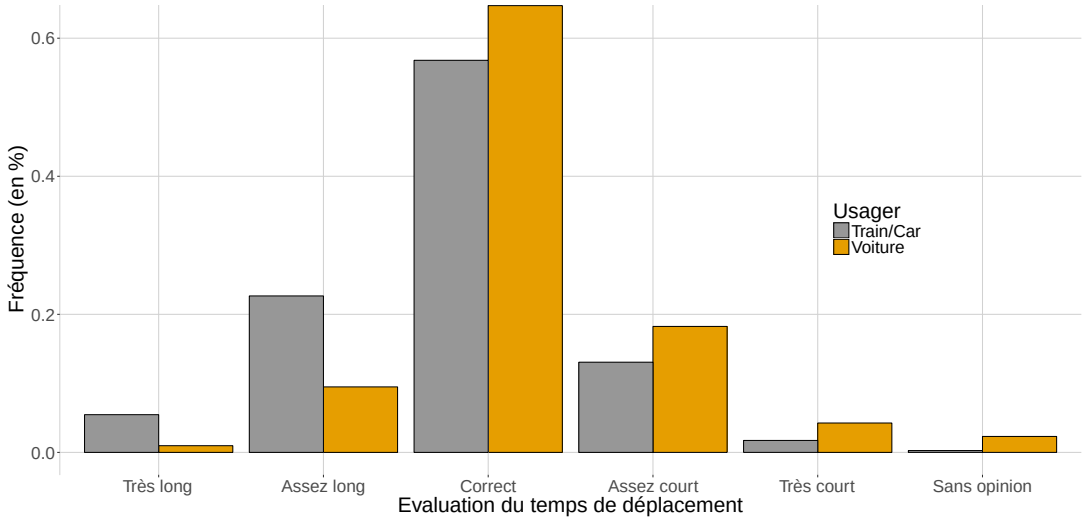


FIGURE 3.5 – Evaluation du temps de déplacement par type d'usager

3.3 Modèles de choix discret

Une première section (3.3.1) présente les résultats des estimations de modèles logit multinomiaux[5]. Une deuxième section (3.3.2) présente les outputs économiques (consentements à payer, équivalent temps, variation compensatoire et élasticités).

3.3.1 Estimations

Un (ou deux) modèle(s) de choix discret ont été estimés sur chacun des trois exercices de choix puis un modèle a été estimé sur l'ensemble des trois exercices.

Le premier exercice de choix fait intervenir le temps et le coût de trajet, ainsi que la probabilité et le temps de retard et la fréquence de passage des trains. La probabilité et le temps de retard peuvent apparaître de multiples façons dans le modèle. Deux solutions sont testées ici. Un premier modèle (MNL1a) considère l'espérance du temps de trajet $\mathbb{E}(t_j) = \underline{t}_j + p_j \times d_j,\ \forall j = 1, 2, 3$, avec j l'alternative, $\underline{t}_j$ le temps de trajet sans retard, p_j la probabilité de retard et d_j le temps de retard. Un deuxième modèle (MNL1b) valorise séparément les probabilités de retard. Sont valorisés d'un côté le retard ayant une faible probabilité d'occurrence (p_j=0,05 ou p_j=0,1) et de l'autre ceux ayant une forte probabilité d'occurrence (p_j=0,2).

Le modèle MNL2 correspond au deuxième exercice de choix et porte sur le temps et le coût du trajet, ainsi que la fréquence de passage des trains et le cadencement.

Le modèle MNL3 correspond au troisième exercice de choix et porte sur le temps et le coût de trajet ainsi que le confort dans le train, mesuré par la garantie ou non d'avoir une place assise dans le train. Seule l'utilité de l'alternative train comprend la variable confort, égale à 1 si la place assise est garantie et à 0 sinon. Le confort est plus ou moins important selon le temps de trajet. Une variable croisée Temps×Confort est donc introduite afin d'analyser l'interaction entre les deux variables.

Le modèle MNL123 estime conjointement les coefficients de l'ensemble des attributs de l'EPD, tous exercices confondus. La formulation pour le retard est la même que pour le modèle MNL1b. En complément des attributs, une indicatrice est incluse spécifiquement pour les questions de l'exercice 3 pour l'alternative train afin de prendre en compte un effet contexte qui permet de corriger le fait que

5. Toutes les estimations de modèles et les simulations ont été réalisées à l'aide du logiciel Python Biogeme (Bierlaire, 2016). Une vérification systématique des résultats a été faite avec R (mlogit) et SAS (proc mdc). D'autres modèles, avec une prise en compte non linéaire de la fréquence de passage des trains et une valorisation du point de probabilité de retard ont également été estimés pour intégration dans le modèle multi-modal régional. Ils ont fait l'objet d'une note spécifique et ne sont pas présentés ici par souci de parcimonie. Des modèles mixed logit et des *scale heterogeneity models* (Fiebig et al., 2010) qui autorisent la variance des termes d'erreur à varier selon les individus ont également été estimés. Ils ne sont pas présentés ici puisque l'objectif est de présenter des modèles les plus simples possibles à visée opérationnelle et qu'ils n'ont pas été nécessaires pour remplir les objectifs de la thèse présentés dans les chapitres suivants.

l'attribut confort n'est intégré que pour le train et dans le cadre de cet exercice de choix. Cette combinaison d'exercices dans un même modèle est réalisée grâce à l'existence d'attributs communs (temps et coût) à tous les exercices. L'hypothèse est faite d'un facteur d'échelle différent pour chaque exercice. Ces facteurs d'échelle (λ_j) sont inversement proportionnels à l'écart-type du terme aléatoire de l'utilité (σ_{ε_j}) et reflètent le degré de dispersion dans les comportements

$$\lambda_j = \frac{1}{\sigma_{\varepsilon_j}},\ \forall j = 1, 2, 3. \tag{3.1}$$

Le facteur d'échelle du premier exercice est normalisé à un ($\lambda_1 = 1$). Si la dispersion des réponses est la même dans les différents exercices, alors le facteur d'échelle est égal à un dans les trois exercices.

Dans tous les modèles, l'alternative de référence est la voiture. Les résultats, présentés dans le tableau 3.8, sont globalement conformes aux attentes et cohérents entre les différents modèles.

Premièrement, les personnes interrogées préfèrent choisir l'alternative voiture plutôt que l'alternative train dans les modèles MNL1a, MNL1b et MNL2. Dans le modèle MNL3, le train est associé à une constante positive car le confort est pris en compte. En revanche, le car n'est pas associé à une constance modale significativement différente de la voiture. Ces résultats se retrouvent dans l'exercice MNL123 : la voiture est préférée au train, excepté quand le confort est pris en compte (exercice 3) ; le car est associé à une constante non significativement différente de zéro. Dans la mesure où les données sont des données d'enquête de préférences déclarées, ces interprétations des constantes ne sont qu'indicatives puisqu'elles sont estimées pour reproduire la part modale observée. Concernant le temps de déplacement, le temps de retard et le coût, les coefficients associés sont négatifs comme attendu. Cela signifie que plus le temps de déplacement ou de retard est long, moins une alternative est choisie. De même, plus le coût d'une alternative est élevé, moins cette alternative est choisie. La fréquence de passage des trains et des cars ainsi que le cadencement de leurs horaires sont associés à des coefficients positifs. En d'autres termes, les alternative de transport en commun ont une probabilité plus élevée d'être choisies lorsque leur fréquence de passage augmente et lorsque leurs horaires sont cadencés. Concernant le confort dans les trains, la variable dichotomique traduisant la garantie ou non d'avoir une place assise n'est pas associée à un coefficient significativement différent de zéro dans le modèle MNL3 et associé à un paramètre significativement positif dans le modèle MNL123. Le coefficient de la variable croisée Time×Confort est positif comme attendu. Cela signifie que plus le temps de trajet est long, plus le confort est important. Parallèlement, quand la place assise est garantie, le temps de trajet influence moins fortement le choix.

TABLEAU 3.8 – Estimation des modèles

Variable	MNL1a	MNL1b	MNL2	MNL3	MNL123
ASC Train	-0.994 (0.259)***	-1.230 (0.269)***	-0.937 (0.272)***	0.462 (0.304)	-0.891 (0.157)***
ASC Car				0.425 (0.302)	0.015 (0.157)
Exo3					0.951 (0.063)***
Coût	-0.095 (0.006)***	-0.105 (0.006)***	-0.122 (0.007)***	-0.130 (0.007)***	-0.113 (0.004)***
Temps (Voiture)	-0.023 (0.001)***	-0.024 (0.001)***	-0.029 (0.001)***	-0.028 (0.002)***	-0.026 (0.001)***
Temps (Train/Car)	-0.020 (0.001)***	-0.022 (0.001)***	-0.024 (0.001)***	-0.0310 (0.001)***	-0.024 (0.001)***
Freq	0.338 (0.018)***	0.352 (0.018)***	0.338 (0.017)***		0.322 (0.012)***
Tretard (p=0,20)		-0.063 (0.005)***			-0.066 (0.005)***
Tretard (p=0,05/0,1)		-0.035 (0.003)***			-0.036 (0.003)***
Cad.			0.528 (0.041)***		0.484 (0.031)***
Confort				0.011 (0.088)	0.219 (0.082)**
Temps×Confort				0.009 (0.001)***	0.005 (0.001)***
Age	0.058 (0.011)***	0.065 (0.012)***	0.044 (0.012)***	0.050 (0.013)***	0.502 (0.067)***
Age^2	-0.0005 (0.0001)***	-0.001 (0.0001)***	-0.0003 (0.0001)***	-0.0005 (0.0001)***	-0.045 (0.007)***
Genre	-0.207 (0.058)***	-0.211 (0.060)***	-0.242 (0.061)***	-0.129 (0.065)***	-0.190 (0.034)***
Motif_obligatoire	0.329 (0.065)***	0.344 (0.068)***	0.221 (0.068)***	0.245 (0.073)***	0.260 (0.039)***
Nb_voit_ad	-0.405 (0.080)***	-0.408 (0.083)***	-0.347 (0.082)***	-0.491 (0.091)***	-0.388 (0.047)***
Type_car	-0.818 (0.148)***	-0.902 (0.152)***	-0.729 (0.154)***	-1.030 (0.195)***	-0.810 (0.092)***
Type_voiture	-3.310 (0.081)***	-3.100 (0.083)***	-3.350 (0.082)***	-3.060 (0.089)***	-3.030 (0.069)***
λ_2					1.12 (0.028)***
λ_3					1.01 (0.031)***
$\mathcal{L}(\hat{\theta})$	-6 680	-6 553	-6 387	-5 327	-18 320
$\bar{\rho}^2$	0.318	0.331	0.331	0.288	0.318
N	8933	8933	8708	6829	24470

Notes : (***) sign. à 1% ; (**) sign. à 5% ; (*) sign. à 10%. ASC: Alternative Specific Constant (constantes modales). Exo3 : variable dichotomique specifique à l'exercice 3 pour le modèle MNL123. λ : paramètres d'échelle. N : taille de l'échantillon.

Le modèle renseigne également sur l'impact des variables individuelles sur le choix modal. Premièrement, plus une personne est âgée, plus elle privilégie les transports en commun (train et car) par rapport à la voiture. Cependant, cet effet diminue avec l'âge puisque la relation est de forme concave (coefficient associé à âge^2 négatif). Par ailleurs, les hommes sont moins enclins que les femmes à utiliser les transports en commun. De même, la probabilité de choisir les alternatives train ou car est plus faible pour les personnes dont le mode de référence est la voiture ou le car ou pour les personnes ayant un nombre élevé de voitures à disposition. En revanche, les personnes se déplaçant pour le travail ou pour les études choisissent davantage les alternatives train et car que celles se déplaçant pour un autre motif.

Dans le modèle MNL123, les deux facteurs d'échelle (λ_2 et λ_3) sont significatifs ce qui signifie que la variance des termes d'erreur (l'inverse du paramètre d'échelle) ou, en d'autres termes, l'hétérogénéité inobservée, varie selon les exercices de choix. C'est dans l'exercice 2 que la variance des termes d'erreurs est la plus élevée, puis dans l'exercice 3 et enfin dans l'exercice 1 (où elle est fixée à 1).

3.3.2 Outputs économiques

Les modèles de choix discret peuvent être utilisés pour calculer des outputs économiques indispensables à l'analyse économique. Par exemple, les consentements à payer sont nécessaires à la réalisation d'analyses coûts-bénéfices, les élasticités permettent d'analyser l'évolution de la demande suite à une modification de l'un des déterminants de la demande (prix, qualité...)[6]. Après avoir défini chacune des notions (consentement à payer, équivalent temps, variation compensatoire et élasticités), les résultats estimés avec les données de l'EPD sont présentés.

3.3.2.1 Consentement à payer et équivalent temps

3.3.2.1.1 Définitions Soient s_j le coût de l'alternative j et $\beta_{cost,j}$ le coefficient qui y est associé. Soient $y_{k,j}$ l'attribut pour lequel on veut connaître le consentement à payer (*Willingness-To-Pay*; WTP) et $\beta_{k,j}$ le coefficient qui y est associé. Soit U_j l'utilité associée à l'alternative j. Les indices individuelles sont omis pour simplifier. Le consentement à payer est la variation nécessaire dans le coût s_j pour garder l'utilité constante suite à une variation de l'attribut $y_{k,j}$. La valeur du temps est un exemple de consentement à payer pour diminuer le temps de déplacement.

6. Les outputs présentés dans cette partie ne prennent pas en compte les poids d'échantillonnage puisque l'EPD n'en contient pas. En effet, l'échantillon est volontairement non représentatif de la population.

La dérivée totale de l'utilité est

$$dU_j = \frac{\partial U_j}{\partial s_j} ds_j + \frac{\partial U_j}{\partial y_{k,j}} dy_{k,j},\ \forall k = 1, \cdots, K;\ j = 1, \cdots, J. \tag{3.2}$$

Le consentement à payer pour l'attribut $y_{k,j}$ est la solution en $ds_j/dy_{k,j}$ de l'équation $dU_j = 0$:

$$\begin{aligned} \text{WTP}_{y_{k,j}} &= -\frac{\partial U_j/\partial y_{k,j}}{\partial u_j/\partial s_j} \\ &= -\frac{\beta_{k,j}}{\beta_{cost,j}},\ \forall k = 1, \cdots, K;\ j = 1, \cdots, J. \end{aligned} \tag{3.3}$$

Au lieu de calculer ce que la personne interrogée serait prête à payer pour compenser une variation de $y_{k,j}$, il est aussi possible de calculer un équivalent temps qui mesure de combien il faudrait diminuer (respectivement, augmenter) le temps de trajet t_j, associé au coefficient $\beta_{time,j}$ pour compenser une dégradation (respectivement, amélioration) de $y_{k,j}$. L'équivalent temps (*time equivalent* ; TE) s'écrit

$$\begin{aligned} \text{TE}_{y_{k,j}} &= -\frac{\partial u_j/\partial y_{k,j}}{\partial u_j/\partial t_j} \\ &= -\frac{\beta_{k,j}}{\beta_{time,j}},\ \forall k = 1, \cdots, K;\ j = 1, \cdots, J. \end{aligned} \tag{3.4}$$

3.3.2.1.2 Estimations L'ensemble des consentements à payer et des équivalent temps sont présentés dans la Table 3.9 puis comparées par rapport aux valeurs trouvées dans la littérature.

Valeur du temps Les valeurs du temps sont relativement homogènes entre les exercices de choix. En voiture, la valeur du temps est comprise entre 13,02 et 14,38 €/heure. Lorsque le confort n'est pas pris en compte, la valeur du temps dans les transports en commun est comprise entre 12,00 et 12,86 €/heure. Elle est un peu plus élevée dans le modèle MNL1a que dans le modèle MNL1b car le premier comprend également les temps de retard qui sont perçus plus négativement par les usagers. Dans les modèles MNL3 et MNL123, la valeur du temps est différenciée, pour le train, selon que la place assise soit garantie ou non. Pour le car et le train sans garantie de place assise, elle est estimée à 14,31 €/heure (MNL3) ou 12,64 €/heure (MNL123). Pour le train avec garantie de place assise, elle est plus faible et estimée à 10,20 €/heure (MNL3) ou 9,94 €/heure (MNL123). La diminution de la valeur du temps dans cette configuration traduit le fait que la place assise permet de pratiquer des activités (travailler, se reposer, etc.). Ces valeurs sont dans les ordres de grandeur du rapport Quinet et al. (2014) (voir Annexe J pour une discussion plus approfondie).

TABLEAU 3.9 – Consentements à payer et équivalents temps des attributs de l'EPD

	MNL 1a	**MNL 1b**	**MNL 2**	**MNL 3**	**MNL 123**
WTP : Valeur du temps (en €/heure)					
Train (conf=0)/Car	12,86	12,51	12,00	14,31	12,64
Train (conf=1)				10,20	9,94
Voiture	14,38	13,66	14,31	13,02	13,70
WTP : Temps de retard (prob.=0,05 ou 0,1 ; en €/heure)					
Tous modes		20,06			19,33
WTP : Temps de retard (prob.=0,2 ; en €/heure)					
Tous modes		36,23			35,10
TE : Temps du retard (prob.=0,05 ou 0,1 ; en minutes)					
Train/Car		1,60			1,53
Voiture		2,47			1,41
TE : Temps du retard (prob.=0,2 ; en min)					
Train/Car		2,89			2,77
Voiture		2,65			2,56
WTP : Fréquence (en €/train supplémentaire/heure)					
Train/Car	3,57	3,35	2,77		2,85
TE : Fréquence (en min/train supplémentaire/heure)					
Train/Car	16,65	16,07	13,85		13,53
WTP : Cadencement (en €)					
Train/Car			4,33		4,28
TE : Cadencement (en min)					
Train/Car			21,64		20,34
WTP : Confort (en €)					
Train (temps = 30 min)				2,14	3,29
Train (temps = 60 min)				4,19	4,63
Train (temps = 90 min)				6,25	5,98
TE : Confort (en min)					
Train (temps = 30 min)				8,96	15,61
Train (temps = 60 min)				17,58	22,01
Train (temps = 90 min)				26,21	28,41

Notes : min=minutes ; prob. = probabilité ; WTP=consentement à payer ; TE=équivalent temps. Pour le confort, « temps » indique le temps de déplacement pour lequel l'output est calculé.

Retard Conformément aux attentes, la valeur du temps de retard est plus élevée que celle du temps de trajet prévu. Cette différence reflète notamment des coûts d'organisation et du stress. Les retards ayant une probabilité d'occurrence de 0,05 ou 0,1 sont associés à une valeur de 19,33 à 20,06 €/heure de retard selon le modèle. Ces ordres de grandeur sont comparables à ceux de la littérature (Börjesson et Eliasson, 2011 ; Pons, 2011, voir tableau 1.3, chapitre 1). Les fourchettes de valeurs sont plus élevées pour les retards ayant une probabilité d'occurrence de 0,2 puisqu'ils sont associés à une valeur de 35,10 à 36,23 €/heure de retard. En transport en commun, avec le modèle MNL3, une minute de trajet prévu est équivalent à 1,6 minute de retard ayant une probabilité de 0,05 de survenir et à 2,89 minutes de retard ayant une probabilité de 0,2 de survenir. Les équivalents temps du retard sont sensiblement inférieurs pour la voiture car la valeur du temps est plus élevée. Une minute de retard ayant une probabilité d'occurrence de 0,05 ou 0,1 (respectivement 0,2) en voiture est équivalente à 1,47 minutes (respectivement 2,65 minutes) de temps prévu. Les valeurs sont sensiblement équivalentes avec le modèle MNL123. Ces résultats montrent que les valeurs ne sont pas linéaires en la probabilité de retard puisqu'une probabilité de 0,05 ou 0,1 ne génère pas des valeurs quatre fois, ni même deux fois, plus faibles qu'une probabilité de 0,2. Ce résultat est similaire à celui de Börjesson et Eliasson (2011) (voir chapitre 1).

Fréquence de passage Les usagers sont prêts à payer de 2,77 à 3,57 € de plus par déplacement pour voir la fréquence de passage augmenter d'un train ou un car par heure. Ils sont également prêts à passer de 13 à 17 minutes de plus dans les transports en commun pour bénéficier d'une telle augmentation de fréquence. Ces valeurs semblent relativement élevées et sont en tous cas supérieures à celles trouvées par Pons (2011) (voir tableau 1.2, chapitre 1).

Cadencement Les usagers sont prêts à payer environ 4,3 € ou à passer 20 minutes supplémentaires dans le train ou le car (modèle MNL2). Les résultats sont similaires pour le modèle MNL123. Si ces valeurs peuvent sembler relativement élevées, elles sont équivalents à celles trouvées par Wardman et al. (2004) si l'on somme les valeurs trouvées pour chacune des composantes du cadencement (régularité des intervalles, passage à minutes fixes, horaires mémorisables) mais restent supérieures à celles trouvées par Pons (2011) (voir tableau 1.2, chapitre 1).

Confort Le consentement à payer pour bénéficier d'une place assise garantie dans le train dépend du temps de déplacement : plus celui-ci est élevé plus le consentement à payer est important. Ainsi, pour un temps déplacement de 30 minutes, le consentement à payer est de 2,14 à 3,29 € selon le

modèle ; pour un temps de déplacement de 60 minutes, le consentement à payer est de 4,19 à 4,63 € ; pour un temps de déplacement de 90 minutes, le consentement à payer est de 5,98 à 6,25 €. Les usagers sont également prêts à passer plus de temps dans le train pour bénéficier d'une place assise : selon le temps de déplacement, de 9 à 26 minutes d'après le modèle MNL3 et de 16 à 28 minutes d'après le modèle MNL123. Il y a encore relativement peu de littérature sur les consentement à payer ou équivalents-temps du confort. Cependant, les ordres de grandeur sont cohérents avec celles trouvées par exemple par Richter et Keuchel (2012) ou RFF (2013) (voir les chapitres 5 et 8 pour une discussion plus approfondie sur les valeurs du confort).

3.3.2.2 Variation compensatoire

3.3.2.2.1 Définition La variation compensatoire permet d'évaluer l'effet d'une politique publique ou de simuler l'effet d'un scénario sur l'utilité. Le calcul requiert la définition des scénarios à évaluer, par exemple le passage d'une situation dans laquelle le niveau d'un attribut passe d'un niveau faible *ex-ante* ($y_{k,j}^{\{0\}}$) à un niveau élevé *ex-post* ($y_{k,j}^{\{1\}}$). Les indices individuels sont omis pour simplifier. En considérant un revenu R, l'effet d'une transition entre scénarios est mesuré en calculant la variation monétaire ΔR nécessaire pour que le niveau de l'utilité correspondant à l'alternative choisie, c'est-à-dire celle ayant l'utilité maximale, soit identique entre la situation *ex-ante* et la situation *ex-post*. L'utilité maximale n'étant pas observable par le modélisateur, la variation compensatoire est définie par l'égalité entre l'espérance des utilités maximales :

$$\mathbb{E}\left[\max_{j\in\mathcal{C}}\left(U_j^{\{0\}}\right)\right] = \mathbb{E}\left[\max_{j\in\mathcal{C}}\left(U_j^{\{1\}}\right)\right], \tag{3.5}$$

en utilisant les définitions suivantes

$$\begin{aligned} U_j^{\{0\}} &= V_j^{\{0\}}(R - s_j, y_{k,j}^{\{0\}}) + \varepsilon_j^{\{0\}}, \\ U_j^{\{1\}} &= V_j^{\{1\}}(R - \Delta R - s_j, y_{k,j}^{\{1\}}) + \varepsilon_j^{\{1\}}, \end{aligned}$$

avec V_j la partie déterministe de l'utilité, s_j le coût de l'alternative j et ε_j le terme aléatoire. Avec un logit multinomial, l'espérance du maximum de l'utilité s'écrit (Small et Rosen, 1981)

$$\mathbb{E}\left[\max_{j\in\mathcal{C}}\left(U_j^{\{l\}}\right)\right] = \ln\sum_{j=1}^{J}\exp^{V_j^{\{l\}}} + A,\ \forall l = \{0,1\}, \tag{3.6}$$

avec A une constante.

L'espérance de ΔR est (Small et Rosen, 1981)

$$\begin{aligned}\mathbb{E}(\Delta R) &= \frac{1}{\beta_{cost}}\left[\mathbb{E}\left[\max_{j\in\mathfrak{C}}\left(U_j^{\{0\}}\right)\right] - \mathbb{E}\left[\max_{j\in\mathfrak{C}}\left(U_j^{\{1\}}\right)\right]\right] \\ &= \frac{1}{\beta_{cost}}\left[\ln\sum_{j=1}^{J}\exp^{V_j^{\{0\}}} - \ln\sum_{j=1}^{J}\exp^{V_j^{\{1\}}}\right], \end{aligned} \tag{3.7}$$

avec β_{cost} le coefficient générique associé au coût des alternatives.

Cette expression de la variation compensatoire fait l'hypothèse qu'il n'y a pas d'effet de revenu, l'utilité est linéaire en $R - s_j$. Sous cette hypothèse, la variation compensatoire est la version hicksienne du surplus du consommateur.

3.3.2.2.2 Estimations Le tableau 3.10 présente les variations compensatoires calculées pour les attributs discrets de chacun des exercices de choix. Pour les attributs continus (temps de déplacement, temps de retard), la variation compensatoire pour une variation marginale entre les scénarios *ex-ante* et *ex-post* est équivalente au ratio de coefficients avec lequel les consentements à payer ont été calculés précédemment. Globalement, les variations de surplus calculés sont inférieures aux consentements à payer calculés avec le ratio de coefficients. Un travail complémentaire serait nécessaire pour mieux comprendre ces différences qui peuvent être dues à la présence de la constante dans le logsum, sans pour autant avoir vraiment de sens lorsque les modèles sont estimés sur la base de données issues d'une EPD[7].

TABLEAU 3.10 – Variation compensatoire (en €) pour la fréquence de passage, le cadencement et le confort

	MNL 1a	**MNL 1b**	**MNL 2**	**MNL 3**
Fréquence : de 0,5 à 1 train/heure				
Train/Car	0,71	0,67	0,57	
Fréquence : de 1 à 2 trains/heure				
Train/Car	1,54	1,45	1,23	
Fréquence : de 2 à 4 trains/heure				
Train/Car	5,13	4,82	4,08	
Cadencement				
Train/Car			2	
Confort				
Train (temps = 30 min)				0,63
Train (temps = 60 min)				1,30
Train (temps = 90 min)				2,05

Notes : min=minutes. Pour le confort, « temps » indique le temps de déplacement pour lequel l'output est calculé.

7. Les valeurs estimées pour le modèle MNL123 ne sont pas reportées car elles sont aberrantes pour le confort, peut être en raison de la constante modale qui n'est pas correctement estimé avec des données d'EPD.

Afin d'évaluer l'équivalent monétaire d'une augmentation de la fréquence de passage des trains et des cars, quatre scénarios ont été simulés : passage d'un train toutes les deux heures, passage d'un, deux ou quatre trains par heure. Par exemple, avec le modèle MNL1a, le passage d'un scénario avec un train ou car toutes les deux heures à un scénario avec un train ou car par heure est valorisé à 0,71 €, la transition de un à deux trains ou cars par heure est valorisé à 1,54 € et celle de deux à quatre trains ou cars par heure est valorisé à 5,13 €. Des ordres de grandeur similaires, quoique plus faibles, sont trouvés dans les modèles MNL1b et MNL2.

Cadencement Le passage de grilles horaires cadencés à des grilles horaires non cadencés génère un surplus de 2 € par déplacement.

Confort Pour un temps de déplacement de 30 minutes, les usagers seraient prêts à payer leur trajet 0,63 € de plus pour avoir la garantie d'avoir une place assise. Si le temps de trajet est de 60 minutes (respectivement 90 minutes), alors la variation de surplus est de 1,30 € (respectivement 2,05 €).

3.3.2.3 Elasticites

3.3.2.3.1 Définitions L'élasticité $\mathrm{e}_i(y_{k,j})$ mesure de combien varie (en %) la probabilité π_i de choisir l'option i suite à une variation du $k$ème attribut $y_{k,j}$ associé à l'option j. Si i =j, alors on parle d'élasticité propre et si $i \neq j$, alors on parle d'élasticité croisée. Comparativement aux effets marginaux, les élasticités présentent l'avantage de ne pas dépendre de l'unité de la variable $y_{k,j}$.

Dans le cas où l'utilité de l'option i est linéaire relativement à l'attribut $y_{k,i}$ associé au coefficient $\beta_{k,i}$, alors l'élasticité pour l'individu n (élasticité point) s'écrit pour tout $j = 1, \cdots, J$; $i = 1, \cdots, J$; $k = 1, \cdots, K$; $n = 1, \cdots, N$:

$$\begin{aligned} \mathrm{e}_{n,i}(y_{n,k,j}) &= \frac{\partial \pi_{n,i}}{\partial y_{n,k,j}} \times \frac{y_{n,k,j}}{\pi_{n,i}} \\ &= y_{n,k,j}\beta_{k,j}(\mathbf{1}_{(i=j)} - \pi_{n,j}), \end{aligned} \tag{3.8}$$

A l'échelle de l'échantillon, il est possible de calculer des élasticités agrégées, $\mathrm{we}_i(y_{k,j})$, pondérées par les probabilités de choix, pour tout $j = 1, \cdots, J$; $i = 1, \cdots, J$; $k = 1, \cdots, K$:

$$\mathrm{we}_i(y_{k,j}) = \frac{\sum_{n=1}^{N} \mathrm{e}_{n,i}(y_{n,k,j})\pi_{n,i}}{\sum_{n=1}^{N} \pi_{n,i}}. \tag{3.9}$$

3.3.2.4 Estimations

Le tableau 3.11 présente les élasticités propres et les élasticités croisées pondérées pour le temps et le coût de déplacement des modèles 1a et 3 [8].

TABLEAU 3.11 – Elasticités propres et élasticités croisées (temps et coût)

Alternative i	Attribut alternative j	MNL 1a	MNL 3 (confort=0)	MNL 3 (confort=1)
Elasticités propres : temps				
TrainA		-0,68	-1,55	-0,75
TrainB/Car		-0,64	-1,34	
Voiture		-0,27	-0,46	
Elasticités propres : coût				
TrainA		-0,34	-0,61	
TrainB/Car		-0,31	-0,69	
Voiture		-0,20	-0,37	
Elasticités croisées : temps				
TrainA	TrainB/Car	0,54	0,57	
	Voiture	0,41	0,46	
TrainB/Car	TrainA	0,64	0,83	0,82
	Voiture	0,38	0,93	
Voiture	TrainA	0,18	0,37	0,35
	TrainB/Car	0,17	0,23	
Elasticités croisées : coût				
TrainA	TrainB/Car	0,24	0,25	
	Voiture	0,30	0,38	
TrainB/Car	TrainA	0,23	0,35	
	Voiture	0,27	0,35	
Voiture	TrainA	0,11	0,21	
	TrainB/Car	0,11	0,14	

Notes : Le modèle MNL1a contient deux alternatives train, la première alternative présentée est dénommée « trainA » et la deuxième « trainB ». Pour le modèle MNL 3, les élasticités pondérées sont différenciées selon le niveau de confort. Les alternatives sont le train, le car et la voiture.

Elasticités modèle MNL 1a Les élasticités pondérées sont conformes à la littérature. Les élasticités propres sont inférieures à 1, ce qui indique une relative inélasticité de la probabilité de choix de chacune des alternatives relativement au temps ou au coût. Selon Ortuzar et al. (2011), l'élasticité coût du bus est généralement de 0,3. C'est également l'ordre de grandeur estimé par Glerum et al. (2014) pour les transports public (train et car) avec des données issues d'une enquête de préférences révélées réalisée en Suisse. En Région Rhône-Alpes, les estimations sont conformes à ces références puisqu'une augmentation de 10% du coût de déplacement en train induit une diminution de 3,1% (trainA) à 3,4% (trainB) de la probabilité de choisir l'option train.

Les élasticités temps sont plus élevées : une augmentation de 10% du temps de déplacement en train

8. Les résultats sont similaires pour les modèles MNL1b et MNL2. Par contre, les élasticités estimées avec le modèle MNL123 ne sont pas cohérentes, vraisemblablement pour la même raison que pour la variation compensatoire. Ceci montre que les élasticités estimées à l'aide d'une EPD doivent être considérées avec précaution (Hensher, 2007 ; Louviere et al., 2000).

induit une diminution de 6,4% (trainA) à 6,8% (trainB) de la probabilité de choisir une alternative train.

Conformément aux intuitions et à la littérature, l'élasticité coût (respectivement temps) de la voiture est plus faible que celle du train puisqu'une augmentation de 10% du coût (respectivement temps) de la voiture induit une diminution de 2,0% (respectivement 2,7%) de la probabilité de choisir ce mode de transport. L'élasticité temps de la voiture est parfaitement cohérente avec celle estimée par Glerum et al. (2014) : -0,28. En revanche, l'élasticité coût de la voiture est relativement élevée puisqu'elle par exemple estimée à -0,07 par Glerum et al. (2014).

Une fois les pondérations prises en compte, les élasticités croisées montrent que la substitution entre les deux alternatives train est plus forte que celle entre le train et la voiture. Par exemple, une augmentation de 10% du temps de trajet pour le trainA induit une augmentation de 6,4% de la probabilité de choix du trainB et de 1,8% de la probabilité de choix de la voiture.

Une variation de 10% du temps de trajet a davantage d'impact sur les probabilités de choix qu'une même variation du coût de trajet, ce qui est également cohérent avec la littérature (voir, par exemple, De Jong et Gunn, 2001). Ce constat vaut pour les élasticités propres et les élasticités croisées.

Elasticités modèle MNL 3 Dans le modèle MNL3, les élasticités propres sont élevées comparativement au modèle MNL 1a, ce qui est vraisemblablement dû à la prise en compte du confort. Pour analyser les effets d'une variation du temps du train, deux situations doivent être distinguées, selon qu'une place assise est garantie ou non. Si la place assise n'est pas garantie, alors l'élasticité propre est plus élevée que si elle est garantie. Cela signifie qu'une personne ayant potentiellement à rester debout dans le train est plus sensible à une variation du temps de trajet qu'une personne ayant la garantie de pouvoir rester assise. En revanche, les élasticités croisées sont relativement similaires quel que soit le niveau de confort.

L'augmentation du temps de trajet en voiture engendre un report modal vers le car supérieur au report vers le train. En revanche, une variation du coût de la voiture génère un report relativement uniforme vers les deux alternatives de transport en commun.

Les autres résultats sont relativement similaires à ceux du modèle MNL 1a : 1) les élasticités temps sont supérieures aux élasticités coût ; 2) la probabilité de choisir la voiture est moins sensible aux variations de son temps et de son coût que les probabilités de choix des transports en commun.

Part II

Modeling mode choice (habits) with latent variables - contributions to an economic and behavioral perspective

Introduction

> As in life, what matters is not the final destination, but all the interesting things that occur along the way.
>
> Ingold (2016), Lines: A Brief History

Objectives

Building on behavioral theories, the overall objective of Part II is to explain mode choice (habits) by means of latent variables, and to investigate how the integration of latent variables in Discrete Choice Models (DCMS) can be used in economic analyses. Indeed, behavioral theories require the use of latent variables, that is, variables that cannot be directly observed but that are inferred from observed variables. This raises the issue of the measurement of latent variables, which has to be addressed with a specific tool, a Structural Equation Model (SEM). Using DCMs and SEMs, this part has, more specifically, three objectives:

1. to review the mode choice literature on the integration between discrete choice and latent variable models by means of Integrated Choice and Latent Variable (ICLV) models, that is, models which incorporate a SEM in a DCM. This work will help to understand how environmental psychology theories can be integrated in DCMs and how latent variables may be used to investigate heterogeneity in behaviors;
2. to estimate an ICLV model and highlight the heterogeneity in value of time depending on perceived and objective comfort during travel by public transport. The perceived comfort is modeled with latent variables;
3. to analyze the direct and indirect effect of environmental concern on mode choice habits by means of SEMs; indirect effects are mediated by the indirect utility of travel.

Besides the literature review, the focus of this part is therefore on comfort in public transport and the indirect utility of travel.

Motivation

Motivation for explaining mode choice with latent variables

Chetty (2015, p.1) identifies three levers by which behavioral economics improves the definition of public policies: "by offering new policy tools, improving predictions about the effects of existing policies, and generating new welfare implications". This insight can be extended to the inclusion of latent variables in DCMs. Firstly, latent variables can help to define new policy instruments, such as nudges. An example concerns social influence. If social comparison is an important determinant of mode choice (e.g., Kamargianni et al., 2015; Maness et al., 2015; Dugundji and Walker, 2005), [9] then the emphasis of a certain behavior performed by the majority of the social circle (neighbors, family, colleagues, and so on) such as the proposition "75% of the inhabitants of the town go to work by walking, cycling or using public transit" may encourage people to behave the same way to comply with social norms.

Secondly, latent variables help to better understand the impact of existing public policies. This can be illustrated with two examples. The first example took place in Germany, where the rail operator offered a discount card with a higher discount rate than before but decreased the flexibility of last minute booking. This pricing policy was followed by a drop in the train market share, which could have been avoided if the importance of the flexibility had been taken into account (Temme et al., 2007). The second example is in France, where the impact of the environmentally motivated bonus/malus fiscal system has a price effect (the less polluting car is subsidized and the polluting car is taxed), but also a social norm effect (the less polluting car is classified as pro-social and the polluting car as antisocial) (Hilton et al., 2014). Taking into account both effects, and not only the fiscal effect, helps to better forecast the effect of this policy.

Thirdly, latent variables may help to better measure the experienced utility. Indeed, in some contexts, behavioral biases, such as myopia or inattention, "generate differences between welfare from a policy maker's perspective, which depends on an agent's experienced utility (his actual well-being), and the agent's decision utility (the objective the agent maximizes when making choices)" (Chetty, 2015, p.2). These differences between decision and experienced utilities, discussed by Kahneman et al. (1997), Kahneman and Sugden (2005) and Kahneman and Thaler (2006), have welfare implications, since decision-makers make sub-optimal choices. Public policies may correct these internalities [10] but

9. None of these papers deals directly with social comparison as a determinant of mode choice within a latent variable framework, but they provide insight in this direction, which leaves room for future research.
10. See Herrnstein et al. (1993); Chetty (2015) for a definition.

need to measure the experienced utility, unobservable by nature, since even the decision-maker is not able to make an estimation. "Happiness" indicators are an efficient tool to achieve this goal.[11] For instance, Abou-Zeid and Ben-Akiva (2011) and Abou-Zeid et al. (2012) rely on happiness indicators to measure subjective well-being after a temporary change in the mode of travel to work, and therefore contribute to a wider understanding of the experienced utility.[12]

The use of latent variables is therefore a key step forward for the design of efficient public policies. The first two public policy implications discussed above involve identifying which latent variables enter the mode choice. Two interrelated strands of literature have incorporated latent variables to explain mode choice. Firstly, mode choice has been modeled by means of environmental psychology theories, based on the idea that behavior is driven by unobservable internal mental states, such as values, norms or attitudes (see Appendix A). The main theories used nowadays find their roots in theories developed in the 1970s, such as the norm activation model (Schwartz, 1977) or the theory of interpersonal behavior (Triandis, 1977). They have since been applied to explain car or public transport use (Hoffmann et al., 2017; Gifford et al., 2011). Qualitative or quantitative tools can be used to construct and test these theories.[13] Structural equation modeling, whose theoretical foundations were also established in the 1970s (Jöreskog, 1970; Hauser and Goldberger, 1971; Jöreskog, 1973), is a powerful tool to test and validate psychological theories, since it allows for relationships between observed and latent variables to be determined and quantified. SEMs are now seen as an efficient analytical tool in a variety of research contexts (transport, but also education, health and others) (Hoyle, 2012). For instance, using a SEM, Bamberg et al. (2003b) show that the components of the theory of planned behavior (attitudes, subjective norms, and perceived behavioral control) accurately predict intentions and use of public transport.

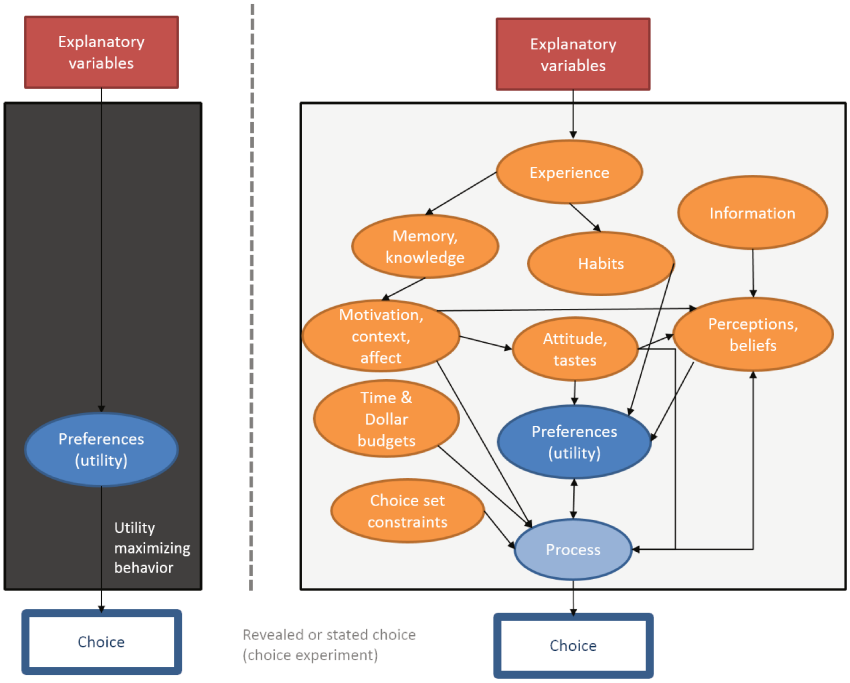
Secondly, apart from a few exceptions, DCMs have long considered only the observable characteristics of the alternatives or individuals and have failed to incorporate other relevant choice components. In the 1980s, Mc Fadden (1986) and Ben-Akiva and Boccara (1987) began to work on the

11. See Di Tella and MacCulloch (2006) for a literature review on the use of happiness indicators in economics.

12. Differences between experienced and decision utilities have also been analyzed by Frey and Stutzer (2014) in a transport context. Based on the observation that some intrinsic needs (human relationships, competence or autonomy, for example) are under-evaluated by comparison to extrinsic material needs (high income or brand car, for example), they show, based on a happiness indicator, that when people choose whether or not to accept a remote job offer, the increase in the commuting travel time is not fully compensated by the increase in income. Nevertheless, this incomplete compensation is not taken into account when the choice is made, since the increase in income is overweighted and the increase in the commuting time is underweighted.

13. Beirão and Cabral (2007) use interviews to analyze the influence of attitudes on the intensity of car use and public transport use. Anable (2005) divides users into six groups according to their attitude to the car, alternative modes to the car, the environment, and according to their environmental behavior (excluding mobility). Molin et al. (2016) conduct a latent class cluster analysis to identify (multi)modal groups from the frequency with which they use the different modes. Collins and Chambers (2005) use multiple regressions to explain the intention to use public transport on the basis of the model developed by Stern et al. (1999) and of situational variables, such as home-to-university cost or commuter time. Steg (2005) uses exploratory factor analyses to identify different motives for car use.

idea of combining DCMs with latent variable models to analyze individuals' behavior in more detail (see Raveau et al., 2012; Walker, 2001, for historical elements). This approach contrasts with the economic treatment of individuals' behavior, which regards the consumer as an "optimizing black box" (Morikawa et al., 2002, p.30). Fifteen years later, Ben-Akiva et al. (2002b) set the theoretical and empirical background of ICLV models, that is, "a general specification and estimation method [...], which provides complete flexibility in terms of the formulation of both the choice model and the latent variable model" (Walker, 2001, p.83). Figure 3.6 shows the gap between DCMs and the complexity of behaviors and represents how the integration of interrelated latent variables, which constitutes the core of behavioral theories, helps to bridge this gap. ICLV models are a means to operationalize the lessons from the environmental psychology theories in an economically sound framework. Indeed, by comparison with SEMs, ICLV models make it possible to explain and forecast mean individual choices, and therefore the demand, as well as to calculate economic outputs and simulate the effects of public policies. By comparison with DCMs with no latent variables, ICLV models allow for a better reflection of individual heterogeneity.



Source: Adapted from Walker (2001)

FIGURE 3.6 – The gap between basic Discrete Choice Models (left) and the Complexity of Behavior (right)

Despite the availability of this tool for fifteen years, the practice of modeling latent variables with a SEM is still not systematic among economists. Numerous authors directly integrate the indicators in

their models (e.g., Ramos et al., 2016) or a mean of indicators (e.g., Millock and Nauges, 2010). This practice has many disadvantages, since indicators are not causal, they are not available for forecasting, results are highly dependent on the phrasing of the survey question and multicollinearity is likely to be increased. Structural equation modeling makes it possible to explicitly account for measurement errors and obtain a closer correspondence between theory and empirics. As observed by Folmer and Johansson-Stenman (2011, p. 351), even if these principles are basic "in sociology and psychology, they still hardly play a role in current economics".

Motivation for a literature review on ICLV models

In the literature, "Integrated Choice and Latent Variable models" is the term most commonly used to denote models that combine a SEM and a DCM, but other terms, such as "hybrid choice models" have also been used.[14] Since the introduction of this term in J. Walker's PhD thesis (Walker, 2001), around 200 published, conference or working papers have used it, half of which have been published since 2014. ICLV models have mainly been used to model transport related behaviors: mode choice; road pricing (Di Ciommo et al., 2013); choice of electric vehicles (Daziano and Bolduc, 2013b; Jensen et al., 2014); route choice (Bekhor and Albert, 2014); departure time (Thorhauge et al., 2016); traffic information (Tsirimpa and Polydoropoulou, 2015); choice of airport and airline (Jung and Yoo, 2016); boat travel (Márquez et al., 2014), and so on. These models are now percolating outside the transport area and are being applied to environment related surveys that focus, for instance, on the landscape externalities of wind turbines (Mariel et al., 2015) or coastal water quality (Hess and Beharry-Borg, 2012).

ICLV models have been widely applied to mode choice (e.g., Johansson et al., 2006; Paulssen et al., 2014; Daziano and Rizzi, 2015). However, there is a high level of disparity in the way these models are applied, in terms of mathematical hypotheses but also in terms of the definition of variables, as well as the behavioral and economic level of interpretation. For instance, although economic indicators, such as value of time, are regularly derived from ICLV models, their economic interpretation and what they tell us about heterogeneity in behaviors, and therefore public policy design, is little discussed. There is therefore a need for a literature review to structure what has been done so far and suggest future avenues for research (see Chapter 4).

14. Hybrid choice models usually denote a broader framework in which ICLV models as well as latent class models are included (see Figure p.306 in Walker and Ben-Akiva, 2002).

Motivation for studying perceived comfort and indirect utility of travel

In the SEM and ICLV literature, the impact of some latent variables on mode choice has been widely studied (environmental concern, habits, and so on), while little attention has been given to others (feelings, perception of time use, and so on). In the two empirical chapters, the focus is on the perceived comfort and indirect utility of travel (Chapters 5 and 6) and on environmental concern (Chapter 6). The motivation for the focus on perceived comfort and indirect utility of travel is two-fold. The first motivation relies on the evidence from the literature, which suggests that mode choice is driven by instrumental motives (to reach a destination to perform activities), but also non-instrumental motives (activities conducted during the journey, and the activity of traveling itself). These concepts can be found in the economic literature (e.g., Mokhtarian and Salomon, 2001), in which the non-instrumental motives are referred to as the positive or indirect utility of travel, and in the psychological literature (e.g., Steg, 2005), in which the non-instrumental motives are referred to as affective and symbolic motives. The affective and symbolic motives for car use have been taken into account in the SEM (Lois and López-Sáez, 2009) and ICLV literature (Hurtubia et al., 2014; Fernández-Antolín et al., 2016). For public transport, the indirect utility is related to the concept of comfort. Indeed, Watts and Lyons (2010) show that in order to increase the productivity of passengers, both in terms of economically productive work and in terms of their travel time use, the creation of an environment with the necessary affect is needed. How people feel about, and engage with, the travel environment around them is crucial to productive as well as pleasurable travel time. However, as noticed by Lyons et al. (2013, p.1), the descriptive methods used by G. Lyons and his colleagues "cannot offer revised 'values of time' as alternatives for the process of economic appraisal". In that regard, Chapter 5 offers a new contribution to this strand of literature by offering values of time that take into account how the travel environment is perceived.

The second motivation relies on the lessons, on car use and public transport use, learned during the qualitative analysis conducted prior to the stated preferences survey. Among the persons interviewed, preferences for car use are mainly justified by instrumental motives, in particular flexibility, but affective and symbolic motives are also outlined. For instance, one of the persons interviewed, who had once used the train, felt embarrassed about having to ask friends to drive him between the train station and their house. He considered that using public transport, and thus not being autonomous, was demeaning, degrading. Others compare cars to pedigree dogs [15] and dream of having a luxurious

15. In the qualitative analysis prior to the stated preferences survey, the game of "Chinese portrait" has been used, according to which people were asked to compare different transport modes to an animal or an object, for instance (see Appendix B).

car. The car is also seen as a personal space, in which one is free to smoke, listen to loud music and have the windows open. Some people also use their car to cruise around, with no specific purpose other than to enjoy a ride with the family.

Train users speak about comfort in terms of objective characteristics: main line trains in the Rhône-Alpes Region are air-conditioned and quiet, and they provide travelers with tables, comfortable seats and legroom (unlike coaches), and the possibility of moving about (unlike cars). Trains are compared to elephants or beds to underline their comfort. The multitasking activities that can be performed on board is pointed out with comparisons to Swiss army knives or bags, which allow the traveler to carry everything necessary to travel (books, work, food, and so on) to avoid wasting time. The perception of digital tools and a Wi-Fi connection is heterogeneous among travelers. Some feel that they need this connection and criticize trains of the Rhône-Alpes region for not offering a continuous connection and causing them to waste time. Others are happy to have a time to relax, to read a book, or to work efficiently with no access to disturbing emails. One train user even described her travel time as an "entertaining" time, a "third half-time, a small space between work and home". [16] The eminently collective character of trains has been outlined as both an advantage and a drawback. Some travelers are unwilling to talk to their neighbor and even feel overwhelmed by the presence of other people, in terms of noise, odors or disrespectful behaviors. Others are happy to meet other travelers, and some commuters even have "train friends" with whom they have developed a long-term friendship. The relationship of train users with the outdoor environment is also specific. Being able to watch the landscape is important and some even describe their ideal train as one with outdoor decks. By contrast, car users who only occasionally use the train, for family, financial or material reasons, perceive it as a constraint, and do not feel comfortable when traveling by train. The constraint is all the more difficult to manage during travel with luggage or children. The choices of one couple interviewed illustrate this idea: although doing the same job with the same travel time by either train or car, the man commutes by train, while his wife commutes by car, the primary reason being the files they have to carry for their work. This example shows that two different choices may be made depending on the perception of constraints.

In short, the qualitative analysis combined with the literature review outlines comfort as a multidimensional attribute, which can be described with objective attributes (a seat or

16. Watts and Lyons (2010, p.8 and p.15) report similar findings. The comments made by two participants in their focus groups illustrate that idea: "I like to have the time. I don't want to teleport, 'cause otherwise you end up going from meeting to meeting to meeting, and you've not had time to think about the next meeting[...]" ; "my ideal office is on the train... especially if I've got things to work out [and I'm] at that point when you're just getting creative with it. It's my most creative place to work because it's that, you get a bit stumped, and you get to look out of the window for half an hour... see things...".

air-conditioning, for instance) and objective time use (to relax or read, for instance), but which also encompasses subjective perceptions of time use (a wasted time or, on the contrary, an efficiently used time), feelings (overwhelmed, relaxed) as well as the perception that traveling by train is easy or difficult. Therefore, mode choice may be well explained by a model incorporating objective and perceived measures of comfort.

Motivation for studying environmental concern

The motivation for the focus on environmental concern to explain mode choice habits is two-fold. Firstly, environmental concern is widely integrated in SEM and ICLV models. However, depending on how it is measured, and depending on the context of application, the conclusions on the impact of environmental concern on mode choice vary. Moreover, even when a positive relationship is found, environmental concern seems to be a weak determinant of mode choice. [17] Environmental concern has mainly been studied to explain either an intention (e.g., De Groot and Steg, 2007), which remains hypothetical, or a one-off choice (e.g., Paulssen et al., 2014; Bamberg and Schmidt, 2003), which is dependent on contextual factors (travel time, travel cost, and so on) and habits. Therefore, the effect of environmental concern on mode choice may be better understood by linking environmental concern to mode choice habits, which reflect a general and actual behavior. Secondly, as outlined above, mode choice is driven by the indirect utility of travel, and each traveler derives a different indirect utility from car or public transport use, which may be influenced by environmental concern. Indeed, the qualitative analysis outlined that people who refer to ecological reasons to justify their choice to use public transport modes are also the ones who seem to experience more positive feelings during travel by public transport or to appreciate the potential for work. Conversely, people who referred to affective and symbolic motives for car use and described bad experiences during travel by public transport, seem to refer less to the environmental advantages of public transport. These qualitative conclusions need to be confirmed by means of a quantitative model, and SEM is well designed for such a purpose.

17. For instance, according to Bamberg and Schmidt (2003), personal norms measured by awareness of the consequences of car use and ascription of responsibility for pollution and traffic, only explain a small percentage (14%) of car use, in the context of the norm-activation model, while intention, in the context of the theory of planned behavior, alone explains 45% of car use. Intention and habit together, in the context of the theory of interpersonal behavior, explain 51% of the behavioral variance.

Outline

Part II is divided into three chapters. Chapter 4 conducts a literature review on ICLV models concerned with mode choice. ICLV models are complex models involving four interrelated dimensions: econometric, data collection, behavioral and economic, which feed off each other. Each dimension is discussed theoretically and empirically to underline the good practices and highlight ICLV models' contribution to the enhancement of mode choice analysis, from a behavioral and an economic point of view.

Chapter 5 investigates how an ICLV model can be used to highlight behavioral patterns of mode choice, translate it into an economic output (value of time), and derive public policies. More specifically, this chapter investigates heterogeneity in value of time with respect to objective comfort, perceived behavior control, perceived use of time and feelings experienced during travel by public transport.

Chapter 6 explains car and public transport mode choice habits by means of a SEM exhibiting a mediation relationship. It studies how the impact of environmental concern on habits is - partially - mediated by the indirect utility of travel, measured as affective and symbolic motives for car use, as well as by perceived behavior control, feelings and perceived use of time during travel by public transport.

Discussion and limitations

The definition and the measurement of the latent variables used in ICLV models are highly heterogeneous, and thereby prevent comparisons and hamper efforts to systematize the lessons learned from the literature review. In an attempt to facilitate future work, the literature review of Chapter 4 presents different measurement methods and categorizes the latent variables used in ICLV models, according to the theme they address, their behavioral interpretation and whether they concern the individual in isolation or take into account her social context. Five main themes are identified: environmental concern, comfort, safety/security, convenience/accessibility, reliability and flexibility. Beyond the thematic characterization, three types [18] of latent variables are further identified:

1. Alternative-specific latent variables which reflect the decision-maker's perception of the supply.

18. Concerning the first two types, a similar categorization is found in Bhat and Dubey (2014, p. 77) who simulate five explanatory variables and say that "the first four are qualitative attributes specific to two modes (air and bus), while the last variable is an individual-specific qualitative attribute (that does not vary across modes)". Bahamonde-Birke et al. (2017) also discuss the first two types and show that perceptions may affect the decision making process and that they are able to capture a significant part of the variability that is normally explained by alternative specific constants. They also show that attitudes may be a reason for systematic taste variations.

These variables may be more important determinants of the demand than the objective characteristics of the supply;

2. Latent variables describing the individual in terms of norms, attitudes and preferences. These variables are the internal mental states described in the environmental psychology theories;
3. Latent variables describing a specific behavior (for example, recycling), generally in a non-transport context. However, pro-environmental behaviors in the transportation field are generally weakly correlated with other kinds of environmentally friendly behavior. Unless the objective is to underline such a connection, the use of these variables should therefore be avoided.

Because of the inclusion of these latent variables, ICLV models are supposed to "allow for more behaviorally realistic representations of the choice process, and consequently a better understanding of behavior, improvements in forecasts, and valuable information regarding the validity of simpler model structures" (Walker, 2001, p.12). Regarding previous applications of ICLV models, as well as the one presented in Chapter 5, they theoretically fulfill their promises by means of a robust mathematical formulation and integrated estimation methods, well served by improvements in software capabilities. However, our reading of the literature is that there is still some way to go to fully exploit the behavioral and economic potentialities of ICLV models.

One explanation is that most papers using ICLV models have focused on econometric issues, either to improve (Daziano, 2015; Bhat and Dubey, 2014) or compare (Raveau et al., 2010) the estimation methods, discuss identification (Vij and Walker, 2014) or endogeneity issues (Fernández-Antolín et al., 2016), or to compare the performance of different specifications (Yanez et al., 2010). Only a few published papers or conference papers (Johansson et al., 2006; Paulssen et al., 2014; Maldonado-Hinarejos et al., 2014; Daziano and Rizzi, 2015; Scagnolarie et al., 2015; Idris et al., 2015; Kamargianni et al., 2015; Sottile et al., 2014, 2013) primarily discuss behavioral, economic or public policy implications in a mode choice context. Although not fully exploited, the economic and behavioral potentialities of ICLV models are promising. Firstly, since the choice process remains based on random utility maximization, ICLV models can be used to derive economic outputs and explain how these economic outputs depend on latent variables. Secondly, even if ICLV models generally do not allow for an improvement in the prediction of outcomes (Vij and Walker, 2016), they can be useful for explanatory purposes.

Now that ICLV models are mathematically and computationally well defined, the behavioral and economic issues will be easier to address. A preliminary step is to develop databases specifically

designed to explore these issues in the context of mode choice modeling, which would be facilitated by a closer collaboration between discrete choice modelers and behavioral scientists. Indeed, in many works, the limitations of the available data precluded the possibility of more elaborate behavioral models. Consequently, the SEMs of the ICLV models do not usually include structural relationships between latent variables, although it is a key feature of behavioral models and SEMs are well designed to model the direct, mediated or moderated relationships that are found in these models (Hayes, 2013b). Only then will it be possible to take advantage of the complementary approaches of the economic and behavioral theories for representing the choice process and mode choice modeling. Collaboration with behavioral scientists may also be fruitful to support the definition of latent variables with an appropriate scale, since the choice of indicators to measure latent variables is a broad area of research in psychology and management research and plays an important role in the development of ICLV models. Additionally, it would be interesting to enlarge the application context with surveys collecting data from the USA and from European countries (France, Spain, and so on). Indeed, since several of the articles reviewed use the same datasets, there is a concentration of applications in some countries (Chile or Switzerland, for instance).

The rest of the dissertation does not claim to answer the broad question of a deeper integration between environmental psychology theories and DCMs. This question is left for future research. Chapter 5 focuses on an original feature of ICLV models: the potential to analyze the heterogeneity of economic outputs, and in particular willingness-to-pay indicators, according to behavioral determinants. This work illustrates how ICLV models can be used to study economic issues and derive public economic policies by means of a better understanding of behaviors, in comparison to standard DCMs. By means of a quantitative measure of the interaction between time and comfort, the estimated models show that during travel by public transport modes, value of time decreases with objective and perceived comfort. When a seat is guaranteed in trains, the value of time is lower than the value of time by coach or by train with no guarantee of a seat. Value of time by car is in-between. People who perceive the use of coach or train as easy are more likely to use these public transport modes. However, once on board, perceived behavioral control does not influence their value of time. People who perceive their travel time as useful or who experience positive feelings during travel by public transport are more likely to choose these transport modes and are also ready to spend more time traveling. When the three measures of perceived comfort are included in the same model, the robust results are that perceived behavioral control regarding the use of public transport modes plays a role in mode choice, while objective comfort in trains (guarantee of a seat) and experiencing positive

feelings during travel by public transport decrease value of time.

Scenarios are constructed to simulate how changes in perception of time use and feelings impact value of time. The simulated values for the latent variables are chosen within the range of observed levels, such that the values are actually reachable and refer to plausible scenarios. For instance, if communication campaigns, accompanied by appropriate infrastructure, manage to raise positive feelings to the higher level observed in the sample, then value in public transport would substantially decrease.

The importance of perceived comfort for mode choice is confirmed by Chapter 6 in a broader perspective by means of a SEM. An economics-related concept, indirect utility, is used to encompass the affective and symbolic motives for car use as well as the perceived comfort of public transport. The estimated models [19] show that positive perceptions and feelings during travel foster public transport habits. People with high environmental concern have a better perception of the utility of public transport and are more likely to use these transport modes. Conversely, people with low environmental concern have more affective and symbolic motives for car use. However, the impact of affective and symbolic motives for car use has not been robustly proven. Such a model could theoretically be integrated in a DCM using new estimation methods, such as the ones proposed by Bhat and Dubey (2014) and Daziano (2015) to avoid the curse of dimensionality. However, particular attention should still be paid to the balance between the complexity of the models and what the data can really tell us. Furthermore, the estimated models assume that attitudes influence behavior, which is fully consistent with most leading psychological theories. This direction of causality is nonetheless challenged by a recently published paper by Kroesen et al. (2017), who show that the reversed direction of causality from behavior towards attitudes is stronger. What they argue is likely to play a role in this study as well: cognitive dissonance predicts that people whose attitudes are not consistent with their behavior are more likely to change their attitudes than their behavior. Although both relationships may be true, causality cannot be demonstrated with the available data, and this is one of the limitations of the model and, more broadly, of the data. In the same way, the finding that living in an urban area significantly increases public transport habits and reduces car use habits, may be explained by self-selection. People who wish to travel by public transport are more likely to choose a residence in an urban area. Panel data, or data distinguishing between treated and untreated people (for instance, in a natural or a laboratory experiment), would be necessary to establish a causal

19. Prior to this work, different estimation methods (Weighted Least Squares with Mean and Variance correction, Maximum Likelihood and Maximum Likelihood with robust standard errors) along with different methods for missing data (Full information maximum likelihood and Multiple imputation methods) have been tested and compared. The results are not presented here, as they still need further analysis.

relationship. In conclusion, the estimated models convey causal assumptions based on previous theoretical models, and on the qualitative analysis, but do not produce validated causal conclusions.

To form new habits and induce a modal shift from car to public transport, several tools may be used. Gärling and Schuitema (2007) show that non-coercive measures alone are not efficient and that coercive measures, such as increasing cost, are subject to public opposition. However, combined with non-coercive measures providing attractive public transport and communicating the benefits of car-use reduction to the public, coercive measures are likely to become more effective and socially acceptable. To provide attractive public transport, the results of the SEM and ICLV models show that public policies must not only rely on the reduction of travel time or travel cost, but also offer a comfortable environment to allow for positive experiences and a polychronic use of time. Another way to enhance the social acceptability of pricing policies is to use reward schemes, [20] since reward seems to be at least as effective as punishment in terms of its effectiveness in influencing behavior (Balliet et al., 2011; Khademi and Timmermans, 2014). Such a scheme has been implemented in the Netherlands (Ben-Elia and Ettema, 2009; Tillema et al., 2013) to avoid peak hour driving. The results show that this policy has long-term effects to encourage off-peak driving, but also teleworking or changing route options (Khademi and Timmermans, 2014). A comparable reward scheme could encourage people to avoid using their car in general, or, more specifically, in the case of a peak pollution event. To accompany such a measure, information on the health consequences of pollution could be effective. According to the SEM model estimated in Chapter 5, if people were more aware of the consequences of pollution, they would be more aware of the utility of choosing public transport and be more inclined to use it. The reward system would act as a trigger to change habits.

Concerning travel time, the usual perspective is to consider that less time spent traveling is assumed to convert "unproductive" time into economically valuable time. An alternative perspective, supported by the research in Chapters 5 and 6 but also previous authors (Lyons et al., 2007; Berliner et al., 2015; Malokin et al., 2017), is that, rather than being wasted, travel time can and does possess an indirect utility. The current economic appraisal of the value of time should therefore take into consideration the quality component of travel time and not only its quantitative component. [21] For instance, since the quality component of travel time has been shown to be higher in trains with the guarantee of a seat than in coaches, then the replacement of train lines by coaches may induce a shift to car because coaches do not offer as good conditions as trains to perform activities on board. Furthermore, as shown by the

20. Two questions remain open on the application of reward schemes: how should the system be financed? How is it perceived by citizens who do not in any case use their car?

21. Differenciations of the VOT by mode or purpose is a way of achieving such a goal.

aforemetioned qualitative analysis and Watts and Lyons (2010), travel time is not only a time that can be considered as an (almost) ordinary activity time, both being potential substitutes for each other. For some people at least, travel time is more than that and has its own specificity; being disconnected from the demands of work or children, this time is unique, efficient and desirable as such. The "travel remedy kit" imagined by Watts and Lyons (2010, p.18), contains suggestions, guides, and maps to transform "potentially bored and inactive passengers, into equipped and alert wayfarers", empowering them to engage in novel interactions with the vehicle, different activities, and suggesting ways of making travel time more fluid. An interesting avenue for future research would be to make a comparable intervention into travel time and quantitatively measure how it changes value of time. Indeed, as noticed by Watts and Lyons (2010), such an intervention is beneficial to passengers but of low cost to the transport operator.

CHAPTER 4

Integrated choice and latent variable models: a literature review on mode choice[1]

Abstract

Mode choice depends on observable characteristics of the transport modes and of the decision maker, but also on unobservable characteristics, known as latent variables. By means of an Integrated Choice and Latent Variable (ICLV) model, which is a combination of structural equation model and discrete choice model, it is theoretically possible to integrate both types of variables in a psychologically and economically sound mode choice model. To achieve such a goal requires clear positioning on the four dimensions covered by ICLV models: survey methods, econometrics, psychology and economics. This paper presents a comprehensive survey of the ICLV literature applied to mode choice modelling. I review how latent variables are measured and incorporated in the ICLV models, how they contribute to explaining mode choice and how they are used to derive economic outputs. The main results are: 1) the latent variables used to explain mode choice are linked to individual mental states, perceptions of transport modes, or an actual performed behaviour; 2) the richness of structural equation models still needs to be explored to fully embody the psychological theories explaining mode choice; 3) the integration of latent variables helps to improve our understanding of mode choice and to adapt public policies.

1. This chapter is a paper that has been submitted to *Transport Reviews* and is currently under review.

4.1 Introduction

Mode choice is a multiparametric decision process determined by the characteristics of the transport alternatives and a combination of individual socio-economic, psychological and spatial factors. Psychological factors are unobservable, latent, variables and remain among the less studied factors in discrete choice models (DCMs), even though, when integrated, they often prove to be significant (De Witte et al., 2013). At the same time, psychological theories propose a behaviourally rich representation of mode choice by modelling relationships between latent variables, such as norms, values or attitudes. Structural equation models (SEMs) are a powerful tool to model these relationships. However, since they are not based on economic theory, SEMs cannot be used in transport models or to derive economic outputs, such as value of time or elasticities, necessary for project appraisal. One solution is to incorporate behavioural theories in discrete choice models by means of an Integrated Choice and Latent Variable (ICLV) model, which theoretically allows the analyst to benefit from the economic and behavioural foundations of both approaches.

ICLV models involve, at the same time, survey theory, econometric theory, psychological theory and economic theory, and the promise of ICLV models is to embrace and integrate these multiple dimensions in an integrative model. Previous work has provided insight into some specific dimensions; for instance, Vij and Walker (2014) focus on identification issues, Bhat and Dubey (2014) on estimation methods, and Vij and Walker (2016) on the operational use of ICLV models. However, there has been no attempt to address the multidimensional nature of ICLV models or to combine theory and practice for each dimension with a synthesis and analysis of the results produced by ICLV models. This paper intends to fill that gap.

The aim of this paper is to provide, for the survey, econometric, psychological and economic dimensions of ICLV models, a comprehensive review of the literature on how mode choice and the determinants of mode choice can be better modelled. All dimensions involved in ICLV models are addressed to emphasize and structure the progress that has been made in the literature in recent years, as well as to identify future research needs. Specifically, the contribution of this paper to the current knowledge on mode choice is three-fold. Firstly, to my knowledge, this paper is the first to review how latent variables that explain mode choice are measured and to propose the use of psychological and management science methods to improve this measurement. Secondly, to my knowledge, this paper is the first to review the different structures of SEMs that can be incorporated in ICLV models to improve the behavioural representation of mode choice. Thirdly, to my knowledge, this paper is (one of) the

first to propose an analysis grid to structure the latent variables used in ICLV models and to review how these latent variables explain mode choice.

The rest of the paper is organized in five sections. Section 4.2 presents the four dimensions of ICLV models and the search strategy. Section 4.3 is dedicated to the survey dimension and Section 4.4 to the tool itself with the econometric dimension. Section 4.5 discusses the psychological dimension of ICLV models with a specific focus on latent variables. Section 4.6 reports on the use made of ICLV models and, in particular, the economic dimension. Section 4.7 is the conclusion.

4.2 Methodology

4.2.1 Overview of the dimensions addressed in ICLV models

To develop an ICLV model requires clear positioning on four dimensions. Firstly, the data collection process has to be adapted to collect the indicators measuring the latent variables. Special attention also needs to be paid to the sample size, since estimated parameters are more numerous and may result in identification issues.

Secondly, ICLV models require an estimation of a SEM and a DCM. Depending on the papers, different terminologies, such as latent variable model, MIMIC model [2] or LISREL models, [3] are used to refer to the SEM. In this paper, the terminology "SEM" is used intentionally to reflect the richness of the models that can be included in ICLV models and to build the bridge with psychological and management science models. Structural equation modelling is a methodology for representing, estimating, and testing a network of relationships between variables (measured or latent variables, exogenous or endogenous variables). They are able to represent a wide variety of relationships between variables, in particular, mediation and moderation processes (Hayes, 2013b).

Thirdly, the latent variables included in ICLV models are drawn from psychological theories, which describe the cognitive mechanisms resulting in travel mode choice. These latent variables and theories are various and their integration in the deterministic part of the utility is supposed to improve the behavioural realism of mode choice models.

Fourthly, ICLV models are modelled using economic random utility theory. The decision rule, that is, choosing the option that maximizes the utility, remains the same as in the other DCMs. Because of their economically sound formulation, ICLV models can be used to derive economic policies or economic outputs, such as elasticities, willingness-to-pay indicators and market shares.

2. Multiple Indicators Multiple Causes (MIMIC) models are specific SEMs.
3. LInear Structural RELations (LISREL) is a statistical software package used for structural equation modelling.

4.2.2 Search strategy

The search for papers was conducted with the aim of being as exhaustive as possible as regards ICLV models on mode choice. The sources used to search for literature are web-based search tools (Web of Science, Researchgate, Google Scholar and the database of the French National Center for Scientific Research). After this first step, cited references were tracked to create a domino effect. The search was conducted in August and September 2016 using the search terms ("mode choice" or "modal choice") and ("integrated choice and latent variable(s)" or "hybrid choice model(s)" or 'hybrid modelling approach" or "hybrid choice modelling"). The resulting output was filtered with a focus on papers containing an application of ICLV models (with real or simulated data) concerned with individual inland mode choice. [4] Freight, boat and air applications have been excluded.

Eventually, 35 papers were retained for further analysis and processed for this review. They were published or presented in conferences between 1998 and 2016 but 50% of them were published after 2013. All papers contain an application but three use only simulated data (Bhat and Dubey, 2014; Daziano and Bolduc, 2013a; Raveau et al., 2012). These papers have been retained because they contain valuable insight into the method.

4.3 Survey dimension

This section provides an overview of the surveys used in ICLV models before focusing on the collection of data to measure the latent variables.

4.3.1 Overview of the surveys

Two thirds of the surveys used for ICLV models are revealed preferences surveys, the remaining third being stated preferences surveys. Some datasets are common to several studies. For instance, Glerum et al. (2014), Atasoy et al. (2013) and Fernández-Antolín et al. (2016) all use the same Swiss survey. Yanez et al. (2010) and Raveau et al. (2010) use the same Chilean survey. Datasets are heterogeneous in terms of the number of persons interviewed (see Figure 4.1). The median is around 800 with a mean around 2,400 due to two large revealed preferences surveys (La Paix Puello and Geurs, 2015; Roberts et al., 2014), which include, respectively, 12,000 and 13,141 observations.

All datasets were collected in western countries (Canada, Chile, Europe), with the exception of one

4. In one paper, mode choice is made conjointly with purchase intentions of alternative fuel vehicles (Daziano, 2015). Another paper models mode shift defined as the choice of departing from the car mode (Idris et al., 2015).

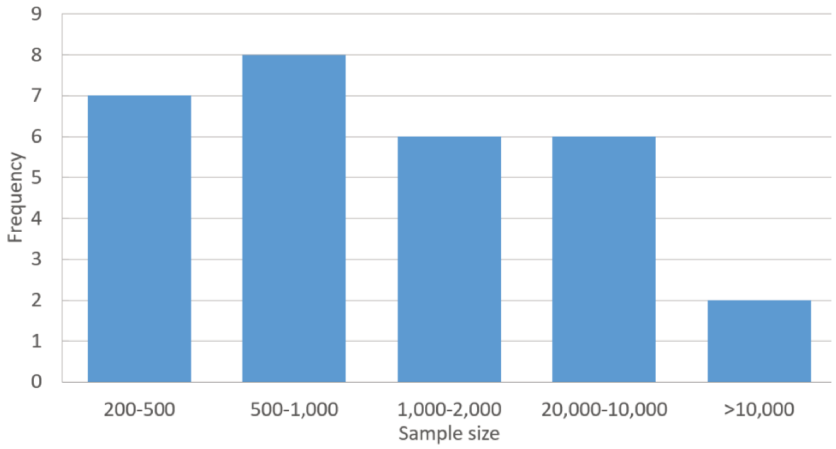


FIGURE 4.1 – Distribution of the size of samples (persons interviewed) used for ICLV models on mode choice (excluding simulated and duplicated datasets)

collected in China (Walker et al., 2013). Some studies collected data only for one city (Habib and Zaman, 2012) or a specific inter-urban corridor (Daziano and Rizzi, 2015; Johansson et al., 2006). Some other studies concern all cities in a country (Daziano, 2015) or the whole country (Roberts et al., 2014). Some surveys specifically restrict their field of study to employees of a university (e.g., Tudela et al., 2011; Córdoba and Jaramillo, 2012), which limits the generalization of their results to the whole population.

With regard to journey purposes, four situations emerge from the review: 1) all purposes are considered (e.g., Glerum et al., 2014; Yanez et al., 2010); 2) the focus is on work trips (e.g., Idris et al., 2015; Daziano, 2015; Johansson et al., 2006); 3) only trips to school or university are considered (e.g., Kamargianni et al., 2015), and 4) only trips to go out with friends are considered (Scagnolarie et al., 2015). Depending on the purpose, respondents to the surveys are also naturally heterogeneous in terms of age.

The selection of respondents can also be made according to their (potential) use of specific modes. For instance, Glerum et al. (2014) only interviewed people living in towns or villages connected to the postal buses. Sottile et al. (2013) restricted participation in their survey to the (potential) users of the park and ride facility they study. The car plays a key role in choice sets, whether or not a distinction is made between car as driver and car as passenger. Apart from Daziano and Rizzi (2015), all choice sets include the car alternative, either as a specific mode or as part of an aggregate. The car is often chosen as the reference alternative. Other modes considered are:

- private motorized modes, such as carpool (Daziano, 2015; Idris et al., 2015) or powered two wheelers / motorcyles (Kamargianni et al., 2015);
- public transport modes, such as park and ride (Sottile et al., 2013), bus or trains (Anwar et al.,

2014);

- active modes: walking and cycling (Idris et al., 2015; Daziano, 2015).

The papers reviewed on mode choice therefore account for a wide variety of contexts and represent what has been done so far, in terms of survey method, geographic area and transport modes.

4.3.2 Psychometrics: theory and application

Psychometrics is the field of study concerned with the theory and technique of psychological measurement. Measurements are made by constructing a scale that assigns numbers to objects or events according to rules (Tinsley and Brown, 2000). The typical scale consists of a set of stimuli (items) designed to elicit particular kinds of responses (the indicators used in the SEM), a set of response alternatives and a set of directions on how to respond to items. Over time, different types of scales have been described in the literature (Zikmund et al., 2013), among them multiple item scales.

An overwhelming majority of studies use four- to ten- point Likert scales (Likert, 1932) to measure latent variables. Anchors of the scale can denote agreement (e.g. Kamargianni et al., 2014), frequency (Sottile et al., 2014), similarity (Temme et al., 2007), importance (Daziano, 2015), or good/bad representation (Morikawa and Sasaki, 1998). If the rating scale is anchored at each end by antonym adjectives, then they are named semantic differential scales (Osgood and Luria, 1954). To measure attitudes towards alcohol consumption, Scagnolarie et al. (2015) use a seven-point semantic differential scale with adjectives, such as pleasant/unpleasant, funny/boring.

Two studies (Idris et al., 2015; Tudela et al., 2011) use the Verplanken scale to measure habits (Verplanken et al., 1994). This scale measures habitual behaviour intensity by counting the number of times a given behaviour would be repeated under specific circumstances without too much reasoning. For instance, for non-work activities, Idris et al. (2015) ask the respondent to provide the mode she would use and then count how many times she had mentioned each mode to perform different activities.

Besides structured scales with closed questions, spontaneous answers to open questions can be used to construct scales. A common question in French travel surveys is to ask for three spontaneous adjectives best describing a transport mode. Numeric data, and therefore scales, can be generated from theses words (e.g. Kaufmann et al., 2010). Although they do not mention it explicitly in their study, Glerum et al. (2014) use the method proposed by Thurstone (1927) to construct a scale based on adjectives. Using a Swiss survey, they asked 25 evaluators to rate the adjectives related to comfort on public transport. A positive number is associated with a positive representation of the latent variable

and a negative number implies a negative representation. The scale is a five-point scale from -2 to 2.

The technique to define an appropriate scale to measure latent variables is a broad area of research in psychology and management research (John and Benet-Martínez, 2000; Dunlap et al., 2000). If there is no adequate or appropriate existing scale to measure a latent variable, then a systematic seven-step process has to be followed (Hinkin et al., 1997): item generation; test for conceptual consistency; administration; purification of the measure with dimension reduction techniques, such as explanatory factor analysis, and confirmatory factor analysis; internal consistency assessment with reliability indicators (Cronbach α, Joreskog ρ); construct validity (discriminatory and convergent), and replication with a new dataset. Failure to carefully develop and validate the measurement scale can result in invalid and unintegratable data (Hinkin et al., 1997).

With the exception of the Verplanken scales measuring habits, to our knowledge, the scales used in ICLV studies on mode choice are either constructed by the authors or the authors use a part, a compilation or an adaptation of existing scales. For instance, Paulssen et al. (2014) measured respondents' value orientations based on the portraits value questionnaire from Schwartz et al. (2001). This questionnaire originally identifies ten motivationally distinct value constructs but the authors restricted their attention to three: power, hedonism and security. To measure variety-seeking, Rieser-Schüssler and Axhausen (2012) construct a scale that includes some of the items in Niininen et al. (2004) and Mokhtarian and Salomon (2001). Sottile et al. (2014) worked with a team of environmental psychologists to measure all items included in the Theory of Planned Behaviour.

Among the tools designed to construct and validate scales, three are used in ICLV models on mode choice. Some studies report the use of principal component analysis (Sottile et al., 2014) or exploratory factor analysis (e.g. Johansson et al., 2006; Scagnolarie et al., 2015; La Paix Puello and Geurs, 2015) for dimension reduction. Confirmatory factor analysis, that is, the measurement model of the SEM, is reported by Johansson et al. (2006), Tudela et al. (2011) and Roberts et al. (2014). Concerning reliability, Johansson et al. (2006), Scagnolarie et al. (2015) and Politis et al. (2012) use Chronbach's α to check reliability.

4.4 Econometric dimension

This section describes the ICLV tool and reviews the general formulation of ICLV models, before describing some components of the model. It also discusses the identification issue and the methods used to estimate ICLV models.

4.4.1 General formulation of ICLV models

This section describes the general specification of ICLV models, before clarifying some hypotheses that can be conducted on error terms, the measurement model of the SEM and the interaction terms in the structural equations.

4.4.1.1 General specification

Mathematically, ICLV models consist of two components: a DCM (Equations (4.1) and (4.4)) and a SEM (Equations (4.2) and (4.3)):

$$\mathbf{U_n} = \mathbf{B}\mathbf{x_n} + \mathbf{\Gamma}\boldsymbol{\xi_n} + \boldsymbol{\varepsilon_n}, \tag{4.1}$$

$$\boldsymbol{\xi_n} = \mathbf{A}\mathbf{x_n} + \boldsymbol{\eta_n}, \tag{4.2}$$

$$\mathbf{z_n^\star} = \mathbf{D}\boldsymbol{\xi_n} + \mathbf{v_n}, \tag{4.3}$$

$$c_{n,j} = \begin{cases} 1 & \text{iif } U_{n,j} \geq U_{n,j'} \text{ for } j' \in \{1, \cdots, J\}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

Equation (4.1) is the structural equation of the DCM. $\mathbf{U_n} = (U_{n,1}, \cdots, U_{n,J})$ is the $(J \times 1)$ vector containing the utilities of the alternatives faced by each individual n ($n = 1, \cdots, N$). $\mathbf{x_n}$ is the $(K \times 1)$ vector of individual and alternative-specific observable explanatory variables. The Q latent explanatory variables are in the $(Q \times 1)$ vector $\boldsymbol{\xi_n}$. $\mathbf{B}$ is the $(J \times K)$ matrix with the parameters of the observable variables. Some terms may be equal to zero if the set of observable variables $\mathbf{x_n}$ is not the same in Equation (4.1) and in Equation (4.2). $\mathbf{\Gamma}$ is the $(J \times Q)$ matrix with the parameters of the latent variables. $\boldsymbol{\varepsilon_n}$ stands for the $(J \times 1)$ vector composed of error terms.

Equation (4.2) is the structural equation of the SEM. $\mathbf{A}$ is the $(Q \times K)$ matrix with the parameters denoting the effects of observable variables on latent variables. Some terms may be equal to zero if the set of observable variables $\mathbf{x_n}$ is different in Equation (4.1) and (4.2). $\boldsymbol{\eta_n}$ is the $(Q \times 1)$ vector composed of Q independently and identically distributed (i.i.d) normal distributed error terms with $\mathbf{\Phi}$ the covariance matrix.

Equation (4.3) is the measurement equation of the SEM. $\mathbf{z_n^\star}$ is the $(P \times 1)$ indicators' vector used to measure the latent variables. They are assumed to be continuous and normally distributed. $\mathbf{D}$ is a $(P \times Q)$ matrix with the parameters associated with the latent variables. $\mathbf{v_n}$ is the $(P \times 1)$ vector composed of P i.i.d. normal distributed error terms with $\mathbf{\Psi}$ the covariance matrix. The error terms $\boldsymbol{\varepsilon_n}$, $\boldsymbol{\eta_n}$ and $\mathbf{v_n}$ are considered as mutually independent.

Equation (4.4) is the measurement equation of the discrete choice part of the model. $c_{n,j}$ is the choice indicator, which is equal to one if alternative j is chosen and zero otherwise.

The model is sometimes written in its reduced form, by integrating Equation (4.2) into Equation (4.1) :

$$\begin{aligned} \mathbf{U_n} &= \mathbf{B}\mathbf{x_n} + \mathbf{\Gamma}(\mathbf{A}\mathbf{x_n} + \boldsymbol{\eta_n}) + \boldsymbol{\varepsilon_n} \\ &= (\mathbf{B} + \mathbf{\Gamma}\mathbf{A})\mathbf{x_n} + \mathbf{\Gamma}\boldsymbol{\eta_n} + \boldsymbol{\varepsilon_n}. \end{aligned} \tag{4.5}$$

The joint density of $\mathbf{c_n} = (c_{n,1}, \cdots, c_{n,j}, \cdots, c_{n,J})$ and $\mathbf{z_n^\star} = (z_{n,1}^\star, \cdots, z_{n,p}^\star, \cdots, z_{n,P}^\star)$ is

$$f_{c,z^\star} = \int_\xi f_c(\mathbf{c_n}|\mathbf{x_n}, \boldsymbol{\xi_n}; \mathbf{B}, \mathbf{\Gamma}) f_{z^\star}(\mathbf{z_n^\star}|\boldsymbol{\xi_n}; \mathbf{D}, \mathbf{\Psi}) f_\xi(\boldsymbol{\xi_n}|\mathbf{x_n}; \mathbf{A}, \mathbf{\Phi}) d\xi, \tag{4.6}$$

with f_c the probability distribution function associated with the choices, $f_{z^\star}$ the probability distribution function assoxiated with the measurement indicators and f_ξ the probability distribution function associated with the latent variables. The expressions of the functions f_c, $f_{z^\star}$ and f_ξ depend on the hypotheses made on the associated error terms [5] $\boldsymbol{\varepsilon_n}$, $\boldsymbol{\eta_n}$ and $\mathbf{v_n}$.

4.4.1.2 Distribution of the error terms

Although not always explicitly mentioned in the papers reviewed, a common choice is to define the elements of $\boldsymbol{\varepsilon_n}$ as i.i.d. Gumbel distributed across alternatives and individuals (Raveau et al., 2012; Atasoy et al., 2013). This hypothesis leads to the well-known multinomial logit kernel. Alternative distributions lead to mixed logit (Scagnolarie et al., 2015; Maldonado-Hinarejos et al., 2014; Idris et al., 2015; Anwar et al., 2014; Sottile et al., 2014; Tudela et al., 2011; Yanez et al., 2010), cross-nested logit (Paulssen et al., 2014; Habib and Zaman, 2012; Habib et al., 2011), and probit models (Daziano, 2015; Kamargianni et al., 2015; Bhat and Dubey, 2014; Roberts et al., 2014; Politis et al., 2012; Johansson et al., 2006). The mixed logit specification makes it possible to associate random coefficients with both observable and latent variables, whose effect on mode choice may be particularly heterogeneous in the population. For instance, Yanez et al. (2010) estimate three models, a mixed logit without latent variables, a mixed ICLV model with a random parameter associated with cost and a mixed ICLV model with random parameters associated with cost and the two latent variables (accessibility and comfort/safety). The last in the list proved to be the best in terms of estimation and forecasting (market shares reproduction). Tudela et al. (2011) associate random parameters with habits, attitudes and affective latent variables and estimate two ICLV models including, or not, the affective variable. They note that when all three latent variables are integrated, the standard deviation of the random

5. See, for instance, Chapter 5.

parameters is smaller, which means that the mean value is more certain and randomness is captured by the affective factor.

Regarding probit specifications, two reasons may explain this choice. Firstly, some studies (e.g. Roberts et al., 2014; Politis et al., 2012) use software originally developed for SEM estimations (M+, AMOS), which usually support only normal error terms. Secondly, recent papers (Daziano, 2015; Kamargianni et al., 2015; Bhat and Dubey, 2014) use normally distributed error terms in order to 1) reduce the estimation time with alternative estimation methods (Bayesian estimator or maximum approximate composite marginal likelihood), and 2) enable more flexible substitution patterns across alternatives by specifying a general covariance matrix (Kamargianni et al., 2015).

4.4.1.3 Ordered probit in the measurement model

The indicators measuring the latent variables (Equation (5.5)) are almost exclusively measured with scales leading to ordinary variables. Nevertheless, the ordinal nature of the indicators is usually overlooked and they are considered as continuously normally distributed (e.g. Idris et al., 2015; Scagnolarie et al., 2015; Glerum et al., 2014; Paulssen et al., 2014; Habib and Zaman, 2012). The usual framework in the SEM literature (Bollen and Hoyle, 2012) is to add a threshold model equation to the measurement model of the SEM. In Equation (4.7), $z^{\star}_{n,p}$ is a latent variable reflecting the "true" value accorded to the evaluated statement and $z_{n,p}$ is its discrete translation

$$z_{n,p} = \begin{cases} 1 & \text{iif} \quad & z^{\star}_{n,p} & \leq \bar{z}_{1,p} \\ 2 & \text{iif} \quad \bar{z}_{1,p} < & z^{\star}_{n,p} & \leq \bar{z}_{2,p} \\ \dots & & & \\ L_p & \text{iif} \quad \bar{z}_{L_p-1,p} < & z^{\star}_{n,p}. & \end{cases} \tag{4.7}$$

L_p is the total number of categories of $z_{n,p}$ and the $\bar{z}$'s parameters are thresholds or cutoff points for $z^{\star}_{n,p}$ that determine the probabilities of observing each category of $z_{n,p}$, with $\bar{z}_{1,p} \leq \bar{z}_{2,p} \leq \cdots \bar{z}_{L_p-1,p}$. The probability that the indicator takes the value l is therefore equal to

$$\Pr(z_{n,p} = l) = P(\bar{z}_{l-1,p} < z^{\star}_{n,p} \leq \bar{z}_{l,p}).$$

Assuming that the latent response variables are normally distributed, the corresponding model is an ordered probit model. Only a few studies applying ICLV models to mode choice use this framework (Abou-Zeid and Ben-Akiva, 2011; Kamargianni et al., 2015; Sottile et al., 2013).

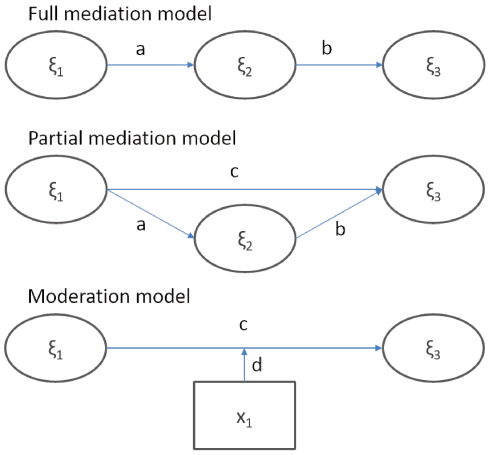
4.4.1.4 Interaction terms in structural equations

Interaction terms between latent variables or between latent and observed variables can be included in the two structural sub-models. In Equation (4.1), the utility may include cross variables to study how latent variables impact the effect of an observed variable. For instance, Fernández-Antolín et al. (2016) define a cross-variable time × car loving attitude and study how value of time varies according to attitude.

In Equation (4.2), latent variables do not interact with each other and depend only on observed variables, as is generally the case in the reviewed ICLV models. However, most psychological and management science theories do involve interactions between latent variables, which requires Equation (4.2) to be changed to Equation (4.8):

$$\boldsymbol{\xi_n} = \mathbf{A}\mathbf{x_n} + \mathbf{A}'\boldsymbol{\xi_n} + \boldsymbol{\eta_n}. \tag{4.8}$$

$\mathbf{A}'$ denotes the $(Q \times Q)$ matrix of parameters translating the effect of some latent variables on other latent variables. The diagonal parameters must be fixed to 0. All structures of SEMs may be estimated with this specification, in particular the full mediation, partial mediation and moderation models (see Figure 4.2 and corresponding matrix), which are extensively used in behavioural theories. The reader is referred to Little et al. (2007) and Hayes (2013b) for more detail on these models. In these three models, there is no feedback between latent variables, $\mathbf{A}'$ is thus a strictly lower triangular matrix. If feedbacks were allowed, then only off-diagonal elements would be non-zero. This could be the case, for instance, in a study of social interactions within networks.



$$\mathbf{A}' = \begin{pmatrix} 0 & 0 & 0 \\ a & 0 & 0 \\ 0 & b & 0 \end{pmatrix}$$

$$\mathbf{A}' = \begin{pmatrix} 0 & 0 & 0 \\ a & 0 & 0 \\ c & b & 0 \end{pmatrix}$$

$$\mathbf{A}' = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ c + d \times x_1 & 0 & 0 \end{pmatrix}$$

FIGURE 4.2 – Mediation and moderation structures in SEMs

4.4.2 Identification issues

Given the degree of complexity of ICLV models, particular attention needs to be paid to their theoretical and practical identification.[6] A model is theoretically identifiable if there are no two distinct sets of parameters that generate the same probability distribution of observable data. Generally, constraints have to be imposed to obtain a unique vector of consistent parameter estimates. Theoretical identifiability has been extensively studied for both DCMs (Train, 2009) and SEMs (O'Brien, 1994; Bollen, 2014). Vij and Walker (2014) and Daziano and Bolduc (2013a) bridge the gap and derive sufficient conditions to ensure theoretical identifiability for ICLV models. However, necessary identification conditions for ICLV models still have to be developed. Even if a model is theoretically identified, a lack of data variability may lead to an empirically unidentified model. Using simulated data, Raveau et al. (2012) show that the discrete choice part of the model is more robust than the SEM model. When variability is low, the parameters obtained in the structural component of the SEM are either underestimated or overestimated. Wrong sign problems may even occur, leading to counterintuitive relationships between variables. The sample size and sampling method are therefore critical to ensure empirical identification.

4.4.3 Estimation methods

Best practice is to estimate ICLV models with a simultaneous approach (Walker, 2001), that is, jointly estimating Equations (4.1) through (4.4). The simulated maximum likelihood estimation method is used by a majority of studies (e.g., Walker et al., 2013; Paulssen et al., 2014; Scagnolarie et al., 2015), since, for a large number of latent variables, the computational complexity increases exponentially and numerical integration is not feasible. However, the sequential approach, which first estimates the latent variables and then includes them in the DCM, provides inconsistent estimates with measurement errors, since it treats the fitted latent variables as non-stochastic. The sequential approach is still sometimes preferred (e.g., Maldonado-Hinarejos et al., 2014; Anwar et al., 2014; Johansson et al., 2006) due to the high computational burden of the simultaneous approach.[7]

To address this issue, Bhat and Dubey (2014) develop the maximum approximate composite marginal likelihood inference approach with a probit specification. The dimensionality of integration in the composite marginal likelihood function is independent of the number of latent variables or the number of indicators in the measurement equations and only depends on the number of alternatives.

6. See Appendix D for more details.
7. See Appendix D for more details.

The models estimated on simulated mode choice data converge in minutes against hours with the standard frequentist estimator. The authors note, however, that larger sample sizes are required to best recover the effects of the latent variables on choice. This approach was further applied by Scagnolarie et al. (2015), with no convergence problems. In Daziano and Bolduc (2013a) and Daziano (2015), Bayesian estimation methods are developed. Daziano proposes a Bayesian estimator that accommodates a panel structure and latent variables that are endogenous and manifested through effect indicators that are discrete, continuous, or both. This estimator "avoids the curse of dimensionality, and addresses other issues, such as having exact (small-sample) properties" (Daziano, 2015, p.2). The statistical properties of this estimator are not affected by the number of latent attributes, with estimation time in the order of minutes.

Roberts et al. (2014) and Politis et al. (2012) rely exclusively on traditional SEM software (e.g., M+, AMOS) and estimate the whole model (including the DCM component) as a SEM. One advantage of this approach is that it allows for alternative estimation methods that are dedicated to ordinal data, such as the weighted least squares method with a mean and variance correction. However, this approach lacks flexibility since, for instance, it requires that all error terms are normally distributed.

4.5 Psychological dimension

After reviewing the use of behavioural theories to explain mode choice, this section proposes a grid analysis to categorize the latent variables used in ICLV models and to discuss how latent variables are explained through the structural model and how latent variables influence mode choice.

4.5.1 Linking ICLV models to behavioural theories

ICLV models were primarily developed to better understand how behaviours are formed and to uncover the "black box" of decision processes. Following an analysis of case studies, Ben-Akiva et al. (2002b, p.35) cautioned users of ICLV models to "first think clearly about the behavioural hypotheses behind the choices, then develop the framework, and then design a survey to support the model". A key step in developing ICLV models is therefore to learn from the theories developed by behavioural scientists to study mode choice.

Behavioural scientists mainly use theories developed to explain pro-environmental behaviour (Gifford et al., 2011). Three types of theories may be distinguished (Table 4.1): theories that explain behaviour with own personal interest and that are based on a rational decision process; theories based

on values and norms, and theories based on habits.[8] A further distinction depends on whether the theory only relies on internal mental states, as in the Norm-Activation Model, or also takes into account contextual factors, as in Attitude-Behaviour-Context theory. The contextual factors may act negatively (for instance, budget constraint) or positively (for instance, monetary incentives) on behaviour.

Five of the previously mentioned theories are used, or at least mentioned, in ICLV models on mode choice. However, even when clear reference is made to a specific theory, the SEM component generally simplifies it and incorporates only some variables directly in the utility of the DCM. The mediation and/or moderation structure of the psychological theories are generally omitted. An exception is Paulssen et al. (2014) who estimate a whole value-attitude-behaviour model by means of a SEM with a full mediation structure, with no direct action of values on behaviour.[9] Two other theoretical examples of mediation or moderation structures can be found in the ICLV models, but the corresponding models are not estimated. Abou-Zeid and Ben-Akiva (2011) develop a theoretical model in which the observed actions of others impact some latent variables, such as attitudes, perceptions or personality. These latent variables, in turn, impact utility for each transport mode, either directly or with the mediation of comparative happiness.[10] In the theory of interpersonal behaviour on which Tudela et al. (2011) build their simplified model, behaviour is determined by intention. The relationship between both variables is moderated by the habit latent variable. This means that, depending on habits, intention has a greater or smaller impact on the effective observed behaviour.

Paulssen et al. (2014, p.875) draw the same conclusions as us, noting that previous studies "tended to simplify significantly the cognitive theories motivating the use of these models, and much of the behavioural richness captured originally in these theories through the complex interplay between different latent psychological constructs has often been lost as a consequence of these simplifications." Roberts et al. (2014, p.3) also note that "ICLV model studies of mode choice generally devote little or no attention to the theoretical model of decision making that underlies the empirical work".

The underuse of behavioural theories may have two explanations. Firstly, incorporating behavioural theories is not always possible, since current estimation methods require long estimation times and raise convergence issues depending on the number of latent variables and indicators. However, the developments mentioned in the previous section seem to overcome these problems and may allow for a better inclusion of behaviourally sound SEMs in ICLV models. Secondly, some surveys may not have

8. See Appendix A for more details.
9. Another exception, regarding departure time choice, is Thorhauge et al. (2016).
10. The mediation is therefore partial (see Figure 4.2).

TABLE 4.1 – Behavioural theories used in ICLV models.

Theory	Summary	Used to explain mode choice in the psychological literature	Mentioned or (partly) used (*) in the ICLV literature.
Rational theories			
Theory of Planned Behaviour (TPB, Ajzen, 1991)	Rests on the assumption that the best predictor of a behaviour is intention, which is determined by attitudes, perceived social norms and perceived behavioural control with respect to the behaviour.	Heath and Gifford (2002); Bamberg et al. (2003a); Bamberg and Schmidt (2003); Bamberg et al. (2007); De Groot and Steg (2007); Wall et al. (2007)	Roberts et al. (2014), Paulssen et al. (2014), Sottile et al. (2014*)
Attitude-Behaviour-Context (ABC, Guagnano et al., 1995) Theory	Behaviour is an interactive product of attitudinal variables and contextual factors. The attitude-behaviour association is strongest when contextual factors are neutral and it approaches zero when contextual forces are strongly positive or negative, effectively compelling or prohibiting the behaviour in question.		Roberts et al. (2014*)
Norms and values based theories			
Norm-Activation Model (NAM, Schwartz, 1977)	Originally developed in the context of altruistic behaviour. Behaviour is predicted by personal norms which are actively experienced "as feelings of moral obligation not as intentions". Personal norms are determined by two factors: awareness that performing the particular behaviour has certain consequences, and feelings of responsibility about performing this specific behaviour	Bamberg and Schmidt (2003); Wall et al. (2007); Bamberg et al. (2007)	Roberts et al. (2014)
Value-Belief-Norm Theory (VBN, Stern et al., 1999)	An adaptation of NAM applied to pro-environmental behaviours. Makes a direct link between values (socio-altruist, egoistic and biospheric) and environmental beliefs which, in turn, determine personal norms and finally act on behaviour.	Nordlund and Garvill (2003); Collins and Chambers (2005)	
Value-Attitude-Behaviour Theory (Homer and Kahle, 1988)	Links values to attitudes and attitudes to behaviour in a full mediation model		Paulssen et al. (2014*)
Habits based theory Theory of Interpersonal Behaviour (TIB, Triandis, 1977)	As in the TPB, intentions are immediate antecedents of behaviour. In the TIB, habits determine behaviour as well as contextual factors. Intention is shaped by three factors: attitude, social factors and affect	Bamberg and Schmidt (2003)	Paulssen et al. (2014),Roberts et al. (2014),Tudela et al. (2011*)

been developed with the specific purpose of estimating ICLV models. Econometric improvements will therefore be fruitful only in cross-disciplinary work, involving psychologists, economists, and transport specialists, from the development of the questionnaire to the interpretation of the results.

4.5.2 Categories of latent variables

Latent variables included in ICLV models on mode choice are related to a wide variety of themes, reflecting the complexity of the mode choice decision process. Almost 60 variables have been identified (Figure 4.3). They refer to personal values (Paulssen et al., 2014), comfort (Glerum et al., 2014), safety and security (Raveau et al., 2010), environmental concern (Kamargianni et al., 2015), convenience and accessibility (Yanez et al., 2010), flexibility (Anwar et al., 2014), reliability (Raveau et al., 2010), habits (Idris et al., 2015), social comparison (Abou-Zeid and Ben-Akiva, 2011), information on stress (Sottile et al., 2014), affective attitude and addiction to the car (Polydoropoulou et al., 2014). These latent variables are very heterogeneous, in terms of meaning and measurement, even when a common term is used. Consequently, comparisons between studies and results are difficult to make since they refer to concepts that are either too dissimilar or that have different measurements. An interpretative framework, i.e. an analysis grid, is therefore necessary to get a deeper understanding of how latent variables explain mode choice.

Following an in-depth qualitative examination of the literature on ICLV models applied to mode choice, a three-entries grid seems well adapted. The first entry relies on the topic to which the latent variables refer. Eight main topics are identified: environmental concern, comfort, safety, reliability, flexibility, convenience, habits and personality traits (Figure 4.3), the first three being incorporated in a third of the ICLV models on mode choice.

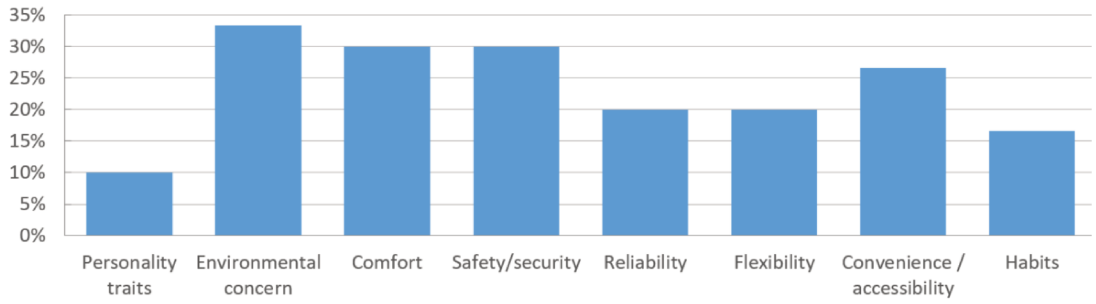


FIGURE 4.3 – Topics related to latent variables included in ICLV mode choice models

The second entry draws upon psychological concepts and distinguishes three categories of latent variables (Table 4.2).

1. The first category of latent variables describes the internal mental state of the individual (norms, values, needs), and the variables are similar to socio-economic variables. These variables are not related to a specific transport mode and they are measured with indicators, such as "How important is it for you to use a convenient and comfortable mode? A stress-free and relaxed mode? A mode on which you don't have to worry about anything while using it?" (Temme et al., 2007; Johansson et al., 2006).
2. The second category of latent variables describes the individual attitudes[11] towards or perceptions of the alternatives; here, the transport modes. In addition to being individual-specific, these variables are also alternative-specific and are similar to the objective attributes describing the alternatives. They are measured with indicators, such as "How do you rate the comfort during your trip by train?" or " In your opinion, how convenient is the coach schedule?" (e.g., Raveau et al., 2010; Daziano and Rizzi, 2015).
3. The third category of latent variables describes an actual behaviour, either related to transport or to other areas, such as waste recycling. These variables are measured with items, such as "Do you wear a bicycle helmet when cycling?" or "Do you compost kitchen refuse?" (Johansson et al., 2006).

The third entry divides the latent variables according to whether the individual is considered in isolation or taking into account her social context. Although the psychological literature underlines the importance of the social context in the shaping of behaviours by means of variables, such as subjective norms in the theory of planned behaviour or social factors in the theory of interpersonal behaviour, almost all the variables reviewed fall into the first individual category. The only application of ICLV models on mode choice to take into account social context variables is Kamargianni et al. (2014) who investigate the impact of parents' walking habits on teenagers' mode choice. Abou-Zeid and Ben-Akiva (2011) study the effect of social comparisons on travel happiness and behaviour but do not estimate the whole ICLV model.

4.5.3 Explaining the latent variables

The effect of latent variables on mode choice depends on the variables chosen in the structural model (Equation (4.2)). However, "in practice, ICLV models often suffer from weak structural equations where the observable explanatory variables are poor predictors of the latent variables, and the latent variables

11. The concept of attitude can be defined as the disposition to respond favourably or unfavourably to a specific object or class of objects (Oskamp, 1977).

TABLE 4.2 – Latent variables grid with examples

Topic	Internal mental states	Perceived alternative-specific variables	Actual behaviour
Personality trait	Temme et al. (2007) Power: concern for social status and prestige, and control or dominance over people and resource: e.g. appreciation of the statement "She/he always wants to be the one who makes the decisions. She/He likes to be the leader."	-	-
Environmental concern	Sottile et al. (2014) Personal norm regarding the environment: e.g. appreciation of the statement "Regardless of what other people do, I feel a moral duty to travel in an environmentally more sustainable way."	-	Johansson et al. (2006) Pro-environmental behaviour: e.g. the respondent's habit of composting kitchen refuse
Comfort	Johansson et al. (2006) Preferences for comfort: e.g. appreciation of the importance of travelling in a calm, non-noisy environment	Daziano and Rizzi (2015) Bus/Train comfort: e.g. appreciation of overall comfort (quality of seats, roominess, etc.)	-
Safety / security	Daziano (2015) Pro-safety consumers: appreciation of safety as a relevant aspect of travel decisions	Kamargianni et al. (2015) Safety Consciousness: e.g. appreciation of the statement "I feel safe when I use the bus"	Johansson et al. (2006) Safety behaviour: e.g. the respondent's habit of wearing a helmet when cycling
Reliability	Anwar et al. (2014) Reliability: e.g. appreciation of the importance of punctuality	Raveau et al. (2010) Reliability: e.g. appreciation of the possibility of calculating travel time before the trip for different modes (car, bus, and so on)	-
Flexibility	Temme et al. (2007) Attitude towards flexibility: e.g. appreciation of the importance of using a means of transport that can be used spontaneously and without planning	-	-
Convenience / accessibility	Johansson et al. (2006) Preferences for convenience: e.g. appreciation of the importance of avoiding queues and congestion while travelling to/from work	La Paix Puello and Geurs (2015) Perception of connectivity: e.g. appreciation of connection to other public transport modes by train	-
Habits	Rieser-Schüssler and Axhausen (2012) Interest in varying one's daily routine: e.g. appreciation of the statement "I like to experience novelty and change in my daily life."	Idris et al. (2015) Habits: for non-work activities (e.g. to visit friends/family, to go to the movies, etc.) ask to provide the mode they would use. Counting how many times the respondent had mentioned each mode to perform different activities.	Kamargianni et al. (2014) Parents' walking habits: e.g. appreciation of the statement "My mother walks for her short-distance trips."

are, in a sense, truly latent" (Vij and Walker, 2016, p.197). The literature review confirms this finding. Firstly, as already mentioned, there are generally no relationships between the latent variables, such that socio-economic variables are the only determinants of latent variables. Secondly, the structural models offer contradictory results. For instance, need for comfort is explained by age and gender in three studies (Anwar et al., 2014; Paulssen et al., 2014; Johansson et al., 2006). Depending on the study, both variables are found to be associated with a negative, positive or non-significant parameter. Another example is income, which has either a positive (Anwar et al., 2014) or a non-significant (Paulssen et al., 2014) effect on need for comfort.[12] These results suggest we should be cautious when interpreting structural models.

4.5.4 Effects of latent variables on mode choice

In most of the papers reviewed, the latent variables studied are significant determinants of mode choice with intuitive effects. The effects of latent variables on mode choice are discussed for the eight topics previously identified.

4.5.4.1 Personality traits

Two papers are specifically concerned with personality traits and personal values. Paulssen et al. (2014) find that personal values (power, hedonism and security) have an impact on attitudes, which in turn explain mode choice, but these values have no direct impact on mode choice. Córdoba and Jaramillo (2012) use the 16 pf psychometric test to measure personality traits, such as emotional stability or spontaneity. Assuming that they took the car as a reference alternative, it seems that people who choose an alternative mode are more pragmatic, shy, insecure and stressed and have a higher level of emotional instability than those who choose the car.

4.5.4.2 Environmental concern

As expected, the overall finding is that people with a higher level of environmental concern or environmental awareness are more likely to choose public transport, park and ride facilities or active transport modes (Rieser-Schüssler and Axhausen, 2012; Polydoropoulou et al., 2014; Atasoy et al., 2013; Sottile et al., 2014). However, Sottile et al. (2014) notice that individuals are less sensitive to long-term effects, such as environmental pollution, than to effects that have immediate repercussions, such as stress. When environmental concern is measured through actual behaviour, the conclusion is the same.

12. Appendix D provides a detailed analysis of how the "comfort" latent variable is explained by the structural models in the literature.

A green lifestyle favours the choice of bus and reduces the probability of choosing motorized modes, such as the car and powered two wheelers (Kamargianni et al., 2015), and increases the likelihood of choosing train over bus (Johansson et al., 2006).

Politis et al. (2012) and Sottile et al. (2013) are the two exceptions to these general results. In particular, Sottile et al. (2013) find that environmental concern, measured as the propensity to adopt environmentally friendly behaviour (waste, energy, food, technologies but not transport), favours the car alternative over the park and ride alternative. The authors justify this contradictory result with cognitive dissonance, which is the inconsistency between attitudes and behaviour. Cognitive dissonance is common in a transportation context. One explanation is that pro-environmental motivation is not enough to make people adopt behaviour that replaces or reduces private car use in favour of sustainable travel modes, since mode shift is not sufficiently easy. Indeed, individuals behave pro-environmentally only when an action is easy to perform. Another explanation is that environmental concern is measured with behavioural indicators related to non-transport fields. However, in the transportation field, environmental behaviour is generally moderately or poorly correlated with other kinds of environmentally friendly behaviour. Compensation effects may strengthen this explanation. Indeed, people with high environmental concern may ease their sense of guilt about using the car with increased environmental behaviours in other areas of life (Johansson et al., 2006).

In a revealed preferences survey, the same authors (Sottile et al., 2014) consider environmental personal norms as part of the theory of planned behaviour. After giving some information about reduction of CO2 emissions and stress, they find that high environmental moral rules and obligations lead individuals to behave in a pro-environmental way, producing an increase in the probability that they will choose park and ride instead of the car. This survey tends to prove an absence of cognitive dissonance. Three differences between the two surveys may explain these contradictory results. The first difference concerns the type of survey (revealed *versus* stated preferences). The second difference concerns the type of information provided (economic and health information for the revealed preferences survey or stress and CO2 information for the stated preferences survey). The third difference concerns the measurement of the latent variables, and this emphasizes the importance of clearly defining in which of the previously defined categories (internal mental states, perceived alternative-specific variables or actual behaviour) the latent variable falls.

4.5.4.3 Comfort

Despite heterogeneity in measures, the underlying idea behind the comfort latent variable is mainly to assess the extent to which public transport offers opportunities for alternative activities, such as relaxation and work. However, Atasoy et al. (2013) measure comfort with items that translate the difficulties of taking public transport (travelling with heavy luggage or children, having transfers, etc.). This can be approached by the notions of convenience or accessibility, which are often paired with the comfort measure (e.g., Morikawa and Sasaki, 1998; Raveau et al., 2010).

All studies involving comfort, either as a need or a perception, converge in a common finding. Comfort is an important factor in explaining mode choice. More precisely, having high comfort (and convenience) needs increases the probability of choosing a public transport mode (Temme et al., 2007). This can be explained by fatigue due to driving, lack of parking space, and the possibility of performing activities during travel by public transport.

Among public transport modes, trains are perceived as more comfortable than buses, which favours the choice of train over bus (Daziano and Rizzi, 2015). Indeed, in trains, it is generally easier to work or rest, since the seats are more comfortable, there is more space and there are fewer shocks and loads. A positive perception of public transport modes increases their utility (Glerum et al., 2014), and sensitivity to travel time decreases with a better perception of comfort on public transport. There also seems to be considerable heterogeneity with regard to the effect of comfort on the choice process since it varies significantly among individuals (Yanez et al., 2010). Moreover, the impact of latent variables on mode choice, and in particular comfort, may increase over time (Anwar et al., 2014, with a comparison between years 2008/2009 and 2010/2011).

4.5.4.4 Safety and security

Paulssen et al. (2014) use the importance of owning the transport mode as a proxy for the safety value. The ownership variable tends to disadvantage public transit modes in comparison to the car. However, with regards to the choice between only ever using the car and driving to access public transport, the effect of ownership is statistically insignificant, since both alternatives require access to a car. For Kamargianni et al. (2015), safety consciousness increases the probability of choosing the car and decreases the probabilities of bus and walking.

Yanez et al. (2010), Raveau et al. (2010), Anwar et al. (2014) and Daziano and Rizzi (2015) evaluate the level of security for each mode integrated in their model. As expected, they find that individuals tend to choose those modes perceived as safer. For instance, the train is preferred to the bus (Daziano

and Rizzi, 2015).

In Johansson et al. (2006), the safety variable, measured in terms of behaviour, is found to be insignificant in both the choice between car and bus and the choice between train and bus. One explanation is that safety and security constitute a combination of two variables: preferences for personal security and traffic safety. These two components work in opposite directions, since the car is usually perceived as being safer in terms of theft and personal safety, while public transport is more reliable in terms of traffic safety issues. Future research will therefore need to distinguish between these two effects.

4.5.4.5 Reliability

Raveau et al. (2010) and Yanez et al. (2010) define reliability as the possibility of calculating travel and waiting time before the trip, and as the availability of suitable information. They find the intuitive result that the higher a mode is valued in terms of reliability, the greater the probability that individuals will choose that mode. La Paix Puello and Geurs (2015) show that a positive perception of connectivity, including punctuality and frequency of trains, is positively associated with bicycle use to go to railway stations. Johansson et al. (2006) include both concepts (punctuality and prior knowledge of travel time) in a global variable but they do not find any significant effect on mode choice. Conversely, in Anwar et al. (2014), reliability is found to have the highest impact on mode choice, along with safety.

4.5.4.6 Flexibility

Latent variables specifically concerned with flexibility measure requirements for a mode that is available at any time, that can be used regardless of weather conditions, and that offers the possibility of shopping or transporting children during a commute to work. Need for flexibility tends to favour the car and metro over carpooling and other public transport modes (Politis et al., 2012; Johansson et al., 2006).

4.5.4.7 Convenience and accessibility

Convenience and accessibility are measured with indicators such as ease of access, frequency of public transport, the presence of queues, wait time and whether the mode is going to the required destination. It is always associated with indicators related to comfort, reliability or flexibility. The general conclusion is that modes that are perceived as convenient are more likely to be chosen (Yanez et al., 2010; Raveau et al., 2010; Anwar et al., 2014).

4.5.4.8 Habits

Car habits, measured with the Verplanken scale, result either in a non-significant effect on mode choice (Tudela et al., 2011) or in a positive effect that favours the car at the expense of the other modes (Idris et al., 2015). The latter result means that the choice of car as a transport mode is mainly determined by habits rather than traditional individual variables and modal attributes. The habits of relatives also determine mode choice. In particular, if parents tend to walk, teenagers are more prone to walk to school (Kamargianni et al., 2014).

Rieser-Schüssler and Axhausen (2012) approach the concept of habits by measuring the extent to which individuals are interested in varying their daily routine. They find that a greater desire for variation in the daily routine is associated with a higher probability of choosing public transport modes over the car.

In conclusion, the more individuals or their relatives are accustomed to a mode, the more likely they are to use that mode. This is especially true for car users, since they seem to be more attached to their routine than public transport users.

4.6 Economic dimension

ICLV models may be used with three objectives: 1) to explain behaviour and derive public policies; 2) to analyze economic outputs, and 3) to forecast market shares.

4.6.1 Explaining behaviour and deriving public policies

ICLV models generally allow for a better description of behaviours than the models to which they are compared (multinomial logit or probit models). Firstly, they outperform models without latent variables in terms of fit to the data. Several papers also report a few shifts in the value and significance of the alternative-specific parameters and state that ICLV models make it possible to detect the real role of these variables (Raveau et al., 2010; Tudela et al., 2011).

Even if the statistical effects may be due to the inclusion of additional variables in the structural model of the SEM component (Vij and Walker, 2016), ICLV models still depict mode choice behaviour with a broader set of determinants. Understanding how these determinants influence mode choice is useful for deriving accurate public policies in two ways. Firstly, the individual variables related to internal mental states may help to understand how individuals will react to a policy and can be used to support the design of an efficient policy. For instance, Temme et al. (2007) report that the German rail

operator changed its pricing policy in 2002. After some adjustments, it offered a discount card with a slightly higher discount rate (55% instead of 50%) but made it a requirement to book a seat seven days in advance. The operator did not expect the drop in market share that followed. The impact of the lack of flexibility due to the booking obligation was indeed underestimated if not actually overlooked. With a model integrating a latent measure of the need for flexibility, the outcome could have been anticipated and the pricing policy adapted to promote train use.

Secondly, the perception of the alternative-specific variables may be as important, or even more important than the objective levels of such variables (Bhattacharya et al., 2014). Indeed, individual perception of the quality of public transport modes differs from the objective measures of service quality (for instance, travel time and reliability) used by planners to make and evaluate decisions. Kaufmann (2002), for instance, shows that the alleged slowness of public transport is not only a question of the objective time it takes to make a journey, but also of the perception of that time. So, the time it takes to travel by car is often underestimated, whereas the time it takes by public transport is overestimated. Measuring and taking into account both objective and subjective measures is therefore necessary to an understanding of mode choice. Moreover, some alternative-specific variables, such as safety or comfort, are difficult to measure objectively. It is therefore all the more important to be able to measure and integrate individual perceptions of the travel mode characteristics. ICLV models could help to understand how the negative perceptions of public transport modes affect their modal shares and to set the priorities for improvement.

4.6.2 Deriving economic outputs

Since they are based on random utility theory, ICLV models may be used to derive economic outputs: market shares, elasticities and willingness-to-pay indicators, such as Value of Time (VoT). Atasoy et al. (2013) compare market shares across three models (ICLV, latent class and multinomial logit models) and observe that they do not vary much. They draw the same conclusion regarding VOT. Conversely, La Paix Puello and Geurs (2015) find that the multinomial logit model overestimates the market share of bicycles in comparison to the ICLV model. They conclude that "missing important unobserved effects may lead to overestimation of demand solution in travel behaviour". Johansson et al. (2006) also find notable differences between a multinomial logit model and an ICLV model in terms of VOT. Yanez et al. (2010) find inconsistent VOTs with a mixed logit without latent variables, reaffirming the importance of including such variables in the model formulation. Glerum et al. (2014) highlight the impact of ratings from different evaluations on the demand by exploring the three indicators, market shares, value of time

and elasticities. Polydoropoulou et al. (2014) calculate ratios for public transport and private motorized vehicles regarding in-vehicle travel time and out-of-vehicle travel time. These examples show the range of economic outputs obtained with ICLV models and highlight the differences from multinomial logit models that might occur.

4.6.3 Forecasting market shares

ICLV models allow for an improvement in the forecast of modal shares, but only on two conditions. The first condition is that forecasts for measurement indicators are available, but that condition is often not met. The second condition is that, in the structural model of the SEM, the observable explanatory variables are poor predictors of the latent variables (Vij and Walker, 2016), which is often the case. If forecasts for measurement items are expected to be unavailable, then ICLV models do not provide further insight for forecasting, compared to mixed logit models (Vij and Walker, 2016). Nevertheless, even if forecasting by means of an ICLV model has no added value from an econometric point of view, it may be interesting from a behavioural point of view. For instance, Politis et al. (2012) simulate the effect of a change of behavioural stage, that is, people's awareness of the necessity or their readiness to change their travel habits, on utility. Daziano (2015) determines that women have higher pro-environment attitudes than men. In the forecasting exercise, they predict that public transport market shares would increase by 16.9% if men were represented as having the same attitudes as women. Paulssen et al. (2014) compare the impact on market shares of two scenarios: a 10% increase of car travel time or a 10% increase of attitudes regarding comfort and convenience. Their model predicts that the first scenario would reduce the market share for the car by 8.4%. In comparison, the second scenario is expected to reduce the market share for the car by only 0.9%. Another approach is to artificially change the values of the observable variables in the structural equations of the SEM. Glerum et al. (2014) decrease the number of cars in each household by one. This leads to a better perception of comfort on public transport and slightly increases (1.8%) the market share of public transport. Anwar et al. (2014) and Yanez et al. (2010) produce a similar argument and they vary the level of income. In conclusion, the simulation of scenarios could help to provide a measure of the scope for improvement of the modal share of public transport modes, or a measure of the respective roles of latent and observable variables in the achievement of a goal, such as the reduction of car use.

4.7 Conclusion

ICLV models are rich and complex models involving multiple dimensions and which therefore represent a fruitful tool for interdisciplinary programs. This review of literature details and explores how the survey, econometric, psychological and economic dimensions of ICLV models may serve to better explain mode choice. The econometric dimension is well addressed by the literature, and recent developments will certainly help to better integrate behavioural theories in discrete choice models. At the current time, although some of these theories are beginning to be explored, there is still room for further integration. There remains considerable scope in research for integrating more complex SEMs based on models previously estimated in travel behaviour research. Moreover, latent variables used in ICLV models on mode are heterogeneous in terms of definition and measure. To structure and facilitate comparisons, an analysis grid is proposed. This paper also suggests adapting the data collection process, with special attention to that part of the questionnaire dedicated to the measurement of the latent variables. It would also be interesting for future research to broaden the latent variables included in ICLV models on mode choice and include social context variables more widely. The usefulness of ICLV models for deriving public policies has been challenged in previous research (Vij and Walker, 2016; Chorus and Kroesen, 2014). This paper provides examples to show that, although forecasting is indeed difficult with ICLV models, they are still a useful tool to explain mode choice, to identify which latent variables exert an influence on mode choice and to suggest how public policies may be designed or adapted to promote public transport use.

CHAPTER 5

Comfort and heterogeneity in values of travel time savings in the Rhône-Alpes Region[1]

Abstract

Based on a choice experiment conducted in the Rhône-Alpes Region, this paper estimates an Integrated Choice and Latent Variable model which addresses the heterogeneity in values of travel time. We evaluate how sensitivity to travel time by public transport is distributed according to the level of comfort. Comfort is modeled as a function of objective attributes and individual perceptions about it: whether a seat is guaranteed, quality of trip experience (feelings experienced), (perceived) use of travel time during the trip, and overall ease of using public transport (perceived behavioral control). The results show that the last two play a significant role in the choice of a public transport mode and that the value of time function is downward sloping with higher levels of objective and perceived comfort. We discuss public policy implications and show that the most effective measure, in terms of economic benefits, would be to optimize and target investments in seat capacity supply.

5.1 Introduction

The main objective of transport investments is time savings, and significant expenditures are made to achieve that objective. Value Of Time (VOT) is therefore a critical value in the appraisal of transport projects. However, Small (2012) brings to light that, despite decades of research on individual time

1. This chapter is a paper co-authored with Matthieu de Lapparent (HEIG-VD, Switzerland) and submitted for presentation at the 2018 Transportation Research Board Annual Meeting. Other versions of this chapter have been presented at a seminar of the "Nouvelle approche économique des mobilités dans les territoires" chair, at the JMA (2017) and AFSE (2017) conferences and submitted to the JMA special issue of *La Revue Economique*.

allocation to travel activity since the theoretical microeconomic frameworks of Becker (1965) and DeSerpa (1971) and hundreds of increasingly sophisticated empirical applications, valuation of travel time is still not well understood and deserves further investigation. In particular, the widespread heterogeneity in how people make the trade-off between time and money needs to be further explored to obtain robust and credible VOTs that can enhance transportation analysis.

To study this heterogeneity, methods that model unobserved heterogeneity or methods that model observed heterogeneity can be used. Firstly, to unveil the unobserved heterogeneity, a popular approach is to capture the distribution of VOT using a random coefficient model (logit mixture) (e.g., Algers et al., 1998; Hess et al., 2005; Hensher, 2006b; de Lapparent et al., 2013). In this approach, the distribution of the time and/or cost coefficients are chosen before the parameters of the distribution are estimated. The distribution of VOT is then derived from the estimated distribution of the coefficients, using simulation. In economics, discussions have focused on the choice of a specific distribution function (Fosgerau and Bierlaire, 2007) and the behavioral realism of allowing positive values in the distribution of the cost coefficient (Daly et al., 2012). Nevertheless, since the cost coefficient enters the denominator of VOT, it may result in arbitrarily large VOTs if the cost parameter is arbitrarily close to zero.[2] Despite this issue, the specification of the model is rarely tested in the model estimation (Börjesson et al., 2012), which questions the relevance of the values from the literature. One solution is to use non-parametric techniques to estimate the heterogeneity in VOT (Fosgerau, 2006, 2007; Börjesson et al., 2012), therefore avoiding the problem of computing the ratio between two distributions.

These methods are relevant to evaluate the extent of variations of the VOT but give no insight into the determinants of human behavior, although identifying these determinants is necessary to build appropriate economic models and design efficient public policies.[3] In that respect, it is necessary to use methods that unveil observed heterogeneity. We identify two methods. Firstly, a comprehensive meta-analysis is a critical tool for capturing heterogeneity in VOT and it has been used at the level of a country (Wardman, 1998, 2001; Wardman et al., 2004; Abrantes and Wardman, 2011), the European Union (Wardman et al., 2012) and world-wide (Shires and De Jong, 2009). Secondly, in the Discrete Choice Models (DCMs), the time variable may be crossed with individual (e.g., age, gender; see for instance, Hossan et al., 2016) or alternative-specific variables (e.g., comfort; see for instance, de Lapparent and Koning, 2016; Haywood and Koning, 2015 or Kroes et al., 2013). Sensitivity to time and, therefore, the VOT are then dependent on these additional variables. This method may also be applied

2. The moments of the VOT distribution, and especially the mean, may therefore not exist for a given distribution.
3. See Appendix E for an overview of these determinants.

to explore how latent variables influence VOT using Integrated Choice and Latent Variable (ICLV) models. The hypothesis is that the perception of and attitude to a specific mode influence perception of time and therefore VOT. The only applications of ICLV models to investigate heterogeneity in VOT use attitudinal variables related to the car (Abou-Zeid et al., 2010; Fernández-Antolín et al., 2016), thus limiting the implications for public policy. Both papers converge in the finding that the higher the car-loving attitude, the smaller the VOT. This paper further investigates this approach, still rarely used, with the originality of focusing on public transport modes instead of the car.

A key determinant of VOT is the quality of the trip experience. Indeed, among the lessons learned from the behavioral literature on travel time, the idea that travel time has its own utility, either because of the activities conducted during travel, or through the pleasure of travel itself, is becoming increasingly validated (e.g., Mokhtarian and Salomon, 2001; Lyons and Urry, 2005; Mokhtarian et al., 2015). This finding tends to qualify the engineer's point of view, which relies on travel time savings. In particular, on public transport modes, travel time can be used to perform activities, such as reading, working or resting (Lyons et al., 2013; Wardman and Lyons, 2016) and this may reduce VOT. To optimize this indirect utility of travel, comfort is critical. A polychronic use of time is indeed favored by a comfortable and pleasant travel environment, which also reduces the stress of commuting and hence negative moods (Li, 2003).

Comfort is a multidimensional concept that can be described through the availability and/or the quality of infrastructures (seats, wireless connection, noise level, cleanliness, and so on). Comfort is also a question of perception and many variables may be used to model the perception of comfort. For instance, Temme et al. (2007) and Johansson et al. (2006) assess the importance of using a convenient and comfortable mode, a stress-free and relaxed mode, or a mode on which you don't have to worry about anything while using it. Daziano and Rizzi (2015) assess the perceived comfort of public transport modes with dimensions such as the ease of travel with children and heavy luggage or the use of time during the trip for activities. Specifically, perception of travel time may be as important, or even more important than its objective use. According to Lyons et al. (2013), UK rail passengers increasingly judge that their travel time is worthwhile, and that may be linked to the improvement in service provision (comfort or delay, for instance) but also to an improvement in terms of how individuals are equipping themselves for travel (laptops and smart phones). People on higher incomes seem to benefit more from the digital revolution; they therefore undertake productive activities during their journey, which reduces its disutility and thus their VOT (Wardman and Lyons, 2016). A study conducted in the city of Lyon (Casals, 2012) also shows that, during travel by urban public transport, travelers seem to feel

mainly positive emotions, such as freedom, a good mood or openness to others, while some, however, feel embarrassed by the proximity to other travelers, noise or smells.

Out of this literature, the intuition emerges that people who benefit from high quality public transport infrastructure and who have positive perceptions of comfort and positive feelings on public transport have a lower sensitivity to travel time and are more likely to choose public transport modes. However, such a relationship has never been quantified in the literature. The aim of this paper is therefore to investigate precisely how behaviors and VOT depend on objective and perceived comfort attributes. To our knowledge, this is the first paper to measure the interaction between time and (perceived) comfort quantitatively and to use it to explore behavioral heterogeneity and explain travel mode choice. Another contribution of this paper is to address the comfort in both its objective and its perceptual dimensions. To address these issues, we use an ICLV model, that is, a model that combines a Structural Equation Model (SEM) to measure the latent perception of comfort and a Discrete Choice Model (DCM), based on random utility theory, to explain the travel mode choice.

The approach adopted in this paper is consistent with recent approaches that integrate latent attitudes and opinions in decision making when faced with discrete travel choices (e.g., Thorhauge et al., 2016; Sottile et al., 2013; Johansson et al., 2006). Indeed, there has been a growing trend in this direction since the seminal work of Ben-Akiva et al. (2002b), which makes it possible to efficiently integrate the qualitative features of travel into DCMs. Researchers have now recognized that decision makers differ significantly according to psychological factors, such as attitudes, perceptions, values, or lifestyle preferences, and that these factors affect the utility of an alternative in a systematic way. Perception of comfort is one of the most widely studied latent variables to explain mode choice, [4] but it has not yet been approached with the dimensions that we believe to be directly linked to VOT: feelings experienced during travel (Morris and Guerra, 2015), perceived use of travel time (Lyons et al., 2013) and overall ease of using public transport. Based on the estimation of ICLV models, scenarios are constructed to simulate VOTs and market shares, which vary according to the three perceptual dimensions, as well as the guarantee of having a seat. By means of these scenarios, it is possible to compare the effectiveness, in terms of VOT or market shares, of public policies based either on investments in seat capacity supply or improved individual well-being during travel.

The rest of the paper is organized as follows. The data are presented in Section 5.2. Model specifications and estimations are developed in Section 5.3. The results are discussed in Section 5.4. Conclusions are drawn and further extensions are proposed in Section 5.5.

4. See Chapter 4.

5.2 Data

Stated Preferences (SP) data were collected between January and April 2015 in the Rhône-Alpes Region (France). In addition to the choice experiment questions, the originality of this survey is that it includes questions about attitudes to and perceptions of public transport modes.

5.2.1 The survey

The survey methods consisted of face-to-face and web-based interviews. The sample of surveyed travelers was compiled from two sources. We first sampled respondents from a large revealed preferences travel survey carried out in the same region. This database of more than 37,000 travelers is geographically stratified (Hurez and Tébar, 2014). Those travelers who declared that they had used the train as a mode of transport on one of their reported trips were asked to answer the web survey. Due to the low rate of regular train users in the population, they were oversampled with a face-to-face survey carried out in regional trains using the quota sampling method (sex, age, motive, travel time and train line).

Respondents were first asked to describe in detail (time, cost, purpose, origin and destination) a journey they had made by coach, train or car during the last month within the area of the Rhône-Alpes Region. This reference journey was then used to tailor the choice questions. Such a strategy is known to minimize the hypothetical bias.

Only respondents living in the Rhône-Alpes Region, aged 18 or over, having a car and a driving license and whose trip was made or could have been made by train or coach, were asked to answer the choice questions. The availability of the alternatives was checked by creating a database with travel time by public transport and car for each of the 8.6 million origin-destination pairs in the Rhône-Alpes Region, within a radius of 10 km around train stations. In total, 1,120 persons answered the whole SP survey (both choice and attitudinal questions, see hereafter). Table 5.1 reports the descriptive statistics for all the variables used in the models. It is important to note that the survey was not designed to be representative of the entire population of travelers in the region but rather to analyze drivers of mode choice.

5.2.2 Choice questions

The choice experiment focused on mode choice. Each respondent had to choose between three transport modes: train, coach or car. Alternatives were described in terms of travel mode, cost, time,

TABLE 5.1 – Descriptive statistics

Variable definition	Label	Mean	S.D.	Min	Max
Alternative-specific variables					
Travel cost by train (in euros)	Train cost	8.96	7.91	1.00	62.00
Travel cost by coach (in euros)	Coach cost	8.91	8.05	1.00	78.00
Travel cost by car (in euros)	Car cost	10.41	8.93	1.00	62.00
Travel time by train (in minutes)	Train time	72.42	52.73	7.00	325.00
Travel time by coach (in minutes)	Coach time	72.87	54.28	7.00	325.00
Travel time by car (in minutes)	Car time	59.13	38.54	4.00	330.00
Comfort in train (1 if a seat is guaranteed, 0 otherwise)	Comfort	0.51			
Individual variables					
Age (in years)	Age	46.15	15.63	19.00	83.00
Number of cars in the household	Cars	1.68	0.72	1.00	5.00
Gender (1 if man, 0 if woman)	Gender	0.52	0.50	0.00	1.00
Presence of children in the household (1 if yes, 0 otherwise)	Child	0.34	0.47	0.00	1.00
Monthly income above 4,000 euros (1 if yes, 0 otherwise)	Income_h	0.29	0.45	0.00	1.00
Car user for the reference trip (1 if yes, 0 otherwise)	Car_user	0.49	0.50	0.00	1.00
Perceived behavioral control					
I'm not comfortable when I travel with people I don't know well.	Pbc1	3.67	1.03	1.00	5.00
It's hard to take public transport when I travel with my children.	Pbc2	2.80	1.10	1.00	5.00
It's hard to take public transport when I travel with bags or luggage.	Pbc3	2.10	1.02	1.00	5.00
Perceived time					
I like seeing people and having other people around me.	Ptime1	3.30	0.90	1.00	5.00
It's time I put up with and I just wait for it to pass.	Ptime2	3.21	1.10	1.00	5.00
I use the time to rest and relax.	Ptime3	3.83	0.89	1.00	5.00
I use the time to do things I wouldn't necessarily do elsewhere.	Ptime4	3.28	1.05	1.00	5.00
I just want to be on my own and undisturbed.	Ptime5	2.85	1.04	1.00	5.00
Given my commutes, the time is too short: I don't have time to do anything.	Ptime6	3.54	0.89	1.00	5.00
It's wasted time.	Ptime7	3.55	1.03	1.00	5.00
Feelings					
I feel a sense of freedom.	Feel1	2.36	1.00	1.00	4.00
It puts me in a good mood.	Feel2	2.50	0.77	1.00	4.00
I feel comfortable and at ease.	Feel3	2.55	0.77	1.00	4.00
I feel I could meet people and get into conversation with them.	Feel4	2.12	0.80	1.00	4.00
I feel I'm doing something, I feel useful.	Feel5	1.83	0.86	1.00	4.00
I find the people, noise and smells disagreeable.	Feel6	3.09	0.68	1.00	4.00
I feel stressed.	Feel7	3.60	0.64	1.00	4.00
I feel harassed.	Feel8	3.73	0.53	1.00	4.00

Notes: The perceived behavioral control and perceived time items are measured on the basis of 5-point Likert scales, which range from "completely disagree" (1) to "completely agree" (5). feelings experienced in public transport are measured on 4-point Likert scales ranging from "never" (1) to "always" (5).

probability and time delay, frequency, clock-face timetable and comfort. To avoid a cognitive burden, variations of the attributes describing the proposed journeys were split into three choice exercises. We

focus here on the one that describes modes of transport that differ with respect to travel time, travel cost and level of comfort. Figure 5.1 depicts one choice exercise. Travel time was defined from origin to destination (sum of access time, egress time, waiting time and in-vehicle time). Travel cost included public transport ticket or pass, gasoline, parking cost and toll. Comfort is defined as a dummy variable that models whether a seat is guaranteed (comfort = 1) or not guaranteed (comfort=0). Respondents were faced with a series of four choice questions. Since a few respondents did not answer all of them, 4,456 observations are available.

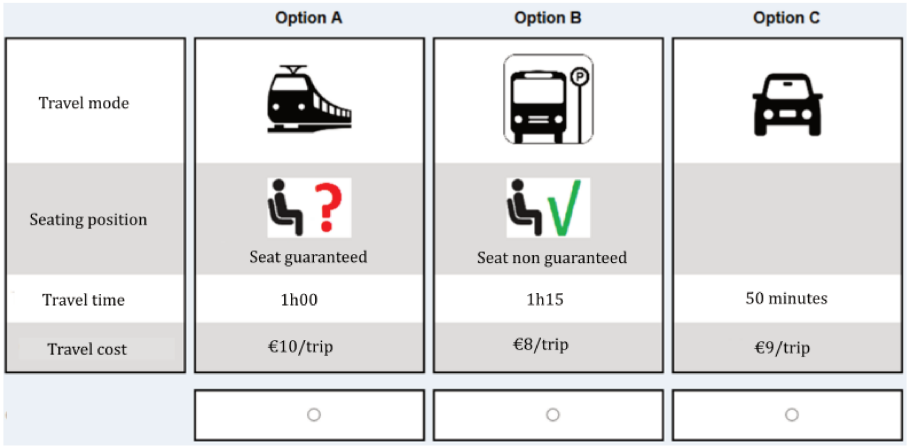


FIGURE 5.1 – Example of choice question in the choice experiment

One of the modes of transport is defined as a status-quo alternative. For the mode of transport reported in the reference journey, the actual travel time and cost attributes were systematically proposed in the choice experiment. For the other alternatives, the levels of the time and cost attributes are pivoted around the collected reference values. To improve the efficiency of the design, a Bayesian efficient design was implemented (Rose et al., 2008). *A priori* weights of attributes were taken from the literature and adjusted during the pilot tests.

5.2.3 Attitudinal variables

The last part of the questionnaire is dedicated to the collection of additional socio-economic characteristics and quantitative information to capture attitudes and psychological constructs about traveling habits. This part of the questionnaire attempts to measure three sets of attitudinal variables: environmental concern, motives for car use and perception of comfort in public transport. A first survey, dedicated to the measurement of these latent constructs, allowed us to refine the phrasing and selection of the measurement items. [5]

5. See Chapters 2 and 6.

To investigate heterogeneity in VOT, we focus here on the variables that pertain to perception of comfort in public transport. They model three main features of perceived comfort during travel by public transport modes: *Perceived Time* in interurban public transport, *Feelings* experienced during journeys made by public transport and Perceived Behavioral Control (*PBC*) on using interurban public transport.

The notion of PBC is part of the theory of planned behavior (Ajzen, 1985). This theory is based on the idea that behavior is driven by internal mental states rather than external conditions, with the assumption that behavior is the outcome of a deliberative conscious process (Savage et al., 2011). Behavior is supposed to be determined by intention, which is, in turn, determined by a combination of three factors: attitudes, social norms and PBC. PBC is defined as the perceived ease or difficulty with which an individual performs a particular behavior, here traveling by public transport.

Table 5.1 lists all the items presented in the survey to measure these three latent variables. The internal consistency of the *Perceived Time* latent variable improves without the item *ptime5*. This item is thus dropped for further analysis. The measurement for *Perception of time* and *Feelings* is based on a local study carried out on public transport in Lyon (Casals, 2012). The items used to represent PBC are based on Atasoy et al. (2011) and Morikawa et al. (1996).

5.3 Integrated choice and latent variable modeling

5.3.1 Specification, identification, estimation

We use the ICLV framework proposed by Walker (2001). Individual n ($n = 1, \cdots, N$) obtains the following level of utility from alternative j ($j = 1, \cdots, J$):

$$U_{n,j} = V_{n,j}\left(\mathbf{y}_{n,j}, \mathbf{x}_n, \boldsymbol{\xi}_n\right) + \varepsilon_{n,j}, \tag{5.1}$$

where $V_{n,j}$ is the deterministic part of the utility, $\mathbf{y}_{n,j}$ ($j = 1, \cdots, J$) is the vector of levels of attributes of alternative j individual n is faced with, $\mathbf{x}_n$ is a vector of her K observed characteristics ($\mathbf{x}_n^T = (x_{n,1}, \cdots, x_{n,K})$), and $\boldsymbol{\xi}_n$ is a vector which contains the Q latent variables ($\boldsymbol{\xi}_n^T = (\xi_{n,1}, \cdots, \xi_{n,Q})$) that model her indirectly observed attitudinal variables. $\boldsymbol{\varepsilon}_{n,j}$ are idiosynchratic shocks, which are assumed to be independently and identically distributed type I Extreme Value across individuals and alternatives: $\varepsilon_{n,j} \overset{\text{iid}}{\rightarrow} EV\left(0,1\right)$. The scale normalization is for identification purposes (for instance, see Train, 2009). What is observed is $c_{n,j}$, the choice of an alternative j made by individual n. Under

random utility maximization, the probability that individual n with characteristics $\mathbf{x}_n$ and $\boldsymbol{\xi}_n$, chooses j when faced with levels of attributes $\mathbf{y}_n$ ($\mathbf{y}_n^T = (\mathbf{y}_{n,1}^T, \cdots, \mathbf{y}_{n,J}^T)$) is given by a logit model. Let us define for all $j = 1, \cdots, J$,

$$c_{n,j} = \begin{cases} 1 & \text{iif } U_{n,j} \geq U_{n,j'} \text{ for } j' \in \{1, \cdots, J\}, \\ 0 & \text{otherwise.} \end{cases} \tag{5.2}$$

Then, we obtain for all $j = 1, \cdots, J$

$$\Pr\left(c_{n,j} = 1 | \mathbf{y}_n, \mathbf{x}_n, \boldsymbol{\xi}_n\right) = \frac{\exp\left(V_{n,j}\left(\mathbf{y}_{n,j}, \mathbf{x}_n, \boldsymbol{\xi}_n\right)\right)}{\sum_{i=1}^{J} \exp\left(V_{n,i}\left(\mathbf{y}_{n,i}, \mathbf{x}_n, \boldsymbol{\xi}_n\right)\right)}. \tag{5.3}$$

As modelers, we do not observe the latent attitudes and opinions of the individuals. We therefore try to approximate them by asking further questions that give proxy measures by means of the measurement model (Equations (5.5) and (5.6)) and assume that they are generated by some causal relations through the structural model (Equation (5.4)). We conjecture that the latent attitudinal variables are generated by a linear combination of individual characteristics

$$\xi_{n,q} = \sum_{k=1}^{K} \kappa_{q,k} x_{n,k} + \omega_q \eta_{n,q}, \forall q = 1, \cdots, Q, \tag{5.4}$$

where $\eta_{n,q}$ are unobserved characteristics. They are independently and identically normally distributed: $\forall q = 1, \cdots, Q, \eta_{n,q} \overset{iid}{\rightarrow} \mathcal{N}(0,1)$. They are weighted by $\omega_q (q = 1, \cdots, Q,)$, which measure the dispersion of these unobserved characteristics across the population of individuals. $\kappa_{q,k}$ are parameters to estimate.

These latent variables in turn generate additional latent variables $z^\star_{n,p_q}$ ($p_q = 1, \cdots, P_q$), for which we obtain discrete observed measures z_{n,p_q} ($p_q = 1, \cdots, P_q$). As discussed in the data collection section, specific questions were asked to capture the perceived comfort when making a trip. Respondents had to reveal their opinions about this by rating several discrete items, defined on Likert scales. Stated differently, each latent variable $\xi_{n,q}$ is measured by a series of P_q associated items z_{n,p_q}. We assume that the observed answers are caused by the associated latent variables $z^\star_{n,p_q}$, which are defined as linear functions of the target latent attitudes and opinions $\xi_{n,q}$ as follows

$$z^\star_{n,p_q} = \alpha_{p_q} + \lambda_{p_q} \xi_{n,q} + \sigma^\star_{p_q} v_{n,p_q}, \ \forall q = 1, \cdots, Q, \ \forall p_q = 1, \cdots, P_q, \tag{5.5}$$

where α_{p_q} and λ_{p_q} are parameters to estimate. We assume that the variables $z^\star_{n,p_q}$ are caused by the perceptions of comfort that we want to capture,[6] but that there is also a superimposed (additive) error term modeling the fact that unobserved item specific characteristics v_{n,p_q} may exist which affect the observed answers. $v_{n,p_q} \overset{iid}{\rightsquigarrow} \mathcal{N}(0,1)$. $\sigma^\star_{p_q}$ $(p_q = 1, \cdots, P_q)$ measure the dispersion of these unobserved characteristics across the population of individuals. In order finally to obtain a one-to-one mapping between the presumed data-generating process and the observed data, we need to add normalization and exclusion constraints. To this extent, for all $q = 1, \cdots, Q$, $\alpha_{1_q} = 0$, $\lambda_{1_q} = 1$, and $\sigma^\star_{1_q} = 1$.

The observed answers to the items are discrete variables that have an inherent underlying ordered pattern because of the Likert scale initially chosen. They can therefore be modeled as ordered censored variables:

$$z_{n,p_q} = \begin{cases} 1 & \text{iif} & & & z^\star_{n,p_q} & \leq & \bar{z}_{1,p_q} \\ 2 & \text{iif} & \bar{z}_{1,p_q} & < & z^\star_{n,p_q} & \leq & \bar{z}_{2,p_q} \\ \cdots & & & & & & \\ L_{p_q} & \text{iif} & \bar{z}_{L_{p_q}-1,p_q} & < & z^\star_{n,p_q}. & & \end{cases} \tag{5.6}$$

L_{p_q} is the total number of categories for item z_{n,p_q} and the $\bar{z}$'s parameters are thresholds or cutoff points for $z^\star_{n,p_q}$ that determine the probabilities of observing each category of z_{n,p_q}. The probability that the item takes the value l is defined as:

$$\begin{aligned} \Pr(z_{n,p_q} = l) &= P(\bar{z}_{l-1,p_q} < z^\star_{n,p_q} \leq \bar{z}_{l,p_q}) \\ &= \Phi\left(\frac{\bar{z}_{l,p_q} - \alpha_{p_q} - \lambda_{p_q}\xi_{n,q}}{\sigma^\star_{p_q}}\right) - \Phi\left(\frac{\bar{z}_{l-1,p_q} - \alpha_{p_q} - \lambda_{p_q}\xi_{n,q}}{\sigma^\star_{p_q}}\right). \end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.[7] The model on observed items is a series of independent ordered probit models.

As statistical inference is based only on observed outcomes, the objective is to model the joint probability of the observed choices $\mathbf{c}_n$ ($\mathbf{c}_n^T = (c_{n,1}, \cdots, c_{n,J})$) and observed answers to the items $\mathbf{z}_n$ ($\mathbf{z}_n^T = (z_{n,1}, \cdots, z_{n,P_Q})$) associated with the attitudinal latent variables. For a given individual n, the contribution to the likelihood function is defined as

$$f_{c,z}(\mathbf{c}_n, \mathbf{z}_n | \mathbf{y}_n, \mathbf{x}_n; \boldsymbol{\Theta}) = \int_{\mathbb{D}(\boldsymbol{\xi}_n)} f_c(\mathbf{c}_n | \mathbf{y}_n, \mathbf{x}_n, \boldsymbol{\xi}_n) f_z(\mathbf{z}_n | \boldsymbol{\xi}_n) f_\xi(\boldsymbol{\xi}_n | \mathbf{x}_n) \, \mathrm{d}\boldsymbol{\xi}_n \,, \tag{5.7}$$

6. This assumption defines a reflective model, as opposed to a formative model, in which the latent variables are the consequences of the measurement items (see Hoyle, 2012, for more details).

7. Due to the symmetry of the normal distribution, only $(L_{p_q} - 1)/2$ thresholds are estimated. For the *Feelings* latent variable with four categories, only one threshold is estimated, zero being the first threshold to which $\bar{z}_{1,p_q}$ is added or subtracted.

where $\mathbb{D}(\boldsymbol{\xi}_n)$ is the space on which $\boldsymbol{\xi}_n$ is defined and $\boldsymbol{\Theta}$ is a vector collecting all the parameters that we want to estimate. Moreover, the following simplified notations are adopted

$$\begin{aligned} f_c(\mathbf{c_n}|\mathbf{y}_n, \mathbf{x}_n, \boldsymbol{\xi}_n) &:= \prod_{j=1}^{J} \left[\Pr\left(c_{n,j} = 1|\mathbf{y}_n, \mathbf{x}_n, \boldsymbol{\xi}_n\right)\right]^{c_{n,j}}, \\ f_z(\mathbf{z}_n|\boldsymbol{\xi}_n) &:= \\ \prod_{q=1}^{Q} \prod_{p_q=1}^{P_q} &\left[\Phi\left(\frac{\bar{z}_{l,p_q} - \alpha_{p_q} - \lambda_{p_q}\xi_{n,q}}{\sigma^{\star}_{p_q}}\right) - \Phi\left(\frac{\bar{z}_{l-1,p_q} - \alpha_{p_q} - \lambda_{p_q}\xi_{n,q}}{\sigma^{\star}_{p_q}}\right)\right]^{\mathbf{1}\left(z_{n,p_q}=l\right)}, \\ f_\xi(\boldsymbol{\xi}_n|\mathbf{x}_n) &:= \prod_{q=1}^{Q} \frac{1}{\sqrt{2\pi\omega_q^2}} \exp\left(\frac{\left(\xi_{n,q} - \sum_{k=1}^{K} \alpha_{q,k} x_{n,k}\right)^2}{\omega_q^2}\right). \end{aligned} \tag{5.8}$$

The loglikelihood function is defined as:

$$\mathcal{L}(\boldsymbol{\Theta}|\mathbf{c}_n, \mathbf{z}_n, \mathbf{y}_n, \mathbf{x}_n) = \sum_{n=1}^{N} \ln f_{c,z}(\mathbf{c}_n, \mathbf{z}_n|\mathbf{y}_n, \mathbf{x}_n; \boldsymbol{\Theta}). \tag{5.9}$$

The loglikelihood function has no closed analytic form. We therefore use the Full Information Maximum Simulated Likelihood (FIMSL) to estimate the parameters of the model. Such an approach is numerically challenging as the loglikelihood function has no global maximum and involves simulation. For this reason, many applied ICLV models are still estimated using a sequential approach (e.g., Maldonado-Hinarejos et al., 2014; Anwar et al., 2014; Johansson et al., 2006) although such an approach provides inconsistent estimates with measurement errors. To use a simultaneous estimation method, we manage the problem as follows:

1. We first estimate the SEM component of our model, i.e. we estimate the structural latent equations and the associated multivariate ordered probit model that are measuring them;
2. we then simulate the values of the latent variables;
3. we continue by estimating the discrete choice model using simulated values of the latent variables;
4. finally, we use all former estimates as starting values for the FIMSL maximization program. There is a further inner step in that we gradually increase the number of Halton draws that are used for simulation from 50 to 1000.

In our application, there are three alternatives ($J = 3$) and three latent variables ($Q = 3$). The *PBC* latent variable is measured with three items ($P_1 = 3$) evaluated on a five-point Likert scale ($L_{p_1} = 5$). The *Perceived Time* latent variable is measured with six items ($P_2 = 6$) evaluated on a five-point Likert scale ($L_{p_2} = 5$). The *Feelings* latent variable is measured with eight items ($P_3 = 8$) evaluated on a

four-point Likert scale ($L_{p_3} = 4$).

The empirical application assumes that the utility functions are defined as follows

$$
\begin{aligned}
U_{n,\text{train}} &= ASC_{\text{train}} + (\beta_1 + \beta_5 \text{comfort}_{n,\text{train}} + \beta_6 \xi_{n,1} + \beta_7 \xi_{n,2} + \beta_8 \xi_{n,3}) \text{time}_{n,\text{train}} \\
&\quad + \beta_2 \text{cost}_{n,\text{train}} + \beta_3 \text{comfort}_{n,\text{train}} + \delta_1 \text{car_user}_n + \tilde{\mathbf{x}}_n^T \tilde{\boldsymbol{\delta}}_{\text{train}} \\
&\quad + \gamma_1 \xi_{n,1} + \gamma_2 \xi_{n,2} + \gamma_3 \xi_{n,3} + \varepsilon_{n,\text{train}} \\
U_{n,\text{coach}} &= ASC_{\text{coach}} + (\beta_1 + \beta_6 \xi_{n,1} + \beta_7 \xi_{n,2} + \beta_8 \xi_{n,3}) \text{time}_{n,\text{coach}} \\
&\quad + \beta_2 \text{cost}_{n,\text{coach}} + \delta_1 \text{car_user}_n + \tilde{\mathbf{x}}_n^T \tilde{\boldsymbol{\delta}}_{\text{coach}} \\
&\quad + \gamma_1 \xi_{n,1} + \gamma_2 \xi_{n,2} + \gamma_3 \xi_{n,3} + \varepsilon_{n,\text{coach}} \\
U_{n,\text{car}} &= \beta_4 \text{time}_{n,\text{car}} + \beta_2 \text{cost}_{n,\text{car}} + \varepsilon_{n,\text{car}} \ ,
\end{aligned}
\tag{5.10}
$$

where $\tilde{\mathbf{x}}_n$ is the $(K-1)$ truncated vector of individual variables without the type of user (car_user_n) such that $\mathbf{x}_n^T = (\text{car_user}_n, \tilde{\mathbf{x}}_n^T)$. The Alternative-Specific Constants (ASC), β's, δ's and γ's are parameters to estimate. In particular, $\tilde{\boldsymbol{\delta}}_{\text{train}}$ (respectively, $\tilde{\boldsymbol{\delta}}_{\text{coach}}$) is the $(K-1)$ vector containing the parameters of the individual variables, except the type of user, associated with the train alternative (respectively, the coach alternative). For all models, the car alternative is the reference alternative.

With regard to comfort, we make the hypothesis that people have different sensitivities to comfort according to the mode. Since for the car and coach modes a seat is always guaranteed, the parameter associated with the comfort variable is not identifiable and its effect on mode choice is included in the alternative-specific constant. Moreover, to take into account the effect of the perceived and objective comfort on the value of time, it is necessary to consider interactions between travel time and comfort as well as between travel time and latent variables. The sensitivity to these interactions is modeled with the parameters β_5, β_6, β_7 and β_8. Moreover, sensitivity to travel time is presumed to differ between car and public transport alternatives

Six models are estimated. Given the normalization and exclusion constraints, there is always a one-to-one mapping between the vectors of unknown parameters and the value of the loglikelihood functions to maximize. The first model ("MNL1") is a multinomial logit model and considers that the utility levels are functions of attributes of alternatives and socio-economic characteristics. Attributes are independently introduced. We assume here that $\beta_5 = \beta_6 = \beta_7 = \beta_8 = \gamma_1 = \gamma_2 = \gamma_3 = 0$.

The second model ("MNL2") is also a multinomial logit model and extends the former model by interacting the seat comfort variable with travel time for the train alternative. We assume here that $\beta_6 = \beta_7 = \beta_8 = \gamma_1 = \gamma_2 = \gamma_3 = 0$.

The next three models are ICLV models, each involving only one of the three latent variables. It is assumed that, with the exception of the type of user, socio-economic characteristics do not directly impact on the levels of utility ($\tilde{\boldsymbol{\delta}}_{\text{train}} = \tilde{\boldsymbol{\delta}}_{\text{coach}} = \mathbf{0}$) but rather on the formation of the latent psychological constructs. [8] If socio-economic characteristics play a role in Model "MNL2", we expect the parameters of the latent variables to be significant when introduced as explanatory variables into the utility function. As the objective is to evaluate how much the latent variables affect sensitivity to travel time in public transport, hence VOT, we also interact them with travel time. The third model ("ICLV Pbc") assumes that $\beta_7 = \beta_8 = \gamma_2 = \gamma_3 = 0$,. The fourth Model ("ICLV Ptime") assumes that $\beta_6 = \beta_8 = \gamma_1 = \gamma_3 = 0$. The fifth Model ("ICLV Feel") assumes that $\beta_6 = \beta_7 = \gamma_1 = \gamma_2 = 0$.

The last Model ("ICLV Full") assumes that the three latent variables may have roles of their own as well as impacts on sensitivity to travel time for public transport modes. [9] As previously, no socio-economic characteristics are directly introduced into the utility functions, such that $\tilde{\boldsymbol{\delta}}_{\text{train}} = \tilde{\boldsymbol{\delta}}_{\text{coach}} = \mathbf{0}$.

5.3.2 Value of Time and Market shares

From among the many economic indicators, we focus here on VOT and market shares. VOT is defined as the budget amount one is willing to pay to marginally save travel time and to keep the same level of utility (for a seminal work on the concept of VOT, see Becker, 1965). It is formally defined as a marginal rate of substitution. Given the specification of the utility functions (see Equation (5.10)) and the different estimated models, we obtain VOT functions that depend on transport modes and, according to the models, on the latent variables

$$\begin{aligned} \text{VOT}_{n,j} &= \int_{\mathbb{D}(\boldsymbol{\xi}_n)} \frac{\beta_1 + \beta_5 \text{comfort}_{n,j} + \beta_6 \xi_{n,1} + \beta_7 \xi_{n,2} + \beta_8 \xi_{n,3}}{\beta_2} f\left(\boldsymbol{\xi}_n | \mathbf{x}_n\right) \mathrm{d}\boldsymbol{\xi}_n, \ \forall j \in \{\text{train}, \text{coach}\}, \\ \text{VOT}_{n,car} &= \frac{\beta_4}{\beta_2}. \end{aligned} \tag{5.11}$$

Simulation is again required to simulate VOTs for public transport modes, as they depend on the latent variables modeled.

Market shares are not relevant as such, since SP data are used, but the variations between market shares according to the values taken by the comfort variables provide information on the extent to which the car market share could be reduced or increased on the basis of selected public policies (see

8. Additional ICLV models which include additional socio-economic characteristics in the utility function have been estimated. However, these models are not necessary to study heterogeneity in VOT. We therefore prefer to keep in line with a simple specification for reasons of parsimony.

9. ICLV models that take into account the correlation between the latent variables by means of a Cholesky decomposition of the variance-covariance matrix, as well as ICLV models that take into account a hierarchical structure between the latent variables have been estimated. These models do not improve the results, possibly due to the data collection process, which is not designed to take such relationships into account.

next section for more details on these policies). As for VOT, market share functions depend on transport modes and, according to the model, on the latent variables:

$$\text{MS}_j = \frac{1}{N}\sum_{n=1}^{N}\int_{\mathbb{D}(\boldsymbol{\xi}_n)} \Pr\left(c_{n,j}=1|\mathbf{y}_{n,j},\mathbf{x}_n,\boldsymbol{\xi}_n\right) f\left(\boldsymbol{\xi}_n|\mathbf{x}_n\right) \mathrm{d}\boldsymbol{\xi}_n, \quad \forall j \in \{\text{train},\text{coach},\text{car}\}. \tag{5.12}$$

5.4 Results

All models were estimated using Python Biogeme (Bierlaire, 2016). Note that for the ICLV models, the robustness of the results to the initial values was tested (and confirmed).

5.4.1 Estimates

Estimates are reported in Tables 5.2 for the DCMs, 5.3 for the measurement latent variable models, and 5.4 for the structural latent variable models. The goodness-of-fit indicators are displayed and discussed in Appendix F.

With regard to the DCMs, the alternative-specific constants show that, all things being equal, public transport modes have higher probabilities of being chosen. This result is the outcome of oversampling public transport users. The cost and time parameters are negative and significant in all models: demands for transport modes are downward sloping with respect to money and time budgets.

In Model "MNL1", the train comfort dummy variable is introduced as a fixed effect. Its parameter is positive and significant: the probability of choosing the train increases if a seat is guaranteed. For the other models, we further interact this dummy variable with travel time. The results are in line with empirical findings from de Lapparent and Koning (2016), Haywood and Koning (2015), Kroes et al. (2013) and many others: the main effects of objective comfort attributes are not significant but they play a role in the interaction with travel time.[10] The significant positive sign of the parameter associated with the cross-variable means that the more comfortable the journey by train, the higher the probability of choosing the train mode. Due to the linear formulation of the interaction, the longer the travel time, the more important the level of comfort.

10. Moreover, a Likelihood-ratio test confirms that the "MNL2" Model outperforms the "MNL1" Model (p-value<0.01).

TABLE 5.2 – Estimation results of the DCM component

	MNL1	**MNL2**	**ICLV Pbc**	**ICLV Ptime**	**ICLV Feelings**	**ICLV Full**
ASC_{Coach}	1.080 (0.223) ***	1.370 (0.228) ***	0.391 (0.215) *	1.070(0.139) ***	1.280 (0.121) ***	0.386 (0.230) *
ASC_{Train}	2.020 (0.215) ***	2.240 (0.217) ***	0.495 (0.215) ***	1.180 (0.143) ***	1.380 (0.125) ***	0.487 (0.231) **
Time (Train+Coach)	-0.027 (0.001) ***	-0.030 (0.001) ***	-0.032 (0.003) ***	-0.033 (0.002) ***	-0.0314 (0.0015) ***	-0.030 (0.003) ***
Time (Car)	-0.027 (0.002) ***	-0.027 (0.002) ***	-0.028 (0.002) ***	-0.028 (0.002) ***	-0.0279 (0.0019) ***	-0.028 (0.002) ***
Cost	-0.114 (0.008) ***	-0.117 (0.008) ***	-0.121 (0.009) ***	-0.126 (0.009) ***	-0.126 (0.009) ***	-0.126 (0.009) ***
Comfort	0.612 (0.068) ***	0.006 (0.108)	-0.020 (0.108)	-0.041 (0.106)	-0.00274 (0.105)	-0.004 (0.106)
Comfort×Time (Train)		0.008 (0.001) ***	0.009 (0.001) ***	0.009 (0.001) ***	0.009 (0.001) ***	0.009 (0.001) ***
Car user (Train+Coach)	-2.820 (0.096) ***	-2.820 (0.096) ***	-2.850 (0.095) ***	-2.900 (0.093) ***	-2.770 (0.095) ***	-2.780 (0.098) ***
Age (Train)	-0.0786 (0.0330) **	-0.082 (0.033) ***				
Age (Coach)	0.010 (0.035)	-0.003 (0.035)				
Child (Train)	0.0451 (0.095)	0.064 (0.096)				
Child (Coach)	-0.169 (0.109)	-0.189 (0.109) *				
Cars (Train)	-0.305 (0.064) ***	-0.300 (0.064) ***				
Cars (Coach)	-0.053 (0.066)	-0.064 (0.066)				
Income_h (Train)	-0.102 (0.097)	-0.097 (0.098)				
Income_ (Coach)	0.215 (0.111) **	0.216 (0.110) **				
Gender (Train)	-0.183 (0.084) **	-0.174 (0.085) **				
Gender (Coach)	-0.108 (0.096)	-0.099 (0.096)				
PBC (Train+Coach)			1.15 (0.248) ***			1.100 (0.279) ***
Time×PBC (Train+Coach)			0.001 (0.004)			-0.002 (0.004)
Ptime (Train+Coach)				0.521 (0.196) ***		0.202 (0.258)
Time×Ptime (Train+Coach)				0.006 (0.003) **		0.001 (0.004)
Feelings (Train+Coach)					0.159 (0.061) ***	0.0004 (0.074)
Time×Feelings (Train+Coach)					0.003 (0.001) ***	0.004 (0.001) ***
$\ell(\hat{\Theta})$	-3,647	-3,669	-21,061	-37,349	-37,083	-88,326
$\bar{\rho}^2$	0.252	0.257	0.322	0.218	0.300	

Notes: For all models, the likelihood-ratio test shows that Robust standard errors: in parentheses. P-values: ***=sign. at the 1% level; **=sign. at the 5% level; *=sign. at the 10% level. Ptime: Perceived time. ASC: Alternative-Specific Constant. For numerical reasons, Age is divided by 10.

TABLE 5.3 – Estimation results of the measurement model of the SEM component

	ICLV Pbc	**ICLV Ptime**	**ICLV Feelings**	**ICLV Full** PBC	Ptime	Feelings
$\bar{z}_1$	0.281 (0.007) ***	0.415 (0.009) ***	1.390 (0.031) ***	0.279 (0.007) ***	0.413 (0.009) ***	1.38 (0.031) ***
$\bar{z}_2$	1.350 (0.023) ***	1.660 (0.031) ***		1.330 (0.023) ***	1.650 (0.030) ***	
α_2	-1.450 (0.080) ***	-0.685 (0.072) ***	0.165 (0.016) ***	-1.440 (0.078) ***	-0.701 (0.074) ***	0.165 (0.016) ***
α_3	-2.260 (0.086) ***	0.560 (0.042) ***	0.238 (0.016) ***	-2.280 (0.090) ***	0.549 (0.043) ***	0.237 (0.016) ***
α_4		-0.139 (0.046) ***	-0.410 (0.022) ***		-0.153 (0.047) ***	-0.409 (0.022) ***
α_5			-0.954 (0.033) ***			-0.954 (0.033) ***
α_6		0.541 (0.033) ***	0.926 (0.024) ***		0.543 (0.033) ***	0.922 (0.024) ***
α_7		-0.109 (0.071)	2.110 (0.058) ***		-0.139 (0.074) ***	2.110 (0.058) ***
α_8			2.400 (0.067) ***			2.410 (0.068) ***
λ_2	1.700 (0.102) ***	2.750 (0.172) ***	0.647 (0.016) ***	1.700 (0.099) ***	2.810 (0.180) ***	0.652 (0.016) ***
λ_3	1.700 (0.111) ***	1.510 (0.096) ***	0.616 (0.016) ***	1.770 (0.118) ***	1.540 (0.098) ***	0.617 (0.016) ***
λ_4		1.440 (0.104) ***	0.558 (0.019) ***		1.480 (0.108) ***	0.563 (0.019) ***
λ_5			0.652 (0.024) ***			0.652 (0.024) ***
λ_6		0.406 (0.066) ***	0.366 (0.016) ***		0.392 (0.067) ***	0.362 (0.016) ***
λ_7		2.450 (0.164) ***	0.528 (0.027) ***		2.530 (0.173) ***	0.527 (0.028) ***
λ_8			0.579 (0.028) ***			0.588 (0.029) ***
σ_2^*	0.783 (0.029) ***	0.866 (0.034) ***	0.590 (0.018) ***	0.804 (0.0242) ***	0.873 (0.034) ***	0.587 (0.018) ***
σ_3^*	0.879 (0.029) ***	1.010 (0.025) ***	0.651 (0.020) ***	0.869 (0.0280) ***	1.010 (0.024) ***	0.652 (0.020) ***
σ_4^*		1.200 (0.025) ***	0.868 (0.024) ***		1.200 (0.0246) ***	0.865 (0.024) ***
σ_5^*			1.190 (0.032) ***			1.190 (0.032) ***
σ_6^*		1.120 (0.024) ***	0.767 (0.023) ***		1.110 (0.023) ***	0.773 (0.0223) ***
σ_7^*		0.938 (0.030) ***	1.140 (0.040) ***		0.931 (0.029) ***	1.160 (0.040) ***
σ_8^*			0.958 (0.041) ***			0.975 (0.041) ***

Notes: Robust standard errors: in parentheses. P-values: ***=sign. at the 1% level; **=sign. at the 5% level; *=sign. at the 10% level. Ptime: Perceived time. The $\bar{z}$ parameters are the thresholds estimated in the ordered probit models (Equation (5.6)). The α parameters are the intercepts of the measurement model (Equation (5.5)), the λ parameters the loadings and the $\sigma^\star$ the standard-deviations of the error terms. For the Model "ICLV Full" the results are given for the three latent variables.

TABLE 5.4 – Estimation results of the structural model of the SEM component

	ICLV Pbc	**ICLV Ptime**	**ICLV Feelings**	**ICLV Full** PBC	Ptime	Feel
Intercept	0.898 (0.076) ***	0.457 (0.066) ***	1.840 (0.189) ***	0.876 (0.074) ***	0.480 (0.068) ***	1.810 (0.196) ***
Age	0.024 (0.033)	-0.0004 (0.029)	-0.928 (0.087) ***	0.023 (0.033)	-0.016 (0.030)	-0.925 (0.091) ***
Age^2	-0.005 (0.004)	-0.0003 (0.003)	0.099 (0.009) ***	-0.005 (0.003)	0.001 (0.003)	0.100 (0.010) ***
Child	0.002 (0.022)	-0.027 (0.017)	0.130 (0.051) ***	0.001 (0.022)	-0.030 (0.018) *	0.099 (0.054) *
Cars	-0.090 (0.014) ***	-0.038 (0.011) ***	-0.195 (0.033) ***	-0.088 (0.014) ***	-0.038 (0.012) ***	-0.197 (0.032) ***
Income_h	0.008 (0.021)	-0.002 (0.017)	0.014 (0.051)	0.003 (0.021)	-0.007 (0.018)	0.006 (0.053)
Gender	-0.012 (0.019)	-0.045 (0.015) ***	0.204 (0.043) ***	-0.006 (0.019)	-0.0376 (0.015) ***	0.210 (0.046) ***
ω	0.468 (0.027) ***	0.417 (0.026) ***	1.260 (0.042) ***	0.454 (0.027) ***	0.412 (0.027) ***	1.280 (0.043) ***

Notes: Robust standard errors: in parentheses. P-values: ***=sign. at the 1% level; **=sign. at the 5% level; *=sign. at the 10% level. The ω parameters are the standard-deviations of the error term of the structural model (Equation (5.4)). For numerical reasons, Age and Age^2 are respectively divided by 10 and 100.

Another robust result across all specifications is that car users have lower probabilities of choosing a public transport mode. From Models "MNL1" and "MNL2", we also find that the age, the number of cars in the household and being a man decrease the probability of choosing the train alternative relative to the car alternative. People living with children have a lower probability of choosing the coach alternative (only in "MNL2") and people living in a household with a high income are more likely to choose the coach than the car alternative. The lessons from the literature, which are summarized by De Witte et al. (2013), are partly in line with these results. Firstly, there seems to be no real consensus in the literature on the impact of age or gender on mode choice. Our results can therefore be neither confirmed nor infirmed by the literature. Secondly, car availability is a very important determinant influencing mode choice. Our finding is consistent with the idea that the higher the number of cars available to the household, the less competition there is for cars among the household members, and consequently, the lower the tendency to use public transport modes. Thirdly, the presence of children increases the utility of car use, which has a significant negative impact on public transport use. This is consistent with our finding that children decrease the probability of choosing the coach alternative. Fourthly, income is generally found to entertain a positive relationship with car use and an inverse relationship with public transport use. The positive relationship we found with coach choice is therefore surprising and will need to be further explored.

Given the empirical specifications, these socio-economic characteristics also play a role in models with latent variables but in an indirect way. Indeed, we assume that the socio-economic characteristics determine latent variables, which, in turn, determine levels of utility for modes of transport. With regards to the fixed effects of the latent variables, the three models with only one latent variable show that the latent variables increase the probabilities of choosing public transport modes. Whether we consider *PBC*, *Perceived Time* or *Feelings* as psychological constructs that drive mode choice, any improvement in these variables increases the probability of choosing a public transport mode. Simply stated, travelers who perceive traveling with public transport as an easy task and who are not bothered by traveling with strangers or with children or luggage, have higher probabilities of choosing public transport modes than other people. Similarly, travelers who perceive that their time spent traveling by public transport is worthwhile or who experience positive feelings during travel by public transport are more inclined to choose the coach or train alternatives.

The effects of these latent variables on interaction with travel time are, however, more mixed. When considering *PBC* as the latent variable affecting utilities for modes of transport (Model "ICLV Pbc"), we find that it does not affect sensitivity to travel time. The results show that the *PBC* latent variable

affects utility for public transport modes only as a fixed effect. One may think that better behavioral control (e.g., low agoraphobia, ease in using public transport modes with shopping or luggage items, smooth traveling with children, and so on) decreases sensitivity to travel time, as it reduces anxiety when traveling by a public mode of transport. The data do not confirm this hypothesis: *PBC* appears to be a generic motive for using public transport modes but it does not change how time is perceived by travelers of whatever kind. *PBC* has to be understood in terms of prior beliefs about the ease of traveling by public modes of transport but not as a driver of sensitivity to travel time.

Note also that the items used to measure this latent variable only address a general resistance to using train or coach but say nothing about what is actually experienced during a trip experience. The associated measurement equations show that travelers with higher levels of *PBC* have higher probabilities of rating public transport modes favorably for all items. Likert scales were recoded prior to estimation to model larger outcomes as rejections of negative statements and acceptance of positive statements.

In explaining the latent variable *PBC* by individual characteristics, in addition to the intercept and the standard deviation, the only significant variable is the number of cars in the household. Having direct access to this mode of transport decreases the overall perception of confidence and convenience in using public transport, which decreases the probability of using these modes.

As stated before, the latent variables (*Perceived time* and *Feelings*) are focused on latent psychological constructs that explain subjective factors affecting trip experience using a public mode of transport. The "ICLV Ptime" Model shows that the *Perceived time* latent variable increases the utility of the public transport mode and, therefore, the probability of choosing these modes, both as a fixed effect and in interaction with travel time. Thus, the VOT decreases as the perception of a worthwhile travel time increases. The same result is observed when considering instead the *Feelings* latent variable in the "ICLV Feelings" Model. The more positive the feelings experienced during a trip by public transport, the higher the probability that those modes will be chosen and the lower the VOT.

As for *PBC*, the measurement equations show that the higher the values of the *Perceived time* and *Feelings* latent variables, the higher the probabilities of answering high outcomes for all items. Travelers who have more positive feelings and/or a better time use for activities during travel will have a higher probability of disagreeing with negative statements and a higher probability of agreeing with positive statements.

With the exception of the intercept and the variance of the error term, the significant socio-economic characteristics that explain the *Perceived time* latent variables are the household motorization rate and

the gender of the traveler. Therefore, having more cars in the household decreases the level of perceived fruitful time. This has two effects: it makes the traveler more sensitive to travel time by public transport modes and it also decreases the baseline level of utility for public transport modes. Being a man also yields the same effects.

Regarding the *Feelings* latent variable, men are less sensitive to inconveniences than women when traveling by a public transport mode. Their baseline level of perceived comfort is higher than women's when it pertains to the collective use of public transport modes. Household motorization rate, not having children, and age also have significant negative effects on the *Feelings* variable.

Some, yet not all, results are robust for the ICLV model that includes all three latent variables (ICLV Full). Firstly, the only latent variable which directly impacts utility is the *PBC* variable. Secondly, the only latent variable which has an effect on sensitivity to travel time is the *Feelings* variable. These results can be interpreted as follows: *PBC* generates a generic acceptability of public transport, which affects the probability of travelers choosing these modes, and, once on board the train or coach, the feelings they experience play a role in sensitivity to travel time. Despite some differences in the significance levels, the results of the structural and measurement models of the SEM are robust across the three separate ICLV models and the full ICLV model.

5.4.2 Values of Time

Aside from estimated VOTs, scenarios are defined to simulate the sensitivity of the VOT to specific public policies (see Table 5.5). In addition to a change in comfort in terms of seat availability, the idea is to simulate how the latent variables change following demographic changes in the population of interest and how that, in turn, impacts VOT. Given the structure of the models and the variables that we use, such changes can be related to three types of public policy:

- public policies may act on car ownership levels, e.g., discourage vehicle purchase and encourage other travel alternatives (including car sharing or car pooling);
- public policies may act on the public transport side and, more specifically, on comfort by improving infrastructure (seats, Wi-Fi, intimacy, places for luggage and children, and so on) to enhance the public transport experience and therefore *PBC*, *Perceived time* and *Feelings*;
- public policies may act directly on perceptions with appropriate communication and advertising campaigns. For example, the "Gender" scenario (see hereafter) studies how behavior would change if women behaved like men. Depending on the results, it may indicate a specific group of persons who should be the focus for communication campaigns.

The first scenario is the baseline scenario: VOTs are estimated according to the observed sample for the latent variables and no guarantee of a seat in the trains (comfort=0), as is the case during peak hours for the main train lines of the Rhône-Alpes Region. The second scenario is the "Seat" scenario, which also considers the observed values for the latent variables and the guarantee a seat in the trains (comfort=1). Note, however, that the VOT when a seat is not guaranteed (comfort=0) in the trains is the same as the VOT for the coach alternative. This observation remains true for the four following scenarios which simulate changes in the latent variable values. The following scenarios therefore consider the seating position as guaranteed in the trains and the VOT with no seat guaranteed is simply derived from the coach VOT. The third scenario is the "Max LV" scenario: what would happen if everyone reached the (in-sample) maximum of overall self-confidence in using public transport (*PBC*), potential productive (in its broad meaning) time use during travel by public transport (*Perceived time*), and best customer experience when using public transport for travel (*Feelings*). This scenario provides insight into the effects of a high quality service at constant price-time-comfort(seat) levels. The fourth scenario is the "Cars" scenario. We simulate a decrease of one car per household.[11] The fifth scenario is the "Gender" scenario. We here imagine that there is no gender difference in latent psychological constructs related to mode choice behaviors. The sixth and last scenario is the "Age" scenario: travelers' age is reduced by 20 years (new age = max(age-20, 18)).

TABLE 5.5 – Scenarios for simulation

Scenario	**Description**
1 : Baseline scenario	All variables are set to their initial value, except the objective comfort set to 0 (no guarantee of a seat in the trains). Serves as a comparison.
2 : Seat scenario	All variables are set to their initial value, except the objective comfort in the trains set to 1 (guarantee of a seat in the trains).
3 : Max LV scenario	Latent variable is set to the highest value observed in the sample. Useful to analyze the potential for development of public transport if perceptions of comfort evolve. Comfort in the trains set to 1.
4 : Cars scenario	For each traveler, reduces the motorization of the household by one. Comfort in the train set to 1.
5 : Gender scenario	Considers the behavior of women as the same as men. Comfort in the train set to 1.
6 : Age scenario	Considers that the behavior of older people becomes the same as younger people (with new_age = max(age-20,18)). Comfort in the train set to 1.

We recall that the utility functions (Equation (5.10)) are defined so that the VOT for the car is invariant with respect to the scenarios, and the VOT for train and coach only differs in that the "seat

11. In the sample, all households have at least one car.

guaranteed / seat not guaranteed" comfort variable is specific to the train alternative, since the objective "Comfort" variable is not in the utility function for the coach alternative. This means that sensitivity to travel time is the same for both travel with no guarantee of a seat in a train and travel with a seat in a coach. Below, we discuss, firstly, the extent of the VOTs for the scenarios that set the latent variables to their initial values. Secondly, we discuss the extent to which these values change according to the objective comfort variable. Thirdly, we discuss how changes in the latent variables impact VOTs.

In the scenarios that set the latent variables to their initial values ("Baseline" and "Seat" scenarios), the mean VOT is estimated to range from €10.39/hour to €15.40/hour, depending on the mode, the guarantee or not of having a seat in the train, and the model (see Table 5.6). These values are consistent with the existing literature. In a SP survey, Arentze and Molin (2013) find values between €14.4/hour and €17.4/hour for travel by train. Glerum et al. (2014) estimate VOTs to be around CHF 12/hour (€10/hour) for travel by train or coach in Switzerland. In their European meta-anaysis, Wardman et al. (2012) find that, in France, the VOT of train commuters is between €4.5/hour and €9.4/hour. Although our results are in line with the values of this meta-analysis, they seem to stand on an upper bound. The French official values of VOT for interurban travel are heterogeneous (Quinet et al., 2014). Depending on the distance traveled (from less than 20 km to 400 km), they range from €7.9/hour to €15.2/hour for car, from €7.9/hour to €28.0/hour for coach, and from €7.9/hour to €26.2/hour for train.

In the models that account for the influence of the objective comfort on the sensitivity to time (all models, except "MNL1"), the VOT for the car alternative (€13.12/hour to €13.71/hour) is lower than the VOT for the "coach" alternative or VOT for the "train (comfort=0)" alternative. However, the VOT for the car alternative is always higher than the VOT for the "train (comfort=1)" traveling situation. This means that, when a seat is guaranteed in trains, then VOTs of train become lower than VOTs of the car alternative. We find that, if travelers can be sure of having a seat, they have $\text{VOT}_{\text{train}} < \text{VOT}_{\text{car}} < \text{VOT}_{\text{coach}}$. This ranking is not sensitive to either our different modeling assumptions or scenarios; see Table 5.6 and Figure 5.2, which plots VOT along with estimated ICLV models for the different scenarios. Since, by construction, sensitivity to travel time is the same for travel by train with no guarantee of a seat and for travel by coach with a seat, travelers are willing to pay the same to save travel time whether they use a coach with a seat or a train with the risk of having to stand. This assumption has been tested in preliminary models [12] and we could not reject the fact that train and coach alternatives have an equal parameter for time.

12. Full results are available upon request.

TABLE 5.6 – Value of time and market shares according to mode, objective and perceived comfort

Model	Scenario	Value of Time (in €/hour)			Modal share (in %)		
		(Train with comfort=1)	(Coach) or (Train with comfort = 0)	(Car)	(Train)	(Coach)	(Car)
MNL1		14.05		14.42	0.33	0.22	0.45
MNL2		11.12	15.38	14.05	0.33	0.22	0.45
ICLV PBC	Baseline	15.21	15.21	13.71	0.27	0.25	0.49
	Seat	10.85	15.21		0.38	0.19	0.42
	Max LV	10.79	15.15		0.40	0.20	0.40
	Cars	10.86	15.16		0.40	0.20	0.40
	Gender	10.86	15.22		0.38	0.19	0.42
	Age	10.83	15.19		0.39	0.20	0.42
ICLV Ptime	Baseline	14.67	14.67	13.12	0.27	0.25	0.48
	Seat	10.39	14.67		0.38	0.19	0.42
	Max LV	10.24	14.51		0.39	0.2	0.42
	Cars	10.29	14.57		0.39	0.19	0.42
	Gender	10.45	14.72		0.38	0.19	0.43
	Age	10.38	14.65		0.38	0.19	0.42
ICLV Feelings	Baseline	15.40	15.40	13.32	0.27	0.25	0.48
	Seat	11.26	15.40		0.38	0.19	0.42
	Max LV	9.40	13.53		0.43	0.21	0.35
	Cars	10.97	15.10		0.33	0.23	0.44
	Gender	11.12	15.25		0.33	0.23	0.44
	Age	10.03	14.17		0.35	0.24	0.41
ICLV Full	Baseline	15.36	15.36	13.28	0.27	0.25	0.48
	Seat	11.20	15.36		0.38	0.19	0.43
	Max LV	9.21	13.36		0.43	0.21	0.36
	Cars	10.94	15.10		0.40	0.20	0.40
	Gender	11.04	15.19		0.38	0.19	0.42
	Age	9.91	14.10		0.41	0.2	0.39

Notes: The scenarios are those defined in Table 5.5. The VOT for (Train with comfort=0) is equal to the VOT for (Coach); see Equation (5.10). In MNL1, $\beta_5 = 0$ (see Equation (5.10)), the VOT therefore does not depend on the objective comfort and is the same for (Train with comfort=0), (Train with comfort=1) and (Coach).

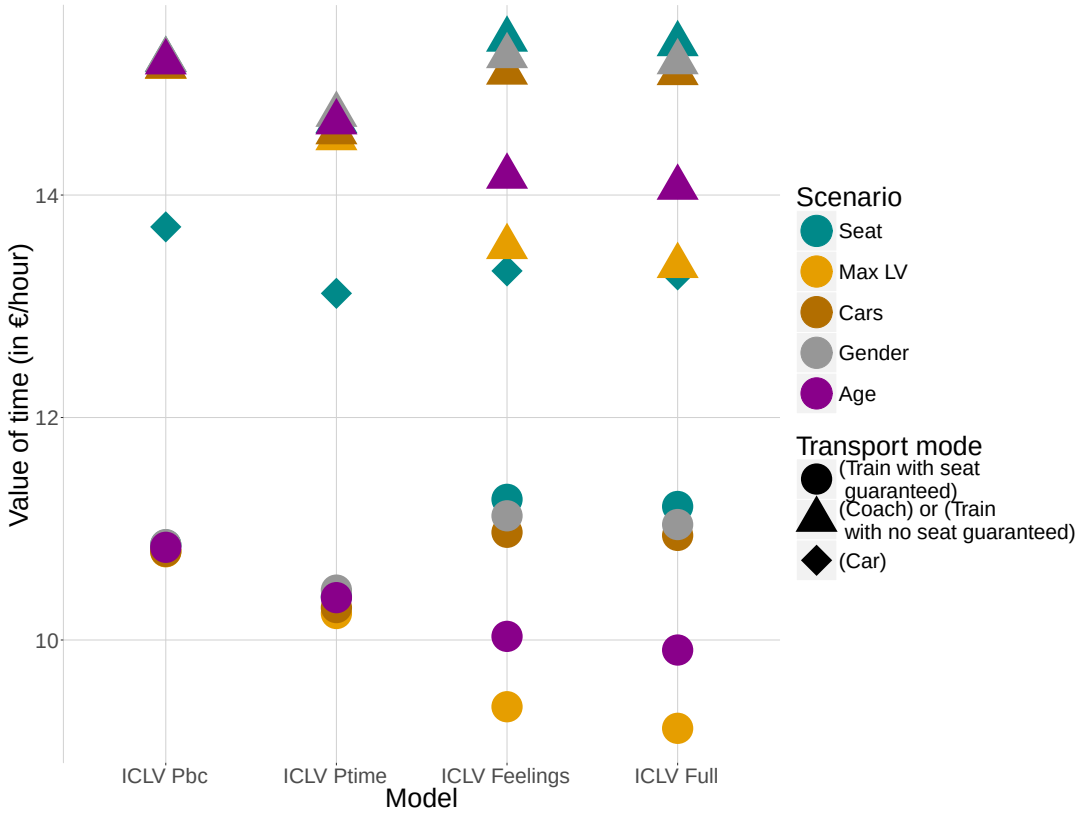


FIGURE 5.2 – VOT heterogeneity according to mode, comfort and latent variables
Notes: The VOTs for the baseline scenario are not displayed, as they are combined with those of the "Seat" scenario for the coach alternative.

Traveling seated by train yields significant reductions in VOT: approximately 30%, from €4.1/hour to €4.6/hour. This is consistent with a British literature review, by Wardman and Whelan (2011) who find that, with a load factor of 100%, VOT has to be diminished by 33% when the user has a seating position instead of a standing position. However, we do not take into account the loading of the trains and consider only that a seat is not guaranteed and not that the traveler is sure to be standing.

For a precise insight into the range of VOTs for the "Seat" scenario, Figure 5.3 displays a boxplot with the bootstraped VOTs simulated with the "ICLV Full" Model. For each individual, a VOT is simulated for the three transport modes, then the quartiles of the individual VOTs are calculated. VOTs vary according to the individual latent variable values. According to Figure 5.3, 50% of the sample has a VOT comprised between €10.92/hour and €11.51/hour for the train alternative with a seat guaranteed. For the coach alternative or the train alternative with no seat guaranteed, 50% of the sample has a VOT comprised between €15.08/hour and €15.67/hour. These results show that although the latent variables play a role in the heterogeneity in VOT, the estimated values remain relatively close to each other. This result is further discussed hereafter.

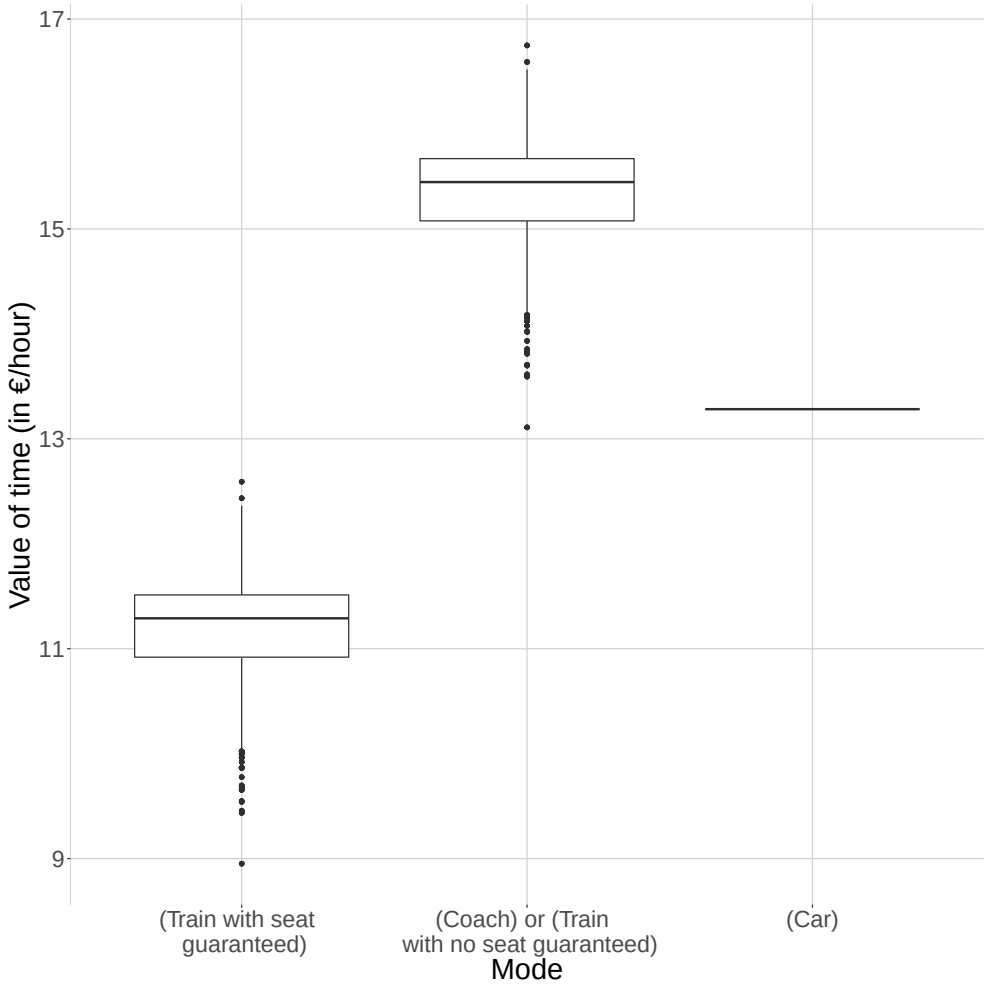


FIGURE 5.3 – Boxplot of VOT estimated with the "ICLV Full" Model

Notes: The "ICLV Full" Model includes the three latent variables. Latent variables: set to the (in-sample) observed value. VOT for train and coach differ in that the "seat guaranteed / seat not guaranteed" comfort variable is specific to the train alternative. VOT for car is constant (depends neither on the objective comfort, nor on the latent variables). The bold line is the median. The ends of the box show the first and third quartiles. The ends of the whiskers show the 1.5 inter-quartile ranges. The points are the outliers.

Turning to the effects of the latent variables on VOT, we find that only the "ICLV Ptime", "ICLV Feelings" and "ICLV Full" models are those where sensitivity to travel time is a function of the associated latent variables (see Figure 5.2). As discussed in Section 5.4.1, the "*PBC*" latent variable does not significantly affect sensitivity to travel time. The small variations in VOT estimates for the different scenarios when applying the "ICLV PBC" Model are only due to simulation noise.

VOTs are not very much affected by the different scenarios, with few exceptions. Only scenarios "Max LV" and "Age" have noticeable results on VOTs when considering the "ICLV Feelings" Model, that is, when assuming that the *Feelings* latent variable is the one that has causal consequences on mode choice. Firstly, the results of the "Max LV" scenario show that public policies that promote optimal well-being in travelers when they use the train (less stressful environment, less sensation of harassment, more seamless travel, convenience, cleanliness, security, quietness, and so on) produce significant reductions in VOT of approximately 12% - by coach or when no seat is guaranteed by train - and 17% by train with a seat, that is, in both cases, around €1.9/hour. For the "ICLV Feelings" Model, having a guaranteed seat during travel by train is at least twice as valuable as a series of policies that improve individuals' well-being while traveling. The same results apply for the "ICLV Full" Model. In terms of effectiveness strategies, it therefore seems preferable to optimize and target investments in seat capacity supply and train operations. Secondly, the scenario "Age", that is, making older travelers behave the same as young ones, yields a reduction in the VOT of €1.2/hour to €1.3/hour. Again, the same results apply for the "ICLV Full" Model.

Although the "ICLV Ptime" Model estimates show that the potential for activities during travel reduces sensitivity to travel time, there is no significant variation of VOT estimates according to the different scenarios.

Unfortunately, mean VOT estimates are not very sensitive to the other scenarios. For instance, one would have expected car ownership restriction policies to influence the VOT of public transport modes. This is not the case whatever latent variable is considered.

5.4.3 Market shares

With regard to market shares, Table 5.6 shows that, according to the ICLV models, if no seat were guaranteed in trains, then the car market share would be 48% to 49% depending on the Model. If a seat were always guaranteed in trains, then the car market share would be 42% to 43%, which is six to seven points lower. Since having a seat is valuable for comfort reasons but also to enable a worthwhile use of time, these values can be compared to those of Malokin et al. (2015). Using revealed preferences data

from Northern California commuters, they estimate that commuter rail and car/vanpool shares would respectively be 0.38 and 3.22 percentage points lower, and the drive-alone share 3.00 percentage points higher, if the option to use time productively while traveling were not available. Our values are higher, certainly because of the use of SP data.

With regard to the effects of variations of the latent variables on market shares, we only interpret them on the two ICLV models with *Feelings* as a latent variable, that is, "ICLV Feelings" and "ICLV Full" Models because they are the most informative models. If, in addition to the guarantee of a seat, all travelers had positive feelings during travel by public transport (see the "Max LV" scenario), then the car market share would drop by seven points and reach 35%. For the "ICLV Full" Model, this drop may theoretically be due to the increase of the other latent variables (*Perceived time* and *PBC*). However, given the minor effects of the "Max LV" scenario on the market shares estimated with the "ICLV Ptime" and "ICLV PBC" Models, we conjecture that it is the *Feelings* variable which yields the observed variations. The other scenarios only have low effects on the modal shares.

With the other ICLV models, the minor effects following a change in the value of the latent variable are consistent with the low impact of the *PBC* and *Perceived time* latent variables on the VOT, and the absence of direct effect of the *PBC* variable on the utility functions.

5.5 Conclusion

This paper investigates how objective and perceived comfort in public transport influences mode choice and VOT. ICLV models are used to study behavioral heterogeneity in some depth by means of the inclusion of latent variables that model the perceived comfort.

The main results are that objective comfort, that is, having a seat guaranteed in trains, and individual feelings during travel by public transport (coach and train) robustly explain variations in VOT. The perceived use of time during travel by public transport also explains heterogeneity in VOT when considered alone but its effect vanishes when considered simultaneously with the experienced feelings. Another robust result is that people who perceive the use of public transport modes as difficult, because they feel they have low control over its use, due to the presence of other people or the difficulty of traveling with shopping or luggage items, are more likely to use their car than public transport modes. When considered separately, the feelings experienced during travel by train or coach, as well as the perception of a worthwhile use of public transport travel time, also explain the choice of public transport modes. However, when all latent variables are considered simultaneously,

only the perceived behavioral control influences mode choice as a fixed effect.

In terms of public policy, the lessons learned from this paper can be summarized as follows. Tangible improvement of the infrastructure, that is, an increase in seat capacity supply, as well as soft measures, that is, conditions to ensure that all travelers have a pleasant experience during travel time by public transport instead of being "hostage" to journey time, both yield economic benefits, in that the two types of measures lead to lower VOT. However, the first measure is much more efficient than the second one, since the the variations in VOT are twice as high. This means that people are more sensitive to physical, direct improvements than to less palpable ones. However, both measures yield a similar drop in the market shares of the car. The comparison of the effectiveness of the measures therefore depends on how the effectiveness is measured (individual economic benefit or overall use of the car). With regard to the cost of these measures, we can conjecture that the marginal cost function of increasing the positive feelings of travelers is upward sloping. For a few travelers, not much modification is needed to improve their travel experience. However, improving the feelings of all travelers is likely to be very costly, or even impossible to implement, since some people are intrinsically opposed to traveling with people they do not know. A recommendation is therefore to combine some non-costly soft measures, based, for instance, on the "travel remedy kit" imagined by Watts and Lyons (2010), with an increase in the capacity or the frequency of trains during peak hours in order to ensure seat availability.

Another lesson of this work, but also of previous research on VOT heterogeneity, is that VOT is highly dependent on individual characteristics. However, while VOT is considered as one key parameter in evaluating transportation investments and related policies to alleviate congestion and other externalities, current practice, at least in some European countries (Mackie et al., 2003; Quinet et al., 2014), still remains based on some representative VOTs, yet accepting that only very limited variations might exist in space and across sub-population of travelers (mostly according to trip purposes and modes). To that extent, there is a room for improvements that take behavioral heterogeneity into account in the economic appraisal of transport projects. Researchers, however, including ourselves in this paper, focus on one or only a few distinct and clear features about valuation of VOT, while an integrative framework that took account of every dimension involved in VOT heterogeneity (socio-economic characteristics, attitudes, spatial indicators, stage of the journey, quality of service, and so on) would be needed. More generally, a broad avenue of research would be to develop a unified tool that allowed for the integration of advanced behavioral modeling of travel choices with a global evaluation of transport projects.

CHAPTER 6

How does environmental concern influence mode choice habits? A mediation analysis[1]

Abstract

Starting from the intuition that people with high environmental concern have a better perception of public transport and therefore a better perception of the utility of public transport, we construct a theoretical model in which the effect of environmental concern on mode choice habits is mediated by the indirect utility of travel. Travel procures the direct utility of providing access to activities, but it also offers an indirect utility that is inherently personal and perceptual. We approach the indirect utility of public transport by measuring perceptions of time and feelings. The indirect utility of the car is approached by measuring affective and symbolic motives. Taking into account car use habits and habits of public transport use, the results show that people who have a high environmental concern perceive public transport use as easier, more useful and more pleasurable than people who do not have that environmental motivation. Such positive attitudes foster public transport use. Conversely, low environmental concern generates non-instrumental motives for car use, such as affective and symbolic motives. However, the relationship between affective and symbolic motives and car use habits is not robust. We can conclude that environmental concern influences mode choice habits and that the effect is partially mediated by perceptions and feelings towards public transport but not significantly by affective and symbolic motives for car use.

1. This chapter is a paper that has been submitted to *Transportation Research Part D* (Bouscasse H., Joly I., Bonnel P.). It has been reviewed and is currently under revision.

6.1 Introduction

In the European Union, the car continues to be the most popular mode of transport. In 2013, passenger cars accounted for 83.2 % of inland passenger transport[2] (Eurostat, 2016) despite the negative external costs involved (air pollution, noise, space usage, and so on). In response to these issues, a growing awareness of the ecological challenges associated with car use is informing the public debate and promoting more sustainable mobility. Yet, the relationship between environmental concern and travel behavior is still a challenging question to which contradictory answers have been given. While a correlation between environmental concern and environmentally-friendly behavior is observed in other fields, such as waste management (Barr, 2007), energy efficiency (Ramos et al., 2016) or water conservation (Millock and Nauges, 2010), the link seems to be more nuanced in the field of transportation. In some studies, the relationship between environmental concern and public transport use, or intention to use, is positive and thus confirms intuition (e.g., OECD, 2014; Garvill, 1999; Nilsson and Küller, 2000). However, other studies do not find a significant relationship (e.g., Walton et al., 2004; De Groot and Steg, 2007), which suggests a potential inconsistency between attitudes towards the environment and mode choice behavior.

An explanation for this cognitive dissonance (Festinger, 1954) is that pro-environmental motivation is not enough for most people to engage in behavior that replaces or reduces private car use to the benefit of sustainable travel modes. In other words, modal shift is not easy enough. This is supported by the theory of planned behavior (Ajzen, 1991), which, through the concept of Perceived Behavioral Control (PBC), states that people are more likely to perform a behavior when they perceive it as easy to perform. To better understand the link between environmental concern and travel behavior, attitudes to travel modes need to be explored as well as the extent to which their use is perceived as easy, pleasant and useful by passengers.

In the theory of planned behavior, the traveler's attitude towards a behavior is portrayed as a cognition process where the rational consumer evaluates pros (positive utility) and cons (disutility) associated with the behavior in question. According to Mokhtarian and Salomon (2001), the positive utility associated with traveling with a specific mode is gained from 1) accessing the activities conducted at destination; 2) activities that can be conducted while traveling, and 3) the activity of traveling itself. The second and third elements reflect an indirect utility arising from traveling, which may explain "excess travel" and unexpected choices of travel modes. An illustration is given by

2. As measured by the number of passenger-kilometers traveled by each mode

Morris (2015), who finds that travel time generally exhibits a positive relationship with life satisfaction. One of his explanations is that the act of travel itself is fulfilling.

To our knowledge, previous works concerned with the relationship between environmental concern and mode choice do not take into account the influence of the indirect utility of travel on mode choice. However, individuals with a higher environmental concern may be more sensitive to the indirect utility provided by public transport and less sensitive to the indirect utility provided by the car. Thus, indirect utility, as well as PBC, may act as a mediator between environmental concern and mode choice habits.

Moreover, previous works (e.g., Walton et al., 2004; Nilsson and Küller, 2000; De Groot and Steg, 2007) mainly focus on single measures (e.g., intention to use a single mode, one-off choice or mode choice for a single purpose), which depend on the journey characteristic indicators such as travel cost or travel time. However, environmental concern and attitudes towards travel modes are not specific to a particular journey. Furthermore, some people use diversified modes of transport and adapt their choice to the context of the trip. In their literature review on the definition and measurement of multimodality, Molin et al. (2016) emphasize the advantage of choosing a wide time window (a year) to capture the multimodal dimension of mobility. An individual may indeed travel by car for commuting purposes but regularly use train for leisure or business trips. The impact of environmental concern on travel behavior thus has to be approached with a general measure, such as mode choice habits, to take into account three dimensions of mode choice: repetitiveness, intensity of travel (excess travel) and trade-off between modes.

The aim of this article is thus to extend previous research by means of an integrative tool (Structural Equation Modeling; SEM). SEM makes it possible to explain a multiple output, car use and public transport habits, with a mediation structure between environmental concern, PBC and the indirect utility of traveling.

The rest of the article is arranged in five sections. Section 6.2 sets out the background and defines the theoretical model and concepts. Section 6.3 proposes a number of hypotheses and suggests a theoretical model to test them. The collected data and the estimation method by SEM are presented in section 6.4. The results of the estimations are discussed in Section 6.5. Section 6.6 offers conclusions.

6.2 Literature review

A growing body of literature lies at the interface between environmental and behavioral economics (Croson and Treich, 2014) and is based on psychological concepts and theories that focus directly on

environmental issues, including travel mode choice.

6.2.1 Environmental psychology

Two behavioral theories, the theory of planned behavior and value-belief-norm theory, are critical to explain pro-environmental behavior, including transportation behavior. They are based on the idea that behavior is driven by internal mental states rather than external conditions, with the assumption that behavior is the outcome of a deliberative conscious process (Savage et al., 2011).

The theory of planned behavior, an extension of the theory of reasoned action (Fishbein and Ajzen, 1975), assumes that individuals behave strategically, evaluating the information they have access to and the implications of those actions. Individuals thus choose alternatives with lowest costs against highest benefits. The theory of planned behavior involves attitudinal dimensions (the degree to which engagement in behavior is positively valued), subjective norms (social pressure from important others to engage in a particular behavior) and PBC, and relates them to the intended behavior that precedes actual behavior. The theory of planned behavior is generally successful in explaining mobility behavior (Heath and Gifford, 2002; Bamberg et al., 2003b; De Groot and Steg, 2007).

Pro-environmental behavior is often associated with higher costs. Individuals are therefore more likely to act pro-environmentally when they subscribe to altruistic or biospheric values beyond their own immediate interests. These values are integrated in value-belief-norm theory developed by Stern et al. (1999) based on a major study on the psychological determinants of pro-environmental behaviors (Stern, 1992). Value-belief-norm theory is based on Schwartz's norm-activation-model (Schwartz, 1977), which stipulates that a behavior is performed by individuals who feel a moral obligation to do so. These personal norms depend on the extent to which individuals are aware of the problems caused by their behavior and feel responsible for the problems and solutions. Value-belief norm theory further proposes that problem awareness is rooted in an ecological worldview, itself determined by personal values. [3] Low-cost environmental behavior (occasional recycling, water and electricty savings) or intentions are well explained by the norm-activation models and value-belief-norm theories (Gärling et al., 2003; Steg, 2005). Yet, in the presence of high behavioral costs or strong constraints on behavior, such as public transport or bike use, the explanation power of these theories is low (Bamberg and Schmidt, 2003). In such contexts, the theory of planned behavior appears to be more powerful in explaining behavior, probably because it takes into account non-environmental motivations as well as PBC (Gifford et al., 2011).

3. Social-altruistic value: the environment is valued for what it brings society. Egoistic value: the environment is valued for what it brings me. Biospheric value: the environment is valued for itself, independently of its value for humankind.

In the theory of planned behavior, PBC takes account of the influence of contextual factors as they are perceived by the user. Objective contextual factors may also be integrated, as in attitude-behavior-context theory (Guagnano et al., 1995), which assumes that a particular behavior occurs when an individual's attitudes are positive and the external conditions are favorable. In attitude-behavior-context theory, contextual factors act as moderators on the attitude-behavior relationship. Using regression models, Collins and Chambers (2005) show that the effect of psychological factors (PBC, biospheric beliefs and so on) on behavior (the intention to use public transport for commuting trips) are indeed moderated by contextual variables.

Neither of these behavioral models includes the role of habits. A behavior driven by habits tends to occur subconsciously, without the person having to consciously think much before performing it. The stronger the habits, the less important the link between attitudes (and other explanatory variables) and behavior. Habits are accounted for by the theory of interpersonal behavior (Triandis, 1977). In this model, habits, intention and contextual factors are the three key factors explaining behavior.

6.2.2 Environmental Concern

There is currently no stable and consensual definition of environmental concern (Le Borgne et al., 2015), which is a concept that carries a wide array of attitudes about environmental issues. These differences in definitions, and thus in measures, may explain the contradictory results on the relationship between environmental concern and travel behavior.

Two examples of definition for environmental concern are: 1) "an individual's insight that humans endanger the natural environment combined with the willingness to protect nature" (Franzen and Vogl, 2013, p. 2), and 2) an inclusive concept, "ranging from a specific attitude toward environmentally relevant behaviour to a more encompassing value orientation" (Fransson and Gärling, 1999, p. 370). A complementary approach is to define environmental concern based on the three dimensions of an attitude: affective, cognitive and conative (Franzen and Vogl, 2013). A person with high environmental concern can therefore be defined as someone who has a good knowledge of (or belief in) the environmental challenges (cognitive dimension), who feels concerned by those challenges (affective dimension) and who has intentions to act to protect the environment (conative dimension).

There are over 700 scales for measuring environmental concern, but the scale most widely used is the New Ecological Paradigm (NEP) Scale (Schleyer-Lindenmann et al., 2014; Dunlap, 2008). This scale has its origins in the 1970s (Dunlap and Van Liere, 1978). In order to address the criticisms acknowledged by the authors themselves and reported in Anderson (2012) (lack of internal consistency, poor correlation

between the scale and behavior, and the "dated" language of the items), Dunlap et al. (2000) updated the original scale. The NEP scale includes three facets: balance of nature, limits to growth, and anti-anthropocentrism (Dunlap, 2008). In value-belief-norm theory, the NEP is the scale used to measure beliefs about human-environment relations (ecological worldview).

Research has examined the individual determinants of environmental concern such as gender, age, education or income with mixed results (Marquart-Pyatt, 2012). Previous studies have proved the effect of age on environmental concern (see Buttel, 1979, for a discussion on the comparative effects of age and education and Liu et al., 2014, for a literature review). Jones and Dunlap (1992) demonstrated that, among 11 socio-demographic and socio-political factors, age was the best predictor of environmental concern, with younger people being more environmentally concerned relative to older people. Literature reviews on the relationship between gender and environmental concern conclude that women tend to have higher levels of environmental concern than men (Liu et al., 2014; Waygood and Avineri, 2016). Three interrelated theories may explain the gender gap (Liu et al., 2014): different expectations for females and males during parenthood and socialization processes; gender-based divisions in the labor market and home, and different value formation processes between women and men.

6.2.3 Indirect utility of car use: affective and symbolic motives

The car is generally perceived as a flexible and low-cost mode of transport that provides rapid access to activities. Such motives for car use, described as "instrumental", are those most widely studied in the literature. However, Steg (2005) argues that behavior models that focus on the instrumental motives for car use are not enough to explain the predominant choice of the car. She builds on the notion of indirect utility and shows that accounting for affective and symbolic motives for car use allows for a deeper understanding of mode choice.

Affective motives for car use account for the pleasure derived from a car, not by virtue of reaching a destination but by virtue of the object (e.g., the enjoyment of owning a fine car) and of using it (e.g., the sensation of speed and movement in Mokhtarian and Salomon, 2001).

Symbolic motives for car use focus on self-image. The mode choice depends, then, on social pressure, on how others see us and on our perception of how others see us. According to Steg (2005), the car can thus be seen as a symbol that reveals who one is (Schlenker, 1980), that allows a person to compare himself with other people (Festinger, 1954) and that allows a person to comply with the norms (Cialdini and Reno, 1991).

Lois and López-Sáez (2009) estimate a structural equation model to analyze the influence of symbolic, affective and instrumental motives for car use. Their results partially coincide with those of Steg (2005), despite differences in the measures of variables. According to their model, neither instrumental nor symbolic motives directly predict car use. Affective motives are the key factor in the prediction of car use and, in turn, they are accounted for by instrumental and symbolic motives.

6.2.4 Indirect utility of public transport use: perceptions of time use and feelings experienced during travel

As for the car, motives for using public transport are diverse. Besides instrumental motives determined by the objective level of service variables (travel cost, travel time, frequency and so on), there is a growing awareness that public transport demand is also driven by perceptual variables. Perceptions may even matter more than reality in travel decisions (Bhattacharya et al., 2014; Kaufmann, 2002). According to Steg (2003), people who usually do not drive think that the car is less safe, more costly, less comfortable and that it offers fewer varied experiences than travel by public transport.

These very notions of comfort and experience are at the heart of a strand of literature concerned with activities, perceived time, and feelings experienced during travel by public transport. In public transport modes, travel time can be used to read, work or rest, for instance (Lyons et al., 2013; Wardman and Lyons, 2016; Joly and Vincent-Geslin, 2016). The digital revolution further reinforces travel-based multitasking and an economically or socially productive use of travel time (see Berliner et al., 2015; Malokin et al., 2017, for an overview on multitasking).

Depending on the user, travel time is perceived as either wasted time or as productive time (Lyons et al., 2013). Public transport users seem to increasingly judge their travel time as worthwhile. In the UK, the proportion of train users considering their time wasted went down by nearly a third between 2004 and 2010 from 19% to 13% of all passengers (Lyons et al., 2013). Correspondingly, the proportion of people making very worthwhile use of their time went up by a quarter – from 24% to 30%. Business travelers are those most likely to judge that they made worthwhile use of their time. The increase in worthwhile use of rail travel time may be linked to improvements in service provision (comfort, delay) but also to improvements in terms of individuals' equipment (laptops, smart phones). Yet, in urban public transport, travel time remains an untapped potential. In Lyon (France), a third of urban transport users judge that their travel time is wasted or too short to do anything (Casals, 2012).

Feelings experienced during travel time also vary. Based on the American Time Use Survey's well-

being module, Morris and Guerra (2015) observed five distinct emotions experienced during travel. They find that bus and train users experience the most negative emotions, though a small part of this can be attributed to the fact that public transport is disproportionately used for the unloved work trip. In France, the above-mentioned local survey (Casals, 2012) provides different results. The majority of urban public transport users experience positive feelings during their travel time. Seventy percent are always or often in a good mood; 60% feel a sense of freedom; 54% feel comfortable and at ease.

6.3 Hypotheses and theoretical model

Based on the literature review, seven hypotheses are formulated on the relationships between environmental concern, motives for car use, perceptions and feelings towards public transport, and mode choice habits. Hypotheses are also formulated on how socio-demographic variables influence the latent variables and how constraints, such as accessibility to public transport, shape mode choice habits.

A first set of hypotheses is built around car use habits. According to these hypotheses, car use habits are stronger when car use generates indirect utility in addition to direct utility. Affective and symbolic motives for car use are a source of indirect utility. Individuals with high environmental concern have lower affective and symbolic motives for car use, car use being mainly motivated by instrumental motives. High environmental concern results in car use to access activities and not for pleasure. Hence, it is expected that affective and symbolic motives for car use will mediate the relationship between environmental concern and car use habits.

- **H1**: **High environmental concern**...

  H1 a) ... reduces affective and symbolic motives for car use;

  H1 b) ... reduces car use habits;
- **H2**: A high influence of **affective and symbolic motives** on car use promotes car use habits.
- **H3**: Affective and symbolic motives for car use play a **partial mediating role** between environmental concern and car use habits.

Similarly, a second set of hypotheses is built around public transport habits. According to these hypotheses, public transport habits are stronger in people who perceive it as easy, productive and pleasant to travel by public transport modes. People with high environmental concern are aware of the advantage of choosing public transport over the car, and they therefore have an additional motive for adopting public transport modes and will do so less reluctantly. They will have more positive

attitudes towards public transport modes, perceive it as easier to travel by public transport, be in a more positive state of mind, and thus perceive more favorably their indirect utility. In short, people with high environmental concern, will experience more positive feelings and judge their time as more productive while traveling by public transport modes. Hence, perceptions and feelings towards public transport modes are expected to mediate the relationship between environmental concern and public transport habits.

- **H4**: **High environmental concern**...

  H4 a) ... has a positive influence on perceptions and feelings towards public transport;

  H4 b) ... promotes public transport habits.
- **H5**: Positive **perceptions and feelings towards public transport** make it more likely to be used.
- **H6**: Positive perceptions and feelings towards public transport play a **partial mediating role** between environmental concern and public transport habits.

The last hypothesis makes the link between public transport and car use habits and reflects the trade-off dimension of mode choice habits.

- **H7**: **Public transport and car use habits** are negatively correlated.

With regard to the socio-demographic variables, the following hypotheses are made. First, older people and men are expected to have lower environmental concern than younger people and women. Second, in terms of mode choice habits, hypotheses are based on the literature review by De Witte et al. (2013). People living in urban areas are expected to have a greater propensity to use public transport since it is more accessible. The size of the household is expected to be positively linked to car use habits. Mixed results are found concerning age and gender and no conclusion is drawn in the literature review.

A theoretical model to test the above-mentioned hypotheses is presented in Figure 6.1. Consistent with the hypotheses, it incorporates a partial mediation relationship, so that there is a direct influence of environmental concern on mode choice habits, but also an indirect influence through the mediation of the two latent variables that reflect indirect utility.

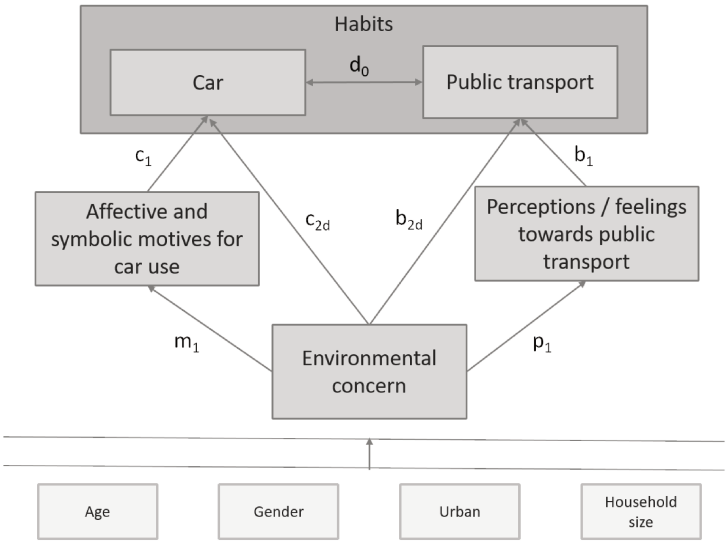


FIGURE 6.1 – Theoretical model

6.4 Data and method

6.4.1 Survey

The data combines a stated preferences (SP) survey and a revealed preferences survey, specifically, a travel survey carried out in the Rhône-Alpes region. The respondents in the regional travel survey were sampled on the basis of stratified random sampling. In order to be sure that all areas in the Rhône-Alpes region were adequately represented, the sample is geographically stratified into 77 selected sample sectors. The regional travel survey was carried out in three phases by phone between 2012 and 2015. The respondents were then contacted to answer an online SP survey, which includes the measure of socio-demographic and socio-psychological variables. The sample includes 1,061 people.

6.4.2 Measure of the socio-psychological variables

A three-stage process was used to choose and measure the socio-psychological variables: qualitative surveys (individual interviews and focus groups), a literature review and a quantitative survey. Based on the qualitative analysis and the literature review, a list of the relevant socio-psychological variables was compiled and included in a dedicated questionnaire answered by 277 persons. On that basis, the first analyses (exploratory factor analysis, measurement model and structural model) were conducted to select the most relevant variables and insert them into the final SP questionnaire.

Below, we record only the items retained in the final survey.[4] All these items are measured on the basis of 5-point Likert scales, which range from "completely agree" to "completely disagree", except feelings experienced in public transport, which are measured on a 4-point scale ranging from "never" to "always".

6.4.2.1 Environmental concern

The three dimensions of environmental concern (cognitive, affective and conative) are measured using items taken from the literature (Table 6.1). The cognitive dimension is measured with three items from the NEP, which are related to the concepts of "ecological crisis" and "balance of nature" (Schleyer-Lindenmann et al., 2014). The affective dimension of environmental concern is measured using four items taken from Steg (2003). The respondents had to say to what extent they felt concerned by certain problems associated with using a car, such as air pollution, quality of life, traffic, odors, use of scant resources and space usage. The conative dimension of environmental concern was measured on the basis of two items involving the legitimacy ascribed to transport policies. Steg (2003) uses them to identify those people most likely to use public transport and how they may be persuaded to use it. We added three items selected from Atasoy et al. (2011) and Hurtubia et al. (2014).

TABLE 6.1 – Items measuring environmental concern

Item	Definition	Dimension	Source
NEP1	The balance of nature is very delicate and easily upset.	Cognitive	Dunlap et al. (2000)
NEP2	The so-called "ecological crisis" facing humankind has been greatly exaggerated.		
NEP3	If things continue on their present course, we will soon experience a major ecological catastrophe.		
Pbcar1	Is concerned by air pollution caused by cars.	Affective	Steg (2003)
Pbcar2	Is concerned by space occupation caused by cars.		
Pbcar3	Is concerned by traffic unsafety caused by cars.		
Pbcar4	Is concerned by reduced quality of life caused by cars.		
POL1	We need to build more parking lots downtown.	Conative	Hurtubia et al. (2014)
POL2	We should increase the price of gasoline to reduce congestion and air pollution.		Atasoy et al. (2011)
POL3	We need more public transport, even if it means higher taxes.		
POL4	Everyone should be free to use their car whenever they want.		Steg (2003)
POL5	The government has the right to reduce car use to safeguard environmental qualities and the urban quality of life.		

4. A list of all items used in the first questionnaire is available on request from the authors.

6.4.2.2 Motives for car use

The measurement for the instrumental, affective and symbolic motives for car use (Table 6.2) comes principally from the work of Steg (2005) and Lois and López-Sáez (2009), themselves influenced notably by Steg et al. (2001) and Jensen (1999). One further item, "When I drive I am afraid of having an accident" was added to measure the instrumental motives for car use. Indeed, the preliminary qualitative analysis emphasized that the risk of an accident comes into the motives for mode choice of some people.

TABLE 6.2 – Items measuring motives for car use

Item	Definition	Dimension	Source
AFF1	I like driving fast.	Symbolic and affective motives for car use	Steg (2005)
AFF2	I am a bit in love with my car.		
AFF3	You can know a person by looking at his or her car.		
AFF4	The better your car is, the more successful you are in life.		Lois and López-Sáez (2009)
INSTR1	I am safe in my car.	Instrumental motives for car use	Steg (2005)
INSTR2	It does not matter to me which type of car I drive.		
INSTR3	I am afraid of having an accident when I drive.		

6.4.2.3 Perceptions and feelings towards public transport

Perceptions and feelings towards public transport are measured by three dimensions: PBC, perceived time in public transport, and feelings experienced during journeys made by public transport (Table 6.3). The variables used to represent PBC were taken from Atasoy et al. (2011) and Morikawa et al. (1996). The more positively a person responds to these questions, the less control they feel they have in using public transport. The measurement for perception of time and feelings is based on a local study carried out on public transport in Lyon (Casals, 2012).

6.4.2.4 Mode choice habits

Mode choice is a decision repeated over time and strongly dependent on habits developed in the past (e.g. Verplanken et al., 1994; Cherchi et al., 2017). Measuring habits is necessarily difficult since behavior generated by habits results from limited awareness. Several scales have been suggested and used in the literature (Eriksson et al., 2008), specifically to appreciate the unintentional nature of behaviors generated by habits. We choose to measure habits by intensity of use, which is therefore restrictive compared with what other authors propose but better adapted to our objective, which is to explain actual behavior.

TABLE 6.3 – Items measuring perceptions and feelings towards public transport

Item	Definition	Dimension	Source
PBC1	I'm not comfortable when I travel with people I do not know well.	Perceived behavioral control	Atasoy et al. (2011)
PBC2	Its hard to take public transport when I travel with my children.		Morikawa et al. (1996)
PBC3	Its hard to take public transport when I travel with bags or luggage.		
Time1	I like seeing people and having other people around me.	Perceived time in public transport	Casals (2012)
Time2	It's time I put up with and I just wait for it to pass.		
Time3	I use the time to rest and relax.		
Time4	I use the time to do things I wouldn't necessarily do elsewhere.		
Time5	I just want to be on my own and undisturbed.		
Time6	Given my commutes, the time is too short: I don't have time to do anything.		
Time7	It's wasted time.		
Feel1	I feel a sense of freedom.	Feelings in public transport	Casals (2012)
Feel2	It puts me in a good mood.		
Feel3	I feel comfortable and at ease.		
Feel4	I feel I could meet people and get into conversation with them.		
Feel5	I feel I'm doing something, I feel useful.		
Feel6	I find the people, noise and smells disagreeable.		
Feel7	I feel stressed.		
Feel8	I feel harassed.		

Intensity of use is measured discretely over a year. The variables introduced to measure car use habits are the number of cars available to the household and the frequency with which a car is used. For public transport, the frequency with which urban public transport and intercity trains and coaches are used is evaluated. This measure has the advantage of translating both a level of use and an implicit trade-off between travel modes. With respect to one-off choice, it has the advantage of being less dependent on specific travel contexts and of encompassing the repetitive dimension of mode choice.

6.4.3 Method

The theoretical model defines the relationships between latent variables and between latent and observable variables. SEM is an appropriate tool to estimate such a model since it can simulateneously handle numerous exogenous and endogenous variables, as well as latent variables specified as linear combinations of the observed variables (Hoyle, 2012). See Golob (2003) for a literature review on the use of structural equation modeling in transport and, for example, Ory and Mokhtarian (2009), Ye and Titheridge (2016) and Zailani et al. (2016) for more recent SEMs on the impact of attitudes on travel behavior.

In the presence of latent variables, an SEM is made up of a measurement model and a structural model (Hoyle, 2012; Esposito Vinzi and Trinchera, 2013). The measurement model relates each latent variable to its measuring items according to a reflexive linear relationship. The structural model describes the relationships between the latent variables and between the latent variables and the contextual variables. In addition, measurement errors of items measuring the same latent variables and with a reverse coding are correlated.

6.4.3.1 Exploratory factor analysis

As a first step, relevant items were selected using an exploratory factor analysis before inclusion in the SEM. The internal consistency of the latent variables was also tested by calculating Cronbach's alpha coefficient (α), for which a value greater than 0.7 indicates high internal consistency (Esposito Vinzi and Trinchera, 2013; West et al., 2012).

6.4.3.2 Mediation analysis

The indirect effect, which constitutes mediation in the strict sense, is calculated as the product of the effect of the independent variable on the mediator and of the effect of the mediator on the output (Hayes, 2013a). The method described in Baron and Kenny (1986) is used to test the existence of a mediation effect and involves the estimation of three sub-models. The first step is to check whether the independent variable has an influence on the outcome (no mediation). The second step is to check that the independent variable affects the mediator and that the mediator influences the outcome (perfect mediation). The third step is to include all paths (partial mediation). If the indirect effect is significant and the direct effect is not, then the mediation is perfect. If both effects are significant, then the mediation is partial.

Since the indirect effects are obtained through the product of two random variables, their standard errors and related test statistics are calculated using the Delta method (Sobel, 1982). To check the robustness of the results, diagonally weighted least square estimations with bootstrapped standard errors (1,000 sample draws) were also performed. Results are robust and available on request from the authors.

6.4.3.3 Definition of the estimated models

First, public transport and car use habits are considered separately to gain insight into the partial mediation relationships of the theoretical model. Then, public transport and car use habits are

considered together. This phased approach helps to progressively investigate the relationships stated in the theoretical model. The use of different models makes it possible to evaluate the effects of the latent variables on mode choice habits, both separately and together, and to check the robustness of the results.

The first model (M1, see Figure 6.2) is built to explain car use habits (H1 to H3). The mediating variable, affective and symbolic motives for car use, allows for an analysis of 1) the direct effects of environmental concern on car use habits (c2d); 2) the indirect effect (c2i=m1$\times$c1) of environmental concern on car use habits through the mediation of affective and symbolic motives for car use, and 3) the total effect (c2t) of environmental concern on car use habits being the sum of the direct and indirect effects. Three sub-models are defined to investigate the partial mediation relationship (M1a: $m1 = c1 = 0$; M1b: $c2d = 0$; M1c: no constraint).

Similarly, the second model (M2, see Figure 6.3) is built to explain public transport habits (H4 to H6). Three sub-models are defined to investigate the partial mediation relationship (M2a: $p1 = b1 = 0$; M2b: $b2d = 0$; M2c: no constraint).

The third model (M3, see Figure 6.1) is the theoretical model, which takes account of both modes and their correlation (H7) together. Nine sub-models are estimated, corresponding to the 3×3 possible mediations for public transport and car use habits (no mediation, partial mediation or perfect mediation).

Table 6.4 provides a summary of all the estimated models according to the nature of the mediation [5].

TABLE 6.4 – Estimated models

Car use habits	**Public transport habits**	**Model**
no mediation	-	M1a
perfect mediation	-	M1b
partial mediation	-	M1c
-	no mediation	M2a
-	perfect mediation	M2b
-	partial mediation	M2c
no mediation	no mediation	M3a
no mediation	perfect mediation	M3b
no mediation	partial mediation	M3c
perfect mediation	no mediation	M3d
perfect mediation	perfect mediation	M3e
perfect mediation	partial mediation	M3f
partial mediation	no mediation	M3g
partial mediation	perfect mediation	M3h
partial mediation	partial mediation	M3i

5. Figures with the 15 estimated models are available on request from the authors.

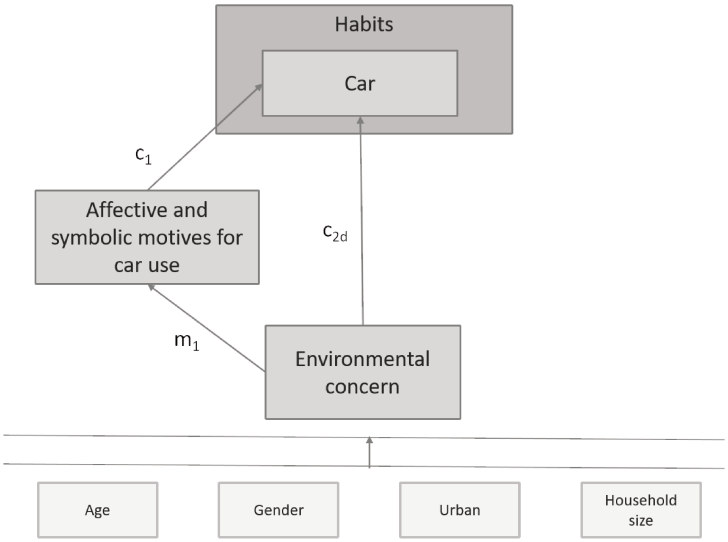


FIGURE 6.2 – Model M1

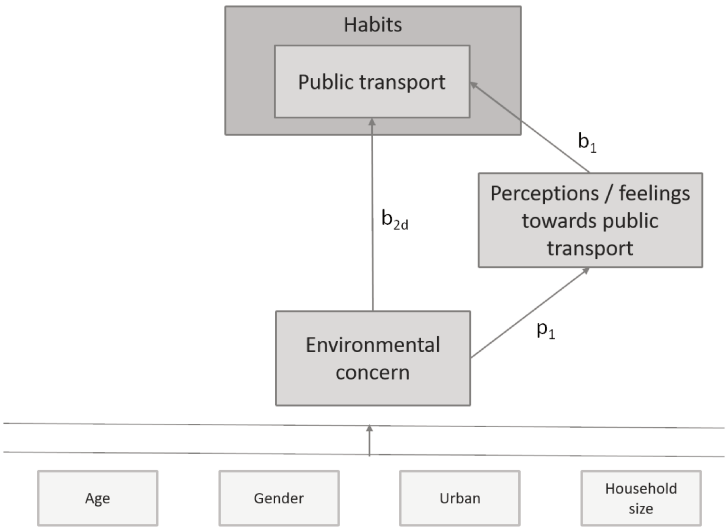


FIGURE 6.3 – Model M2

6.4.3.4 Estimation method

The observed items are ordinal variables with four or five levels. The maximum likelihood estimation is therefore not suitable because it leads to a biased estimate of the parameters and to inaccurate standard errors and test statistics (Rhemtulla et al., 2012). A commonly used alternative is to estimate a limited-information method. Each observed item is supposed to be the manifestation of a latent variable with a Gaussian distribution. A threshold model divides this continuous variable into categories to obtain the ordinal item as well as the matrix of polychoric correlation. The measurement model is then estimated on the basis of the matrix of polychoric correlation. Among the limited-information methods, the weighted least squares method is the most effective but it requires very large samples. We therefore use the diagonally Weighted Least Squares method with a Mean and Variance correction (WLSMV). Indeed, owing to the loss of efficiency, the standard errors need to be corrected (Muthén, 1993) and the test statistic adjusted by the average and variance (Muthén, 1993; Maydeu-Olivares, 2001).

6.4.3.5 Goodness-of-fit indicators

There are dozens of indicators for assessing the Goodness Of Fit (GOF) of a theoretical model to the data. Two GOF indicators commonly used in the literature (West et al., 2012) are: the Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA). The literature defines thresholds in order to assess the GOF of a model to the available data. Although thresholds vary according to the authors, a CFI of more than 0.90 indicates a good fit (West et al., 2012; Hair et al., 2006). The RMSEA should be less than 0.06 or 0.08 (West et al., 2012; Hair et al., 2006). As recommended by West et al. (2012), thresholds are reported as guidelines but not to select a specific model.

6.5 Results

The exploratory factor analysis was done in SAS. The estimation of the SEM uses the R package Lavaan (Rosseel, 2012).

6.5.1 Description of the Sample

The sample consists of 1,061 people, 542 of whom are men (51%, see Table 6.5). The respondents are between 18 and 91 years of age, with an average age of 52. The household consists of an average of 2.5

people with a maximum of 7 in the household. Forty-eight percent of the respondents live in an urban area. On average, the households have access to 1.7 cars. In terms of travel modes used (see Table 6.6), the car is the predominant mode, since 64% of respondents use it every day or nearly every day, whereas that is only the case for 8% of the respondents for bicycles, 7% for urban public transport (bus, tram and underground), 5% for the train or coach and 2% for car-sharing.

TABLE 6.5 – Descriptive statistics

Definition	**Label**	**Mean**	**S.d.**	**Min**	**Max**
Age (in years)	Age	52.27	13.56	18.00	91.00
Gender (1 if man, 0 if woman)	Gender	0.51	0.50	0.00	1.00
Live an urban area (1 if yes, 0 otherwise)	Urban	0.84	0.37	0.00	1.00
Size of the household	Househ.	2.53	1.23	1.00	7.00
Number of cars in the household	Nb_cars	1.74	0.70	0.00	5.00

TABLE 6.6 – Frequency of use of transport modes (in %)

	Car	**Bicycle**	**Urban public transport**	**Train/coach**	**Car-sharing**
Never	0.4	58.8	59.9	51.8	61.6
Occasionally	5.2	20.4	17.7	32.9	22.2
At least twice a month	6.7	5.8	8.0	6.7	8.4
At least twice a week	23.7	7.4	7.9	3.9	5.8
Every day or nearly every day	64.1	7.6	6.5	4.7	2.1

Most of the items measuring the socio-psychological variables are asymmetrically distributed with marked kurtosis (see Figure 6.4). The use of the WLSMV estimation method is thus fully justified.

6.5.2 Exploratory Factor Analysis

An exploratory factor analysis with promax rotation identifies the factors that best distinguish the sample and the observed variables with the greatest loadings. Previously, some coding was reversed so that each observed variable expressed the same idea in the same way.

Four factors have a value higher than one but three are enough to explain 80% of the variance in the data. As the fourth factor explains less than 10%, it was omitted. Also, the interpretability of the first three factors is good and consistent with the theoretical models (see Table 6.7). Only the items with a loading greater than 0.35 were retained in the SEM. Some items are therefore not included in any factor (Instr2, Instr3, Time5 and Time6).

The first factor ($\alpha = 0.83$) refers to environmental concern and involves the three dimensions

FIGURE 6.4 – Distribution of answers on items

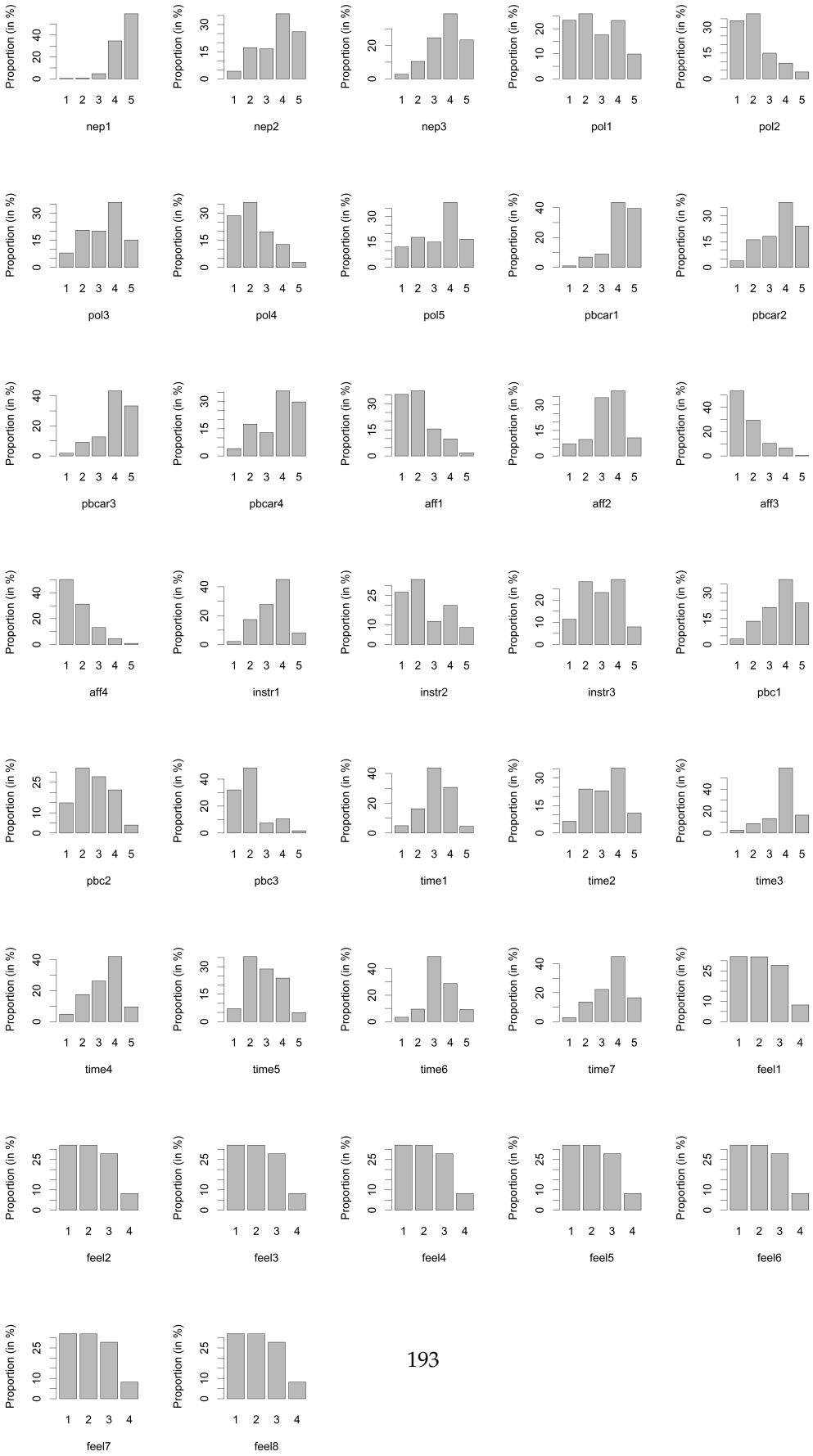


TABLE 6.7 – Exploratory factor analysis loadings

	Factor 1 Environmental concern	**Factor 2** Affective and symbolic motives for car use	**Factor 3** Perceptions and feelings towards public transport
NEP1	48 *	-21	16
NEP2$^{(i)}$	63 *	-15	13
NEP3	59 *	-15	16
POL1$^{(i)}$	41 *	-12	26
POL2	49 *	-4	27
POL3	47 *	-10	26
POL4$^{(i)}$	52 *	-10	24
POL5	61 *	-7	35
Pbcar1	65 *	-17	21
Pbcar2	61 *	-20	21
Pbcar3	39 *	-19	7
Pbcar4	59 *	-18	20
AFF1	-16	43 *	-10
AFF2	-30	47 *	-16
AFF3	-7	66 *	-10
AFF4	-14	70 *	-9
INSTR1	-33	42 *	-11
INSTR2$^{(i+a)}$	-10	33	-7
INSTR3$^{(a)}$	21	-25	-5
PBC1$^{(i)}$	30	-26	51 *
PBC2$^{(i)}$	29	-11	39 *
PBC3$^{(i)}$	28	-9	36 *
Time1	22	-8	57 *
Time2$^{(i)}$	25	-12	59 *
Time3	21	-8	52 *
Time4	20	-6	40 *
Time5$^{(i+a)}$	3	-12	29
Time6$^{(i+a)}$	16	-9	17
Time7$^{(i)}$	19	-16	50 *
Feel1	23	-9	67 *
Feel2	16	-7	66 *
Feel3	20	-8	68 *
Feel4	20	-13	54 *
Feel5	22	-8	51 *
Feel6$^{(i)}$	14	-5	55 *
Feel7$^{(i)}$	14	4	53 *
Feel8$^{(i)}$	19	4	53 *
Cronbach's α	0.83	0.67	0.86

Notes: * loading greater than 0.35. (i) variable for which the coding has been reversed; (a) variable not included in any factor because of too small a loading.

previously established, the New Ecological Paradigm (NEP), the degree of agreement with policies promoting public transport, and awareness of environmental problems.

The second factor refers to the affective and symbolic motives for car use and includes five items. Unlike Steg (2005), the item "I feel safe in my car" is not associated with instrumental motives but shows evidence of affective motives ascribed to the car. The relatively low Cronbach's alpha ($\alpha = 0.66 < 0.7$) suggests caution regarding the interpretation of parameters associated with motives for car use.

The third factor (α=0.86) refers to the perceptions and feelings towards public transport and is made up of three dimensions: PBC, how time spent on public transport is viewed, and people's feelings during journeys made on public transport.

6.5.3 Structural Equation Modeling

Consistent with our methodology, mediation is first evaluated in two separate models for each mode and then evaluated in a joint model. Fifteen models are tested: three sub-models for M1, three sub-models for M2 and nine sub-model for M3. For the sake of simplicity, table 6.10 shows only the results relevant for the discussion on the hypotheses. Tables 6.8 and 6.9 show the relevant results for the mediation analysis in the separate Models M1 and M2. Appendix G shows the detailed estimation results for Model M3.

Table 6.10 also shows robust GOF indices and degrees of freedom for the sub-models whose results are discussed in sections 6.5.3.1 and 6.5.3.2. All the models show a reasonable goodness of fit. Apart from M3i, the CFI of all sub-models is above the threshold value of 0.9. The RMSEA is below the threshold value 0.06 for all sub-models.

The part of the variance of public transport habits explained by sub-models M2c, M3c and M3i is 15%. It is higher for car use habits with 33% to 37% of variance explained by Models M1c, M3c and M3i.

6.5.3.1 Impact of latent variables on mode choice habits

We first look at the analysis of car use habits in Model M1. According to sub-models M1a, M1b and M1c, affective and symbolic motives for car use do not mediate the relationship between environmental concern and car use habits.

In sub-models M1a (no mediation) and M1c (partial mediation), a high environmental concern significantly lowers motives for car use (respectively, m1=-0.483, p.value<0.01 and m1=-0.443, p.value<0.01) and car use habits (respectively c2d=-0.678, p.value<0.01 and c2d=-0.585, p.value<0.01). According to sub-model M1c, affective and symbolic motives for car use do not mediate the

relationship between environmental concern and car use habits. Indeed, the parameter associated with the indirect effect of environmental concern on car use habits is significantly different from zero in sub-model M1b (perfect mediation) (c2i=-0.227 , p.value <0.01) but becomes non-significant in sub-model M1c (c2i=-0.051, p.value > 0.1).

Only the direct effect of environmental concern on car use habits (H1b) seems to be supported by the data. Hypothesis H3, which states that affective and symbolic motives for car use play a mediating role between environmental concern and car use habits, is rejected with our data. However, environmental concern seems to influence affective and symbolic motives for car use (H1a).

An explanation for the absence of an indirect effect is that the relationship between affective and symbolic motives for car use and car use habits (H2) is not robust, since the associated parameter is significant only at the 10% level in sub-model M1c (c1=0.114, p.value<0.1). The direct effect of environmental concern on car use habits seems to overcome both the indirect effect and the affective and symbolic motives effect on car use habits.

TABLE 6.8 – Test of mediation: Model M1

Dependent variable	**Explanatory variable**	**Label**	**M1a**	**M1b**	**M1c**
			no mediation	perfect mediation	partial mediation
Motives (car)	Env. concern	m1		-0.483 (0.065) ***	-0.443 (0.062) ***
Car use habits	Motives	c1		0.471 (0.08) ***	0.114 (0.069) *
	Env. concern (dir.)	c2d	-0.678 (0.094) ***		-0.585 (0.094) ***
	Env. concern (ind.)	c2i		-0.227 (0.041) ***	-0.051 (0.031)
	Env. concern (tot.)	c2t			-0.636 (0.09) ***

Notes: (***) significant at the 1% threshold; (**) significant at the 5% threshold; (*) significant at the 10% threshold. Abbreviation: Env. concern: environmental concern; Motives (car): affective and symbolic motives for car use; dir.: direct effect; ind.: indirect effect; tot.: total effect.

The three sub-models related to public transport habits (M2a M2b, M2c) indicate that perceptions and feelings towards public transport play a partial mediation role.

In M2c, environmental concern has a positive influence on perceptions and feelings towards public transport (p1=0.645, p.value<0.01), which, in turn, increase public transport habits (b1=0.241, p.value<0.01).

Environmental concern promotes public transport habits (b2t=0.546, p.value<0.01), both directly and indirectly. Indeed, the direct and indirect effects of environmental concern on public transport habits are significantly different from zero (b2d = 0.39, p.value<0.01; b2i=0.156, p.value<0.01), which supports the hypothesis of a partial mediation.

All hypotheses related to Model M2 are thus supported by the data. A high environmental concern has a positive influence on perceptions and feelings towards public transport (H4a) and promotes public transport habits (H4b). Positive perceptions and feelings towards public transport make it more likely to be used (H5), and perceptions and feelings towards public transport play a partial mediating role between environmental concern and public transport habits (H6).

TABLE 6.9 – Test of mediation: Model M2

Dependent variable	**Explanatory variable**	**Label**	**M2a**	**M2b**	**M2c**
			no mediation	perfect mediation	partial mediation
Percep. (PT)	Env. concern	p1		0.662 (0.065) ***	0.645 (0.064) ***
	Percep.	b1		0.467 (0.078) ***	0.241 (0.083) ***
PT habits	Env. concern (dir.)	b2d	0.778 (0.127) ***		0.39 (0.109) ***
	Env. concern (ind.)	b2i		0.31 (0.055) ***	0.156 (0.054) ***
	Env. concern (tot.)	b2t			0.546 (0.095) ***

Notes: (***) significant at the 1% threshold; (**) significant at the 5% threshold; (*) significant at the 10% threshold. Abbreviation: Env. concern: environmental concern; Percep. (PT): perceptions and feelings towards public transport; PT habits: public transport habits; dir.: direct effect; ind.: indirect effect; tot.: total effect.

In Model M3, both mode choice habits are considered together. The model tests whether the results which hold with separate habits are still relevant jointly. The nine sub-models provide consistent results, which support previous conclusions.

Table 6.10 shows the estimation results of the theoretical model (M3i) with partial mediation for both mode choice habits. It also displays the results for the Model M3c, which is a model with partial mediation for public transport habits and no mediation for car use habits, with better GOF.

M3i confirms that affective and symbolic motives for car use play no mediating role (c2i=-0.034, p-value>0.1). The relationship between affective and symbolic motives for car use and car use habits is no longer valid once the influence of environmental concern is taken into account (c1=0.091, p.value>0.1). As in Model M1, including environmental concern, both as a determinant of affective and symbolic motives and as a determinant of mode choice habits, reveals its dominant role over affective and symbolic motives for car use. Those affective and symbolic motives for car use are significantly determined by environmental concern in a negative way (m1=-0.376, p-value<0.01).

Finally, the more parsimonious Model M3c (no mediation / partial mediation) provides a better fit and leads to the following results. Environmental concern influences both mode choice habits. More precisely, people with high environmental concern are more prone to use public transport modes (b2t= 0.59, p-value<0.01) and more reluctant to use the car (c2d=-0.563, p-value<0.01). The coefficients

regarding the effects of environmental concern on car use habits are similar in magnitude to those obtained for public transport habits. However, for public transport habits, the total effect is composed of both an indirect effect (b2i=0.152, p-value<0.01) and a direct effect (b2d=0.438, p-value<0.01). Since both parameters are significant, Model M3c confirms that environmental concern is partially mediated by perceptions and feelings towards public transport. Moreover, people with a higher environmental concern assign a higher indirect utility to public transport (p1=0.651, p-value<0.01).

Results are thus robust to the shift from separate to joint models. Taking into account the trade-off between public transport and car use habits does not alter the results even if, as expected, this trade-off is significantly different from zero (d0=-0.233, p-value<0.01).

TABLE 6.10 – Estimation results

Dependent variable	**Explanatory variable**	Label	M1c partial mediation	M2c partial mediation	M3c no mediation / partial mediation	M3i partial/partial mediation
Env. concern	Age	e3	0.02 (0.006) ***	0.025 (0.008) ***	0.024 (0.008) ***	0.024 (0.008) ***
	Age^2	e4	-2$\times 10^{-4}$ (0) ***	-2$\times 10^{-4}$ (0) ***	-2$\times 10^{-4}$ (0) ***	-2$\times 10^{-4}$ (0) ***
	Gender	e5	-0.051 (0.027) *	-0.057 (0.033) *	-0.06 (0.033) *	-0.061 (0.033) *
Motives (car)	Env. concern	m1	-0.443 (0.062) ***			-0.376 (0.051) ***
	Age	m3	-0.015 (0.006) **			-0.015 (0.006) **
	Age^2	m4	9$\times 10^{-5}$ (0)			9$\times 10^{-5}$ (0)
	Gender	m5	0.105 (0.029) ***			0.105 (0.03) ***
Percep. (PT)	Env. concern	p1		0.645 (0.064) ***	0.651 (0.064) ***	0.698 (0.065) ***
	Age	p3		-0.007 (0.008)	-0.006 (0.008)	-0.007 (0.008)
	Age^2	p4		1$\times 10^{-4}$ (0) *	1$\times 10^{-4}$ (0)	1$\times 10^{-4}$ (0) *
	Gender	p5		0.072 (0.038) *	0.077 (0.038) **	0.08 (0.04) **
PT habits	Percep. (PT)	b1		0.241 (0.083) ***	0.233 (0.086) ***	0.221 (0.085) ***
	Env. concern (dir.)	b2d		0.39 (0.109) ***	0.438 (0.113) ***	0.445 (0.117) ***
	Env. concern (ind.)	b2i		0.156 (0.054) ***	0.152 (0.057) ***	0.154 (0.06) ***
	Env. concern (tot.)	b2t		0.546 (0.095) ***	0.59 (0.096) ***	0.599 (0.097) ***
	Age	b3		-0.081 (0.015) ***	-0.082 (0.015) ***	-0.082 (0.015) ***
	Age^2	b4		0.001 (0) ***	0.001 (0) ***	0.001 (0) ***
	Gender	b5		0.053 (0.066)	0.073 (0.069)	0.074 (0.07)
	Urban	b6		0.346 (0.116) ***	0.385 (0.123) ***	0.385 (0.123) ***
Car use habits	Motives (car)	c1	0.114 (0.069) *			0.091 (0.07)
	Env. concern (dir.)	c2d	-0.585 (0.094) ***		-0.563 (0.073) ***	-0.521 (0.077) ***
	Env. concern (indir.)	c2i	-0.051 (0.031)			-0.034 (0.027)
	Env. concern (total)	c2t	-0.636 (0.09) ***			-0.556 (0.072) ***
	Age	c3	0.053 (0.011) ***		0.057 (0.011) ***	0.057 (0.011) ***
	Age^2	c4	-0.001 (0) ***		-0.001 (0) ***	-0.001 (0) ***
	Gender	c5	0.047 (0.052)		0.057 (0.055)	0.048 (0.055)
	Urban	c6	-0.236 (0.075) ***		-0.254 (0.081) ***	-0.249 (0.079) ***
	Househ.	c7	0.168 (0.025) ***		0.155 (0.025) ***	0.16 (0.025) ***
Covariance (PT habits / car use habits)						
		d0			-0.233 (0.03) ***	-0.229 (0.03) ***
Goodness-of-fit indicators						
		χ^2	1042	1551	1734	2370
		Df	229	500	592	777
		Cfi	0.917	0.907	0.906	0.886
		Rmsea	0.058	0.045	0.043	0.044

Notes: (***) significant at the 1% threshold; (**) significant at the 5% threshold; (*) significant at the 10% threshold. Abbreviation: Env. concern: environmental concern; Motives (car): affective and symbolic motives for car use; Percep. (PT): perceptions and feelings towards public transport; PT habits: public transport habits; dir.: direct effect; ind.: indirect effect; tot.: total effect.

Table 6.11 summarizes whether the hypotheses are consistent with our data.

TABLE 6.11 – Summary of hypotheses and results

Hypothesis	Related parameter	Model(s)	Supported by data?
H1: High environmental concern			
H1 a) reduces affective and symbolic motives for car use habits;	m1	M1/M3	Yes
H1 b) reduces car use habits.	c2d/c2t	M1/M3	Yes
H2: A high influence of affective and symbolic motives on car use promotes car use habits	c1	M1/M3	No at the 5% level
H3: Affective and symbolic motives for car use play a **partial mediating role** between environmental concern and car use habits.	c2i	M1	No
H4: High environmental concern			
H4 a) has a positive influence on perceptions of public transport;	p1	M2/M3	Yes
H4 b) promotes public transport habits;	b2d/b2t	M2/M3	Yes
H5: Positive perceptions of public transport make it more likely to be used.	b1	M2/M3	Yes
H6: Perceptions of public transport play a **partial mediating role** between environmental concern and public transport habits.	b2i	M2/M3	Yes
H7: Public transport and car use habits are negatively correlated.	d0	M3	Yes

6.5.3.2 Socio-demographic variables as determinants of latent variables and mode choice habits

Surprisingly, environmental concern increases with age (e3=0.024, p.value<0.01), with a concave relationship (e4=-2$\times 10^{-4}$, p.value<0.01). A similar negative relationship between age and environmental concern was found by Liu et al. (2014) in American national public surveys conducted in 2007 and 2013. They explain this effect with possible generational changes and declining trends of youth attitudes to environmental issues and rely on three theoretical reasons: the rising importance of extrinsic values (such as money); a lack of connection to nature due to more time spent indoors, and a technological worldview that suggests that technology can and will solve all problems, including environmental problems. Women are found to be more concerned by environmental issues (e5=-0.06, p.value<0.01), which is consistent with previous research.

Age reduces public transport habits (b3=-0.082, p.value<0.01) to the advantage of the car (c3=0.057, p.value<0.01). Conversely, those effects are reduced as age increases, since age^2 has a positive coefficient for public transport (b4=0.001, p.value<0.01) and a positive value for the car (c4=-0.001, p.value<0.01). Contextual constraints such as access to public transport also explain behavior, since living in an urban area significantly increases public transport habits (b6=0.385, p.value<0.01) and reduces car use habits

(c6=-0.254, p.value<0.01). The household size has an inverse effect: the larger the household, the more mode choice habits involve car use (c7=0.155, p.value<0.01). With our data, gender is not a determinant of mode choice habits (b4=0.073, p.value>0.1; c4=0.057, p.value>0.1).

6.6 Conclusion

The SEM analysis shows that the relationship between environmental concern and mode choice habits takes different forms: a partial mediation form for public transport habits and a non-mediated form for car habits. By comparison with previous work, the combined effect of public transport perceptions and environmental concern on public transport habits and the existence of a partial mediation relationship enrich the understanding of how pro-environmental behavior is determined by environmental concern. An intuitive interpretation is that people with higher environmental concern are more motivated to use public transport and thus derive more indirect utility from using it. Conversely, motives for car use and car use habits are explained by environmental concern. However, affective and symbolic motives for car use do not mediate this last relationship. These results seem to be robust in the sense that they appear in the separate analyses for each mode use habit and in the joint analysis. The intuitive negative correlation between public transport and car use habits thus provides a more realistic model for people using multiple modes without altering the understanding of the process underlying each mode choice habit.

Even where "objective" constraints, such as residential location or the household size, have an impact, the socio-psychological determinants of mode choice habits also have to be considered. Their inclusion helps to better understand the behavioral process underlying mode choice habits, and can be seen as an opportunity to identify alternative solutions to traditional economic instruments that only work imperfectly (Croson and Treich, 2014). A challenge is thus to design robust policies based on contextual dependent socio-psychological factors. Although the transferability of our results to other urban and cultural contexts needs to be further investigated, we suggest three interrelated directions in terms of infrastructures, communication strategies and nudges.

First, we show a significant influence exerted on mode choice habits by the quality of travel time and by people's feelings while traveling on public transport. This result, also supported by previous research, provides evidence of the importance of designing infrastructures that make time spent on public transport as pleasant and as efficient as possible. Infrastructures and public policies that improve public transport users' emotional experience may be a key determinant to promote the use of public

transport and induce modal shift.

Second, if mode choice habits are partly dictated by environmental concern, affective and symbolic motives and perceptions and feelings, then those dimensions need to be managed, by, for example, publicizing the environmental impact of the car or by promoting the image of public transport.

Third, a better knowledge of the framing of mode choice habits provides an opportunity to research the way in which users could be nudged towards public transport modes. A nudge is any aspect in the framing of a decision problem that can affect people's decisions without changing economic incentives (Croson and Treich, 2014). Possible nudges could be to organize campaigns in which people commit not to use their cars during a couple of months so that new habits and new perceptions can be formed.

Future research could extend our results and investigate their robustness. First, results may be dependent on how the latent variables are measured. It would be interesting to investigate whether the relationships still hold with different items, in particular for environmental concern since its measurement varies according to the studies, and for affective and symbolic motives for car use since our measure may be improved. Second, an interesting avenue of research would be to test whether the results remain true in different urban and cultural contexts, with different public transport infrastructures. Third, other transport modes, such as cycling or walking, also provide indirect utility, such as health benefits or the pleasure of being outside. An extended version of our model could thus study whether our results still hold with these active transport modes.

Part III

Rank-dependent utility theory and reference models - contributions to the modeling of mode choice

Introduction

Objectives

This part draws on two behavioral economics theories, Rank-Dependent Utility Theory (RDUT; Quiggin, 1982) and reference-dependent theory (Tversky and Kahneman, 1991). RDUT is used to analyze mode choice in a risky context. Reference-dependent theory is used to analyze the economic and behavioral premises of a new family of discrete choice models (DCMs), the reference models developed by Peyhardi et al. (2015). However, this part goes behind the use of those theories and meets two objectives:

1. The first objective is to investigate the intra- and inter-individual heterogeneity of mode choice, when travel time is subject to variability. By "inter-individual heterogeneity" is meant that people are different in terms of attitude toward risk and have different functional forms for the utility function. By "intra-individual heterogeneity" is meant that the behavior may be different even when performed by the same individual when faced with a different transport mode;
2. The second objective is to provide an economic and behavioral interpretation for the reference models, these models never having been applied from an econometric perspective.

Motivation

Motivation for studying the behavioral heterogeneity in mode choice with risky travel time

The motivation for studying the behavioral heterogeneity in mode choice when travel time is subject to variability is three-fold. The first motivation relies on the lessons learned during the qualitative analysis conducted prior to the stated preferences survey. The analysis of the individual and group interviews reveals contrasting perceptions of delays. To be delayed during travel by train is indeed not perceived the same way as being delayed during travel by car. Three dimensions highlight this heterogeneity across modes.

- The first dimension concerns the ascription of responsibility for the delay. The use of the car is the sole responsibility of the driver, even in difficult traffic conditions. Conversely, during travel

by train, it is easier to pin the blame for the delay on the rail operator.

- The second dimension concerns the valuation of the delay. By car, the precise departure and arrival time are not known in advance. The delay is therefore defined in relation to the desired arrival time, for instance an appointment or work start time. A five-minute variability will not be perceived as a delay if the user still arrives on time. By train, the schedule is defined and the delay can be measured, whatever the desired arrival time. A five-minute delay is already perceived as a delay by some people.
- The third dimension concerns the feelings experienced during travel. By car, the driver has a sense of being active and of being able to change her route, even in traffic jams. By train, users may feel trapped and have a sense of powerlessness compounded by the lack of information.

Ultimately, delay probabilities for train trips are certainly overweighted by comparison to delay probabilities for car trips. Indeed, a train trip is more easily classified as being delayed, since the delay is easier to assess by train than by car, and it is easier to pin the blame on others (the rail operator) than on oneself. Moreover, because feelings experienced in the case of a delay are more negative for train trips than for car trips, they are more easily remembered. These findings are questionable and need to be measured. They furthermore depend on the transport system and infrastructure. Inter-individual heterogeneity also needs to be emphasized. Overall, train users seem to be well aware of the delay probabilities, some of them even note when their train is delayed to keep track. They are also aware that cars may also be delayed as a result of a breakdown or heavy traffic, which is one reason why they choose the train. With regards to the delay time, although some users consider the first few minutes of the delay to be unacceptable, most of them have a threshold of 15 minutes. Under 15 minutes, they consider it as natural travel time variability, and therefore do not consider the train as being late. Regular train users also tend to schedule activities in the train, such as reading or working, so that a delay has a smaller impact on them, since they have less of a feeling that the time has been wasted than irregular train users.[6] Conversely, drivers seem to be less lenient when it comes to train trips than car trips and consider that train trips are suited for leisure purposes only.

The second motivation relies on the criticisms made with respect to Expected Utility Theory (EUT) (Von Neumann and Morgenstern, 1944). Experimental evidence, including the Allais paradox (Allais, 1953), has shown that the observed behavior of most decision makers is not consistent with EUT. Experimental evidence also shows that the probability of the same event can be perceived differently,

6. However, regular train users generally use the train for work or study purposes and, therefore, have a stricter timetable than the occasional train users traveling for a leisure purpose. Most of them also get tired in cases of accumulated delays over days.

depending on the characteristics of the decision maker and of the decision problem (Slovic, 1987, cited in Cohen, 2015). Two further criticisms concern the lack of explanatory power of EUT (summarized by Cohen, 2015):

- In EUT, the utility function characterizes two distinct attitudes: the attitude toward risk and the attitude toward health. Therefore, if a decision maker has a diminishing marginal utility, that is, her utility function is concave, then she has to be risk averse;
- Several degrees of loss aversion can be distinguished (Chateauneuf and Cohen, 1994; Chateauneuf et al., 2005). In theory, when it is possible to avoid any risk, a decision maker may have a preference for certainty. However, when it is not completely possible to eliminate risk, the preference for certainty is not sufficient to compare the riskiness of two random variables (here, the travel time). A weakly risk-averse decision maker could choose the riskier situation in the hope of a higher gain. Weak risk aversion, that is, preferring the expected value to any random variable, does not imply strong risk aversion, that is, aversion to mean preserving increases in risk. [7]

RDUT answers these criticisms. Although RDUT has been widely discussed and used in the economic literature (e.g., Schmeidler, 1989; Wakker, 1994; Diecidue and Wakker, 2001), to my knowledge, it has rarely been applied to model mode choice. Inter-individual heterogeneity has been investigated in a transport context (e.g., Hensher and Li, 2012, use a mixed logit model to study unobserved inter-individual heterogeneity) or in the more general lottery games (e.g., Bruhin et al., 2010, distinguish "EUT" and "cumulative prospect theory" types of people with a statistical classification procedure). However, to my knowledge, there are no studies evaluating, for the same decision maker, probability weighting functions whose parameters depend on the context (here, the travel mode). The closest study is Conte et al. (2011) who use lotteries to demonstrate inter-individual heterogeneity, in terms of both functional forms and parameters associated with these functionals, and intra-individual heterogeneity, that is, different behaviors for the same individual and the same choice problem. [8]

The third motivation rests on the importance for transport project appraisal of the economic valuation of travel time and delays (Beaud et al., 2016; Mackie and Worsley, 2013). To be accurate, such valuations have to take into account the specific profile of the users of the transport project and,

7. That means that when comparing two risky situations X and Y, with Y being more risky in the sense that it is obtained by adding a noise (independent of X and with a zero mean) to X, then X is chosen. In the same way, it is possible to define weak and strong risk-seeking attitudes.

8. The choice between lotteries was repeated on five different days over a two-week period.

therefore, heterogeneity has to be effectively addressed.

Motivation for studying reference models

The motivation for studying reference models is two-fold. The first motivation is exploratory. Reference models are a family of models generalizing the multinomial logit model by means of a change in the link function, and more specifically, a change in the cumulative distribution function which is part of the link function. Reference models have been developed as Generalized Linear Models (GLMs) with a statistical purpose. Since they are designed to analyze non-ordered discrete response variables, their use as DCMs to explain consumer choice behavior makes sense. However, before using them to that end, it is necessary 1) to provide them with a behavioral and economic framework; 2) to analyze how the different models of the family allow the analyst to better represent, or represent differently, the choice behavior in comparison to the multinomial logit model, and 3) to be able to derive economic outputs consistent with theory and expectations.

An essential characteristic of the reference models is their dependence on the reference alternative, which leads us to the second motivation. The literature underlines the dependence of choices on the context in which they are made (Tversky and Simonson, 1993). Discrete choice modelers have investigated the influence of context on choices, notably through hybrid choice models (Ben-Akiva et al., 2012). Context is a broad concept including the social environment of the decision maker, but also her habits, the conditions under which she answers the survey, the type of survey, and so on. The context determines, in particular, the references to which an individual refers when making choices. Reference-dependence of choices has been proved by Kahneman and Tversky (1979) and Tversky and Kahneman (1991) by means of experimental evidence, and it arises when utility from an outcome depends on comparisons to a relevant reference point. Preferences which are reference dependent can be seen as an economic manifestation of the general comparative nature of human perception and feelings. However, understanding the role of reference points is not only a psychological issue, it is also an important step in explaining real economic behavior (Starmer, 2000).

Reference dependence is therefore a key feature of behavior, which should be a central issue in discrete choice modeling but which has only recently been integrated in DCMS, in four different ways. Firstly, risky choices, for instance route choice decisions subject to travel time variability, have been modeled by applying (cumulative) prospect theory to the travel time variable, before inclusion in a random utility model (e.g. Hensher et al., 2011; Jou and Chen, 2013).

Secondly, some discrete choice modelers have defined specific variables in their model as the difference between the actual level of that variable and its reference level. For instance, Mabit and Fosgerau (2011) model the choice of alternative-fuel vehicles by including a reference-dependent price variable, the reference price being that of a recently purchased vehicle.

Thirdly, reference dependence has been used to analyze the gap between willingness to pay and willingness to accept in a time/cost trade-off,[9] and to analyze loss aversion.[10] In particular, Stathopoulos and Hess (2012) broaden experimental economics work on reference dependence by presenting a multi-attribute framework applied in the context of commuter choices and allowing for gain-loss asymmetry, non-linearity and testing for several possible reference points. This work reveals important gains in model fit and further insights into behavior compared to standard modeling approaches.

Fourthly, reference dependence has been used to develop new types of DCMs. Zhang et al. (2004) set out a framework where utility is defined by the decision context. This includes 1) features of the choice set (alternative or attribute-specific); 2) the background situation (circumstances surrounding the choice), and 3) individual features that influence decision making, including past choice behavior (social/individual reference). Pan et al. (2016) review these "relative-utility" models and propose a systematic comparison with Random Regret Minimization (RRM) models (Chorus et al., 2008; Chorus, 2010). RRM-models postulate that "the evaluation of an alternative depends on its performance relative to a reference point, and that losses compared to that reference point loom larger than gains of the same magnitude" (Chorus, 2012b, p.57). However, the underlying hypotheses between the reference models studied in Chapter 8 and RRM-models are different. RRM-models are also used to calculate probabilities of choice between alternatives depending on multiple attributes but they are based on the postulate that, since trade-offs have to be made by the decision maker, there will be regret in the sense that there will generally be at least one non-chosen alternative that outperforms a chosen one in terms of one or more attributes. The RRM-model defines a random regret function, which is a weighted sum of the difference between the attribute levels of the considered option and all other options. Therefore, there is no fixed reference point: all alternatives are successively considered as the reference point to calculate the associated regret.

9. Hjorth and Fosgerau (2011) and De Borger and Fosgerau (2008) show that the degree of loss aversion in the time dimension is larger than in the money dimension and depends on age and education. According to Fosgerau and De Borger (2008), loss aversion can be rationally explained since it results in the maximization of experienced utility (the experienced utility is named "expected utility" in Fosgerau and De Borger, 2008).

10. De Borger and Glazer (2017) show the impact of reference-dependent preferences and loss-aversion on the individual attitude toward a Pigovian tax aimed at reducing congestion. In the context of choices among alternative-fuel vehicles, Mabit et al. (2015) show that people have reference-dependent preferences that could be explained by loss aversion for price and that these preferences are related to individual attitudes toward car features.

Ultimately, these examples support the idea that reference dependence is relevant in the context of transport decisions, and that reference models offer one of the first opportunities to analyze reference dependence in a discrete choice framework.

Outline

Part III is divided into two chapters. Chapter 7 applies RDUT to analyze behavioral heterogeneity in mode choice. Chapter 8 introduces reference models and characterizes them statistically and as DCMs. This chapter also derives economic outputs from the reference models and compares them with those of the multinomial logit model on the basis of the data from the stated preferences survey. Finally, this chapter provides an economic and behavioral interpretation of reference models by means of simulated data.

To make the reference models easier to understand, Appendix H sets out the background and defines the key concepts of GLMs. In particular, the concept of "link function" is necessary to understand the reference models and the differences with the multinomial logit model. Appendix H discusses the GLMs applied to categorical response variables and makes a direct connection with the DCMs.

Summary and limitations

Chapter 7 provides an example of how a behaviorally sound model can improve economic analysis. Starting from the intuition that probability weightings are different for train trips and car trips, a nested logit model incorporating RDUT is built. The main findings are that the probabilities of delay associated with train trips are overweighted and the probabilities associated with car trips are underweighted. A latent-class model offers a somewhat different view: if, overall, car users do indeed weight the probabilities for train and car trips differently, the train users seem to be risk neutral and consider the objective probabilities that are presented in the stated preferences survey. The values of time derived from the different models with a risk attitude parameter decrease with travel time because the utility is convex. The order of magnitude of the values of time are consistent with the literature. However, users appear to be heterogeneous according to the transport mode they actually use: train users have a considerably higher value of time for car trips than for train trips. Car users have a similar value of time for both modes. However, this result needs to be carefully considered given that the travel time coefficients are not significant.

Since the car is associated with optimism regarding the delay probabilities and a convex utility function, travelers (or at least car users) are strong risk seekers when they travel by this mode. This means that they are ready to take risks in the hope of having a lower travel time. However, on the basis of the work presented in Chapter 7, it is not possible to conclude on the risk attitude when people travel by train. Indeed, as the probability weighting function is concave and the utility function convex, the curvatures of the functions need to be compared before concluding whether people are weakly risk averse or risk seekers. This estimation will have to be made using the work of Chateauneuf and Cohen (1994) and Cohen (2015).

The originality of the work presented in Chapter 7 relies on the richness of the data with different types of users, different travel modes and variability in probability of delay and time of delay. However, it would be interesting to have access to other datasets to test the robustness of the results.

Another theory of choice under risk that might be considered is Cumulative Prospect Theory (CPT; Tversky and Kahneman, 1992), which integrates RDUT into the original prospect theory by using cumulative probability weights instead of separable probability weights. Moreover, CPT defines a reference point, in comparison to which the outcomes are mapped as gains or losses. Therefore, besides attitudes toward risk and cumulative decision weights, CPT recognizes different perceptions of gains and losses, another key feature deviating from EUT. It results in a utility function which is supposed to be concave in the gains domain and convex in the loss domain. A greater slope ratio in losses than in gains is assumed to capture loss aversion. The convex shape in losses results in a diminishing sensitivity to losses, associated with a decreasing value of losses. A drawback of CPT is the need to define a reference point. Although there is no consensus in the literature on how this reference should be chosen, one solution would be to consider the current travel time as the reference.

The results of Chapter 8 show that reference models can be used as DCMs and shed new behavioral and economic perspectives on the modeling of consumer choice behavior. Many experts feel that it pays off (in terms of insights gained) to study discrete choice behavior from different behavioral perspectives (Chorus, 2012b), and the reference models provide a formally tractable way of doing so. Firstly, the reference models can be used to derive economic outputs (willingness to pay, marginal effects and elasticities) with formulae generalizing those derived with the multinomial logit model. In the case of high unobserved heterogeneity, some reference models, such as the one defined with the Student distribution, retrieve consistent economic outputs, whereas the multinomial logit elicits implausible values. Secondly, the choice of the reference alternative impacts the parameter

estimation and also the different economic outputs, such as the elasticities. This dependence on the reference alternative allows the analyst to partially relax the Independance of Irrelevant Alternatives property. Thirdly, in comparison to the multinomial logit model, reference models allow the analyst to better select the explanatory variables and obtain a better fit to the data.

Testing a variety of models generates more flexibility in the modeling process and makes it possible to point out the model that best fits the data. Changing the cumulative distribution function is a way to model different responses to the explanatory variables. The simulated data show that some distributions comply with reference-dependent theory. All distributions, except the logistic one (that is, the one corresponding to the multinomial logit model), lead to reference-dependent models. The heavy tail distribution functions (Student and Laplace) are associated with a diminishing sensitivity to gains and losses. With the asymmetric Gumbel distribution function, the losses loom larger than the gains. Those three distributions also provide the best empirical performances. In conclusion, the reference models that are the most behaviorally sound are also the ones that retrieve the most consistent estimation results and economic outputs. However, the reference models still need to be evaluated using cross-validation methods (holdout method or K-fold cross-validation) to assess and compare the model's predictive performance.

Three empirical developments should also be tested in future research. Firstly, an asymmetric distribution function with heavy tails, such as the non-central Student distribution, should provide even better empirical results than the one already tested and fulfill all three features of reference-dependent theory. Secondly, a latent-class specification of reference models should be tested, since the reference alternative is heterogeneous in the population and *a priori* unknown to the analyst. Thirdly, reference models may be extended to generalize models other than the multinomial logit. For example, it is theoretically possible to include random parameters in reference models and, therefore, to generalize the mixed logit model.

If the main results presented in this part - a better selection of explanatory variables and more accurate outputs than the multinomial logit model - are further confirmed on other real and simulated datasets, then the use of reference models will have operational implications, in terms of marketing strategies and public policy design. Moreover, a collaborative work with Jean Peyhardi is under way to study the connection between reference models and random utility theory.

CHAPTER 7

A rank-dependent utility approach to model intra- and inter-individual heterogeneity in risky choices[1]

Abstract

Using a stated preferences survey, the objective of this paper is to investigate the intra- and inter-individual heterogeneity of mode choice, when travel time is subject to variability. By"inter-individual heterogeneity" is meant that people are different in terms of attitude toward risk and have different functional forms for the utility function. By "intra-individual heterogeneity" is meant that the behavior may be different even when performed by the same individual when faced with a different transport mode. Based on Rank-Dependent Utility Theory, the paper shows that the probabilities of delay associated with train trips are overweighted and the probabilities associated with car trips are underweighted. A latent-class model offers a somewhat different view: if, overall, car users do indeed weight the probabilities for train and car trips differently, the train users seem to be risk neutral and consider the objective probabilities that are presented in the stated preferences survey.

7.1 Introduction

All travelers face travel time uncertainty whatever their transport mode. However, the perception of travel time varies depending on travelers and transport modes, which leads to inter- and intra-individual heterogeneity. A strong hypothesis of the economic theory is that preferences are stable over time and context, which controverts intra-individual heterogeneity. However, behavioral

1. This chapter results from collaborative work with Matthieu de Lapparent (HEIG-VD, Switzerland) and has been submitted for a presentation at the 2018 Transportation Research Board Annual Meeting. This chapter is still a work in progress.

experiments show evidence of context-dependent preferences (Mazar et al., 2014; Tversky and Simonson, 1993). The objective of this paper is to model the two types of heterogeneity by integrating Rank-Dependent Utility Theory (RDUT) in a random utility model to explain mode choice. Inter-individual heterogeneity is taken into account by means of a latent-class model, while intra-individual heterogeneity is considered using probability weightings that depend on travel mode. In doing so, this paper also responds to the criticism of Rasouli and Timmermans (2014) regarding the lack of work on "context-dependent heterogeneity and taste differences among travelers".

Among the wide body of literature on travel time variability, only a few studies are concerned with mode choice (Bhat and Sardesai, 2006; König and Axhausen, 2002; Batarce and Ivaldi, 2014; de Lapparent and Ben-Akiva, 2014; Hensher et al., 2015). Moreover, most of them have been developed within a linear utility framework. The mean-variance and scheduling models, typically applied with stated preferences data, are the two dominant approaches to empirical measurement of the value of travel time variability (e.g., Börjesson et al., 2012; Bates et al., 2001).[2] These approaches make two strong assumptions: people are risk neutral and they consider objective probabilities in their decisions. Such assumptions have been challenged by the economic theory, and transportation research started to incorporate theories such as Rank-Dependent Utility Theory (RDUT; Quiggin, 1982; Yaari, 1987) or (cumulative) prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) a decade ago. de Palma et al. (2008) provide a general overview of choice analysis for both EUT and non-EUT approaches in the context of modeling travelers' risky choices and, in an editorial, de Palma et al. (2015) review various studies on risk and uncertainty in transport and urban economics.

Li and Hensher (2011) conduct a literature review on the use of (cumulative) prospect theory in different fields of transport: route choice, departure time choice, parents' choice of picking up their child and network equilibrium. They notice that most works only draw selectively on some contributions from prospect theory and highlight the absence of willingness-to-pay estimates and consideration of unobserved between-individual heterogeneity. As an answer, Hensher et al. (2011) estimate logit and mixed logit models which incorporate probability weighting and risk attitude in a route choice context and derive estimates of Value Of Time (VOT). Hensher and Li (2012) propose a similar work using RDUT. Razo and Gao (2013) specify a latent-class model which incorporates RDUT to explain strategic route choice under real-time traffic information. Jou and Chen (2013) apply cumulative prospect theory to freeway drivers' route choice behaviors. They study behavioral

2. For appraisal of transport projects, the former is the only relevant option since it relies on statistical measures of variability (for instance, standard deviation). However, the microeconomic foundations of how travel time variability generates costs are hidden. In contrast, the scheduling model is microeconomically sound.

heterogeneity by integrated individual variables to explain the parameters of the model (risk aversion, loss aversion, probability weightings). Xiao and Fukuda (2015) reformulate a general scheduling model employing rank-dependent utility theory and study the uncaptured costs of travel time variability. To my knowledge, there are only two applications of RDUT to mode choice. The first is de Lapparent and Ben-Akiva (2014) who compare different parametric forms to analyze risk attitudes and probability weightings. They conclude that commuters are weakly averse to small-time losses and that leisure travelers are risk neutral to small losses of time. However, in their data, only probabilities vary and not the delay times. The second is Hensher et al. (2015) who conduct a survey in which they ask the respondents to report three perceived travel times for car and public transport, as well as the probability associated with each travel time. They conclude that utility does not depend on the source of uncertainty (see Fox and Tversky, 1995; Tversky and Fox, 1995, for an introduction to this concept) and is the same for risk and uncertainty.

Although the literature using probability weighting functions in the field of transportation is growing, parameter estimations from travel behavior are still limited. The behavioral economics literature mainly bases its estimations on decision making under risk for monetary outputs. However, there are fundamental differences between simple gambling behavior and complex travel behaviors (Timmermans, 2010; Kemel and Paraschiv, 2013). Firstly, travel behavior is characterized by multiple attributes (travel time but also cost, frequency of public transport, comfort, and so on). Secondly, probabilities are generally unknown such that a decision is made under uncertainty. However, due to the repeated nature of travel decisions, the decision maker experiences the consequences of her decision and therefore acquires some information about the travel time distribution, which also makes travel decisions different from other decisions under uncertainty.

The motivation for analyzing heterogeneity stems from an upstream qualitative analysis.[3] A first observation from this analysis is that people seem to have different perceptions of travel time uncertainty depending on the travel mode. With regard to train travel, the departure time and the arrival time are written on a schedule. If the train arrives just a few minutes behind schedule, then some people consider that it is late. Other people are more flexible and consider that the train is late when the delay exceeds 15 minutes. The departure and arrival times of car travel, by contrast, are not written down on a schedule. Therefore, delay is only defined relative to the preferred arrival time (for instance, an appointment time). For instance, for a scheduled travel time of 90 minutes, if a train is 15 minutes late, then the traveler will perceive the delay. In the same conditions by car, the final travel

3. Two focus groups of 13 persons and 11 individual interviews.

time will probably not be categorized as delayed. Trains are therefore perceived as more frequently delayed than cars. A second observation is that, in the event of train delays, most people feel trapped and they have no control over the situation, whereas in the event of car delays, people feel responsible and are active in their car. In some situations, they may even change their route. Delays are therefore experienced more negatively during travel by train than by car, and will therefore be more pregnant in memories (Baumeister et al., 2001). A third observation is that this opposition in perceptions between trains and cars is especially true for car users. Train users seem to be more aware of the travel time distribution for both modes. In other words, the lessons learned from the qualitative phase is that the event "delay" is associated with unknown probabilities and is not accurately but fuzzily measured. To draw a parallel with lotteries, this would correspond to a situation in which the lottery amounts are known with different fuzziness and the probabilities as well.

As there does not appear to be any revealed preferences survey available to analyze heterogeneous behaviors when travel time is uncertain, the analysis is based on a stated preferences survey. The delay time and probabilities are clearly written down and decision makers are put in a situation of risk and not of uncertainty. The two following hypotheses are formulated: 1) people carry their real-life bias into the stated preferences survey, and since they usually misrepresent the delays and probabilities of delay, they will also do so in the survey; 2) the misrepresentation weighs on the probabilities and not on the perception of delays.

The contribution of this paper to the current knowledge on mode choice is four-fold. Firstly, to my knowledge, it is one of the first application of RDUT to mode choice with probabilities and delay times which vary across choice questions. Almost all other applications of RDUT or (cumulative) prospect theory have modeled route choice. In their application to mode choice, Hensher et al. (2015) consider probabilities which are stated by the respondent and, in de Lapparent and Ben-Akiva (2014), the delay times are fixed. Secondly, this work is one of the first to study intra-individual heterogeneity, and, more specifically to show how probability weightings depend on the travel mode. Thirdly, this work is one of the first to differentiate risk attitudes in a RDUT framework with a latent class model. Previous studies have modeled behavioral heterogeneity with alternative methods such as mixed logit (de Lapparent and Ben-Akiva, 2014; Hensher and Li, 2012; Hensher et al., 2011) or by allowing behavioral coefficients to depend on individual variables. For instance, by means of an ordered probit model, de Palma and Picard (2005) show that the level of risk aversion is impacted by key socio-economic factors, such as gender, employment status or purpose of the trip. Unlike mixed logit, latent class models "relax [the] requirement that the analyst make specific assumptions about the distributions of

parameters among individuals" (Greene and Hensher, 2003, p. 681). It is also computationally less demanding and easier to interpret. To my knowledge, two papers have already used latent-class models to analyze variability in travel time but the class does not define risk attitudes. Razo and Gao (2013) differentiate strategic and non-strategic decisions, strategic decisions referring to the traveler's recognition that certain detours can be taken in response to future information. Sun et al. (2012) define three risk-attitude classes for sequential choices concerning activity rescheduling: the risk avoider considers the worst case at each node, the risk seeker considers the more likely outcome and the risk neutral individual uses expected utility. In their study, there is thus no risk attitude parameter or probability weighting function. Fourthly, studying travel time variability taking into account risk aversion within a RDUT or (cumulative) prospect theory framework also enriches the interpretation of the VOT. Previous works have discussed how VOT increases with the size of the time change (Hjorth and Fosgerau, 2012) or the VOT distribution across the population (Hensher et al., 2011, 2015). To my knowledge, this work is the first to estimate VOTs taking into account a risk aversion parameter in a RDUT framework and to show how it evolves according to travel time.

The rest of the paper is organized as follows. Part 7.2 presents the main theories of choice under conditions of risk. Part 7.3 defines the model framework. Part 7.4 presents the data. Part 7.5 discusses the results. Part 7.6 concludes.

7.2 Theories of choice under conditions of risk

Risk-aversion or risk-seeking behaviors are taken into account in Expected Utility Theory (EUT) by means of a transformation $\phi(.)$ of the utility function. However, violations of EUT, revealed by Allais (1953) and others (see Denant-Boèmont and L'Haridon, 2013, for a review of violations to the EUT), suggest that the "psychological weight attached to an event [...] usually differs from the probability of that event" (Wu and Gonzalez, 1999, p.74).

To account for the perceptual translation of probability p_E associated with event E, non-linear probability weighting functions $w(p^E)$ have been introduced. The seminal work in this field is Kahneman and Tversky (1979) who developed prospect theory, within which the notion of risk in decision making is represented by a separable function $w(p^E) \times \phi(t^E)$ which defines the value of a prospect that delivers an outcome t^E with probability p^E. In this theory, people transform probabilities without taking into account the associated outcome. [4]

4. In prospect theory, formulas "turned out not to be sound because they imply violations of stochastic dominance in preferring less money to more money in manners that are not only normatively but also descriptively unwarranted" (de Palma et al., 2008, p.5).

By contrast, RDUT, suggested by Quiggin (1982) and fully described by Yaari (1987), proposes to determine the weighting function of the probabilities as dependent on the rank of the outcomes (see Diecidue and Wakker, 2001, for an intuitive presentation of RDUT). RDUT turns the valuable concept of probability weighting from the psychological literature into "a theory sound enough for economists to use" (de Palma et al., 2008). Outcomes are ranked from the worst to the best and the probability weights operate in a cumulative way since they depend on the rank of the associated outcome and on the probability. The probability weighting function usually observed in empirical works has an inverse S-shape (e.g., Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Wu and Gonzalez, 1999), which means an overweighting of small probabilities and underweighting of high probabilities.

7.3 Model framework

7.3.1 Embedding RDUT within a random utility model

For each alternative j ($j = 1, \cdots, J$), consider the prospect $\mathcal{X}_j = (E_1 : t_j^{E_1}, \ldots, E_S : t_j^{E_S})$ with S mutually exclusive events, which occur each with probability $p_j^{E_1}, \ldots, p_j^{E_S}$.

The program of the decision maker n ($n = 1, \ldots, N$) is to maximize her utility $U_{n,j}$

$$U_{n,j} = \sum_{E \in \mathcal{E}} W(p_j^E) \tilde{V}_{n,j}(\mathbf{x_n}, \tilde{\mathbf{y}}_{\mathbf{n,j}} | E, \tilde{\boldsymbol{\beta}}) + \varepsilon_{n,j}, \tag{7.1}$$

with p_j^E the probabilities associated with each event E ($E \in \mathcal{E} = (E_1, \ldots, E_S)$) for alternative j, $\tilde{V}_{n,j}$ the deterministic part of the utility which depends on a vector of individual variables $\mathbf{x_n}$ and on a vector of alternative-specific variables $\tilde{\mathbf{y}}_{\mathbf{n,j}}$, and $\varepsilon_{n,j}$ error terms which are assumed to be independently and identically distributed type I Extreme Value across individuals and alternatives: $\varepsilon_{n,j} \overset{\text{iid}}{\to} EV(0,1)$. $\tilde{\boldsymbol{\beta}}$ is the parameter vector. For the sake of simplicity, subscripts corresponding to the choice questions are initially omitted.

Under RDUT, the different travel times t_{E_s} ($s = 1, \ldots, S$) are ordered such that the least preferred outcome, that is, the longer travel time, is number 1 and the preferred outcome, the shortest travel time, is number S. For all $j = 1, \cdots, J$, the weighting probability function is

$$W(p_j^{E_s}) = \begin{cases} w(p_j^{E_s}) & \text{, if} \quad s = 1 \\ w(p_j^{E_1} + \cdots + p_j^{E_s}) - w(p_j^{E_1} + \cdots + p_j^{E_{s-1}}) & \text{, if} \quad s = 2, \ldots, S, \end{cases} \tag{7.2}$$

with $w(0) = 0$ and $w(1) = 1$.

The hypothesis is formulated that, for each alternative j, the travel time t_j faced by the decision maker can take two values ($S = 2$): a travel time with delay $t_j^{E_1} = \bar{t}_j$, which is the worst outcome and occurs with probability $p_j^{E_1} = p_j$; and a travel time with no delay, $t_j^{E_2} = \underline{t}_j$, which is the best outcome and occurs with probability $p_j^{E_2} = 1 - p_j$. Therefore, the probability weighting function is

$$W(p_j) = \begin{cases} w(p_j) & \text{, if} \quad s = 1 \\ 1 - w(p_j) & \text{, if} \quad s = 2. \end{cases} \tag{7.3}$$

Since the outcome of the events only impacts travel time, the other alternative-specific variables remain unchanged, such that Equation (7.1) can be re-written

$$U_{n,j} = f(p_j, \underline{t}_j, \bar{t}_j) + V_{n,j}(\mathbf{x_n}, \mathbf{y_{n,j}}|\boldsymbol{\beta}) + \varepsilon_{n,j}, \tag{7.4}$$

with $\mathbf{y_{n,j}}$ the vector of alternative-specific variables except the time, such that the vector $\tilde{\mathbf{y}}_{\mathbf{n,j}}^T = (t_j, \mathbf{y}_{\mathbf{n,j}}^{\mathbf{T}})$, $\boldsymbol{\beta}$ the truncated parameter vector such that $\tilde{\boldsymbol{\beta}}^{\boldsymbol{T}} = (\boldsymbol{\beta}_{time}, \boldsymbol{\beta}^{\boldsymbol{T}})$ and $V_{n,j}$ the truncated deterministic part of the utility function.

The risky component of the utility function is

$$f(p_j, \underline{t}_j, \bar{t}_j) = \beta_{time} \times \left[\left(1 - w(p_j)\right) \times \phi(\underline{t}_j) + w(p_j) \times \phi(\bar{t}_j)\right], \tag{7.5}$$

with $w(.)$ and $\phi(.)$ functions to be defined.

7.3.2 Functional form of $\phi(.)$

$\phi(.)$ is the value function and expresses the risk aversion of the decision maker. Constant relative risk aversion utility functions have been widely used in behavioral economics (e.g., Tversky and Kahneman, 1992; Holt et al., 2002). Specifically, with the following specification for all $t_j \in \{\underline{t}_j, \bar{t}_j\}$,

$$\phi(t_j) = \begin{cases} \frac{t_j^{1-\alpha}}{1-\alpha} & \text{, if } \alpha \geq 0 \text{ and } \alpha \neq 1 \\ \ln(t_j) & \text{, if } \alpha = 1, \end{cases} \tag{7.6}$$

α denotes the coefficient of relative risk aversion.

Under EUT, the risk attitude of a decision maker is completely determined by the curvature of the value function. A concave $\phi(.)$ suggests a risk-averse attitude and a convex $\phi(.)$ suggests a risk-seeking attitude, regardless of the outcome of the probabilities.

7.3.3 Functional form of $w(.)$

The decision weight $w(.)$ may take different functional forms (see de Lapparent and Ben-Akiva, 2014; Rasouli and Timmermans, 2014, for an overview). Given the objectives of the paper, two forms are investigated. [5] Firstly, $w(.)$ is modeled by means of a power transformation

$$w(p_j) = p_j^{\delta}, \delta > 0, \tag{7.7}$$

which denotes an overweighting of probabilities if $0 < \delta < 1$ and an underweighting if $\delta > 1$. δ is therefore expected to be lower than one for train trips to reflect pessimism and greater than one for car trips to reflect optimism. [6]

Secondly, $w(.)$ is modeled by means of the transformation proposed by Tversky and Kahneman (1992, T&K)

$$w(p_j) = \frac{p_j^{\delta}}{[p_j^{\delta} + (1-p_j)^{\delta}]^{1/\delta}}, \delta > 0. \tag{7.8}$$

If $\delta = 1$, then $w(p_j) = p_j$ and EUT applies. If $\delta > 1$, then probabilities are underweighted. If $0 < \delta < 1$, the decision weight function has an inverted S-shape, which means that outcomes with lower probabilities tend to be overweighted whereas outcomes with higher probabilities tend to be underweighted. Put differently, when the probability of delay p_j is below the threshold probability $p^{\star}$, then $w(p_j) > p_j$ and $1 - w(p_j) < 1 - p_j$. Therefore, if the probability of delay observed in the database is below the estimated threshold, that is, $p_j < p^{\star}$, then the travel time with delay weighs more than proportionally compared with the travel time with no delay. This reflects pessimism in perceiving the probability of delay.

The risk aversion attitude has to be interpreted in light of the weighting function form (see Cohen, 2015, for a discussion). For instance, if the dual theory of choice under risk (Yaari, 1987) applies, decision makers are weakly risk averse since the value function is linear in t, that is, $\alpha = 0$, and the probability of delay is overweighted, that is, $\delta < 1$ with the power transformation. If the decision maker is pessimistic enough, she might exhibit weak risk aversion even with a convex value function.

5. The Prelec functional form (Prelec, 1998) has also been tested but it provided inconsistent results, possibly due to the two parameters which have to be estimated with this function. Further research is needed to solve this issue. These three parametric functions - Power, T&K (or Quiggin) and Prelec - are also the ones considered in (de Palma et al., 2008).

6. Usually, in the gains domain, the inverse relationship is considered: $\delta < 1$ denotes an optimistic attitude and $\delta > 1$ a pessimistic attitude.

7.3.4 Tested hypotheses and models

Based on the observations made during the qualitative analysis, three hypotheses are formulated:

- **H1**: Probabilities of delays during travel by car are underweighted since the event "delay" is not well defined;
- **H2**: Probabilities of delays during travel by train are overweighted since the event "delay" is well defined and leads to negative experiences which are more easily retained than the favorable consequences of the event "no delay";
- **H3**: The distinction in the weighting of probabilities is more accurate for car users than for train users who are more likely to consider objective probabilities when making their choice.

To test these three hypotheses seven models are estimated. Firstly, six nested logit models are estimated. The two train alternatives (see Section 7.4) are grouped within a first nest while the car alternative is in a second nest. In the first model, people are supposed to be risk neutral and to consider the expected travel time as stated in the scenarios when making their choice. The second model fits into EUT, with a relative risk aversion coefficient. The third model fits into RDUT, with a power transformation of the probabilities. The fourth model is the same as the third model but estimates two distinct parameters for δ, that is, δ^{car} and δ^{train}, to take account of intra-individual heterogeneity. The fifth model fits into RDUT, with a T&K transformation of the probabilities. The sixth model is the same as the fifth model but estimates two distinct parameters for δ, as in the fourth model.

Secondly, a latent class model (Greene and Hensher, 2003) is specified to take into account inter-individual heterogeneity in risk attitudes, in addition to the intra-individual heterogeneity. Two classes are defined. In the first class, decision makers are risk neutral and base their decision on the expected travel time, as in the first nested logit model. In the second class, decision makers may not be risk neutral and have non-linear decision weights, with a T&K transformation, as in the sixth nested logit model. For decision maker n belonging to class k ($k = 1, 2$), the conditional probability of choosing option j in the choice question q ($q = 1, \cdots, Q$), is

$$\pi_{n,q,j}^{\{k\}} = \frac{\exp\left(\beta_{time}^{\{k\}}\left[(1-w(p_j))\times\phi(\underline{t}_j) + w(p_j)\times\phi(\overline{t}_j)\right] + V_{n,j}(\mathbf{x_n}, \mathbf{y_{n,j}}|\boldsymbol{\beta^{\{k\}}})\right)}{\sum_{i=1}^{J}\exp\left(\beta_{time}^{\{k\}}\left[(1-w(p_i))\times\phi(\underline{t}_i) + w(p_i)\times\phi(\overline{t}_i)\right] + V_{n,i}(\mathbf{x_n}, \mathbf{y_{n,i}}|\boldsymbol{\beta^{\{k\}}})\right)}, \tag{7.9}$$

with $\beta_{time}^{\{k\}}$ the parameter associated with travel time for class k and $\boldsymbol{\beta^{\{k\}}}$ the parameter vector

specific to class k associated with the other explanatory variables. $\phi(.)$ is defined in Equation (7.6) and $w(.)$ is defined in Equations (7.7) and (7.8). For class 1, two constraints are imposed: $\alpha = 0$ and $\delta = 1$.

For decision maker n, the probability $\tilde{p}_n^{\{k\}}$ of belonging to class k is defined by a logit model

$$\tilde{p}_n^{\{k\}} = \frac{\exp\left(\boldsymbol{\theta}^{\{k\}T}\mathbf{x}_\mathbf{n}^\star\right)}{\sum_{k'=1}^{2}\exp\left(\boldsymbol{\theta}^{\{k'\}T}\mathbf{x}_\mathbf{n}^\star\right)}, \tag{7.10}$$

where $\mathbf{x}_\mathbf{n}^\star$ denotes a vector of individual variables which enter the model for class membership and $\boldsymbol{\theta}^{\{k\}}$ is the parameter vector associated with class k.

The assumption is made that the answers to the Q questions given by the same decision maker are independent given the class assignment. [7] The contribution of decision maker n to the likelihood is

$$\pi_n^{\text{LC}} = \prod_{q=1}^{Q}\left[\tilde{p}_n^{\{1\}} \times \prod_{j=1}^{J}\left(\pi_{n,q,j}^{\{1\}}\right)^{c_{n,q,j}} + \tilde{p}_n^{\{2\}} \times \prod_{j=1}^{J}\left(\pi_{n,q,j}^{\{2\}}\right)^{c_{n,q,j}}\right], \tag{7.11}$$

with

$$c_{n,q,j} = \begin{cases} 1 & \text{, iif } U_{n,q,j} \geq U_{n,q,i},\ \forall i = 1, \cdots, J \\ 0 & \text{, otherwise.} \end{cases} \tag{7.12}$$

$U_{n,q,j}$ is the utility function defined in Equation (7.1) which omitted the index q for the sake of simplicity.

The loglikelihood of all observations over all individuals is

$$\mathcal{L}(\hat{\beta}) = \sum_{n=1}^{N} \ln \pi_n^{\text{LC}}. \tag{7.13}$$

7.4 Data

Between 2012 and 2015, a large travel survey was conducted among 36,000 inhabitants living in the Rhône-Alpes region (France) with geographic stratified sampling. Using this data set, 19,601 inhabitants who had recently traveled by train were asked to answer a web stated preferences survey. 2,396 volunteer respondents were asked to describe in detail (time, cost, purpose and origin), a trip they made by coach, train or car during the last month within the Rhône-Alpes region. Only respondents living in the Rhône-Alpes region, aged 18 or over, having a car and a driving license and whose trip was made or could have been made by train or coach were asked to answer the choice

7. To take into account the panel structure of the data, this assumption will be relaxed in a future work in which we intend to estimate the same models with Bayesian statistics.

questions, resulting in 1,215 answers.[8] The feasibility of a modal shift was assessed thanks to a database constructed with the software Musliw (Palmier, 2010) and containing travel time by public transport and car for each origin-destination pair of the region. Due to the low rate of train users in the population, they were oversampled with a face-to-face survey made in regional trains, using the quota sampling method (sex, age, motive, travel time and train line). 1,059 respondents began the questionnaire and 608 finished it. The loss is mainly due to trips which were not made entirely within the Rhône-Alpes region or persons not living in the region. After cleaning the data (i.e. deleting respondents who should not have been interviewed or who gave absurd answers), the answers of 1,788 respondents were retained. Eventually, 54% of the respondents are car users.

Each respondent had to choose between three travel modes, which were described in terms of travel mode, cost, time, probability and time delay, frequency, clock-face timetable and comfort. To avoid a cognitive burden, attributes describing the journey were split into three exercises. In Exercise 1, on which this chapter focuses, the two train alternatives (A and B) and the car alternative vary according to travel cost, regular travel time, delay time, probability of delay, and frequency (only for the two train alternatives; see Figure 7.1).

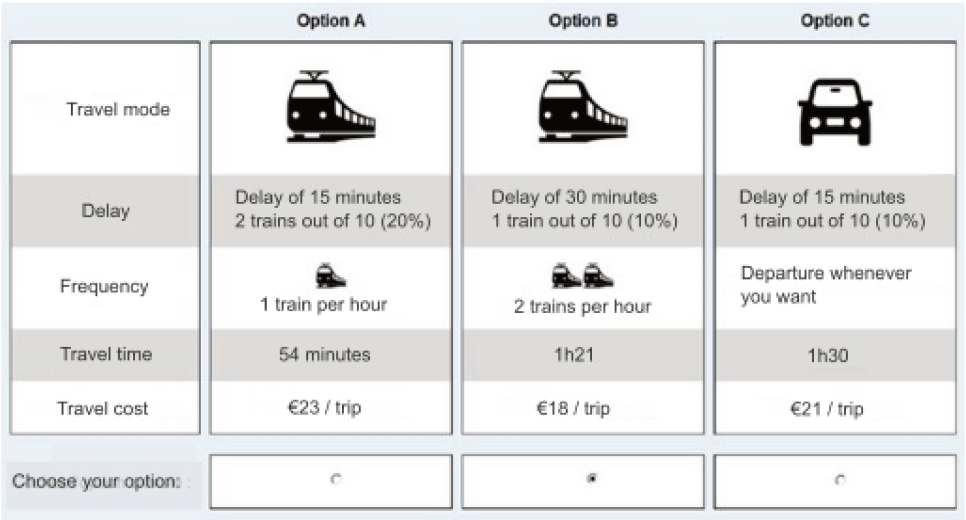


FIGURE 7.1 – Example of choice question for Exercise 1

Travel cost includes public transport ticket or pass, gasoline, parking cost and toll. Travel time is defined from origin to destination (including access time, egress time, waiting time and in-vehicle time). The delay time has one of the following values: 10, 15 or 30 minutes. The delay probability has one of the following values: 5%, 10% or 20%, which reflect what is actually observed for trains in

8. See Chapter 2 for more details on the survey and Chapter 3 for a comparison of descriptive statistics between the samples.

the Rhône-Alpes region. The frequency of trains is one every two hours, one every hour, two every hour and four every hour. To improve the efficiency of the design, a Bayesian efficient design was implemented (Rose and Bliemer, 2007). A priori weights of attributes were taken from the literature and adjusted during the pilot tests. Respondents had to answer to five choices questions in Exercise 1, leading to a database with 8,933 observations, since a few respondents did not answer all five questions. The choice questions are personalized with the data collected for the reference journey. In particular, levels of time and cost attributes are pivoted around the values collected for the reference journey. One of the three alternatives was systematically a *status quo* alternative with the mode, travel time and travel cost identical to the reference journey. The percentage of respondents who chose a train alternative was 47%, while 53% chose the car alternative. The descriptive statistics of the sample are shown in Table 7.1.

TABLE 7.1 – Descriptive statistics

Variable definition	**Label**	**Mean**	**S.D.**	**Min**	**Max**
Alternative-specific variables					
Travel time by train A (in minutes)	Train time	67.53	51.89	5.00	325.00
Travel time by train B (in minutes)	Coach time	67.15	52.43	5.00	375.00
Travel time by car (in minutes)	Car time	59.07	39.04	4.00	330.00
Travel cost by train A (in €)	Train cost	8.57	7.28	1.00	62.00
Travel cost by train B (in €)	Coach cost	8.53	7.35	1.00	78.00
Travel cost by car (in €)	Car cost	10.34	9.17	1.00	71.00
Individual variables					
Age (in years)	Age	47.79	16.10	18.00	90.00
Gender (1 if man, 0 if woman)	Gender	0.50			
Car user for the reference trip (1 if yes, 0 otherwise)	Type_car	0.54			
Trip's purpose is obligatory (work or study) (1 if yes, 0 otherwise)	Obligatory	0.42			

7.5 Results

The estimation of the models uses the software Pythonbiogeme (Bierlaire, 2016). Table 7.2 shows the estimation results. The car alternative is the reference alternative for the mode choice model and the first class is the reference class for the class membership model. In all models, the car alternative is chosen as the reference alternative.

TABLE 7.2 – Estimation results for the seven models

	Nested 1	**Nested 2**	**Nested 3**	**Nested 4**	**Nested 5**	**Nested 6**	**Latent-class 7**	
	Expected time	EUT	Power hom.	Power het.	T&K hom.	T&K het.	Class 1: Expected time	Class 2: T&K het.
Mode choice model								
ASC_{Train}	-0.58	-0.83	-0.86	-0.58	-0.86	-0.76	-2.20	-2.07
Time (Train)	-0.02	-0.36	-0.29	-0.33	-0.35	-0.36	-0.04	-0.73 NS
Time (Car)	-0.02	-0.43	-0.35	-0.38	-0.42	-0.43	-0.10	-0.70 NS
Cost	-0.08	-0.08	-0.07	-0.07	-0.08	-0.08	-0.20	-0.09
Freq	0.19	0.20	0.20	0.20	0.20	0.20	0.35	0.35
Car user	-3.00	-3.11	-3.07	-3.10	-3.10	-3.12	-2.27	-1.97
Age	0.04	0.04	0.04	0.04	0.04	0.04	-0.03 NS	0.05
Age^2	-0.0004	-0.0004	-0.0004	-0.0004	-0.0004	-0.0004	0.0011 NS	-0.0008
Gender	-0.17	-0.18	-0.18	-0.18	-0.18	-0.18	-0.34 NS	-0.18 NS
Obligatory	0.36		0.38	0.38	0.38	0.38	0.81	0.39
α		0.511	0.479	0.497	0.508	0.514		0.567
δ^{All}			0.375		0.783			
δ^{Train}				0.233		0.762		0.726
δ^{Car}				4.780		5.170		3.590
μ_{Train}	2.24	2.16	2.19	2.16	2.11	2.1		
λ_{Train}=$1/\mu_{Train}$	0.45	0.46	0.46	0.46	0.47	0.48		
Class membership model								
ASC							-1.01	
Car user							3.15	
Goodness of fit								
Loglikelihood	-6,641	-6,607	-6,607	-6,601	-6,616	-6,614	-6,540	
AIC	13,304	13,237	13,240	13,230	13,257	13,257	13,130	
$\bar{\rho^2}$	0.322	0.326	0.325	0.326	0.325	0.325	0.331	
Rank	5	3	3	2	4	4	1	

Notes: ASC: Alternative-Specific Constant; hom.=homogeneous; het.=heterogeneous; AIC: Akaike Information Criterion. EUT: Expected Utility Theory; Power: Power probability weighting function; T&K: Tversky and Kahneman probability weighting function. All variables are significant at the 0.1 % level, except the one indicated as Non-Significant (NS). The δ parameters are all significantly different from one at the 0.1 % level. The ranking is calculated with the Ben-Akiva and Swait (1986) test.

7.5.1 Goodness of fit

The adjusted McFadden ρ^2 ($\bar{\rho}^2$) and the Akaike Information Criterion (AIC) are used to rank the models in terms of goodness of fit.[9] The models being non-nested, one solution is to follow the method proposed by Ben-Akiva and Swait (1986), which determines the asymptotic upper bound of the probability that the difference between the $\bar{\rho}^2$ is greater than the observed one.[10] The difference between $\bar{\rho}_2$ is sufficient to conclude that the latent class model outperforms all other models, therefore showing the importance of taking into account intra- and inter-individual heterogeneity. Then comes the nested model with a power transformation of travel time and heterogeneous weighting parameters. Third is the EUT nested model and the nested model with a power transformation model and homogeneous weighting parameters. Both are similar in terms of data fit. The two nested models with the T&K transformation are also similar. The worst ranking model is the nested model, which considers the expected travel time.

7.5.2 Estimation results

The parameter estimates for the nested logit models are robust across specifications and consistent with expectations: all coefficients are significant, with the expected sign. For instance, for Model 1, the probability of choosing an option decreases as travel time and travel cost increase. The higher the frequency of trains, the higher the probability of choosing a train alternative. As age increases, the train alternatives become more attractive but with a decreasing marginal effect. Women and people who travel for work or studies are more likely to choose the train as a transport mode. Moreover, the value (1-λ_{Train}), which can be used as a measure for correlation (Train, 2009), shows that the two train alternatives are not independent. The nested structure is therefore well adapted to the data.

For Models 2 to 7, the relative risk aversion coefficients are positive, significantly different from zero and comprised in a narrow range, between 0.479 and 0.567, which show the robustness of the results across the models. Since the risk refers to a "loss" of time, a positive coefficient is associated with a convex utility function (see Figure 7.2). For instance, people are willing to pay less to save 10 minutes on a 30-minute trip than on a 60-minute trip[11]. These coefficients are consistent with the literature. Indeed, the risk aversion coefficients reviewed in Li and Hensher (2011) are comprised between 0.26

9. The Bayesian Information Criterion (BIC) is not reported here but provides the same ranking.

10. $P(\bar{\rho}_2^2 - \bar{\rho}_1^2 > z) \leq \Phi\{-[-2NzlnJ + (K2 - K1)]^{1/2}\}$, with N the sample size, J the number of alternatives, z the observed difference between $\bar{\rho}^2$, K_l the number of parameters in Model l ($l = 1, 2$), and Φ the standard normal cumulative distribution function. The table with the upper bound of the p-values of the test is available on request from the authors.

11. Based on theoretical assumptions, Beaud et al. (2016) consider risk-averse decision makers with a concave value function. They therefore derive the opposite result: a VOT which increases with travel time.

and 0.89 (and even 2.07 in Senbil and Kitamura, 2005). Hensher and Li (2012) estimate a relative risk aversion coefficient equal to 0.29 with a logit model and 0.36 with a mixed logit model. Hensher and Li (2012) interpret these coefficients as denoting risk-seeking attitudes. However, under RDUT, since risk attitude is probability-dependent, a different interpretation is possible but has not been addressed in the transport literature. One exception is de Lapparent and Ben-Akiva (2014) who discuss weak risk aversion only in the dual theory framework of Yaari (1987). Indeed, although people are willing to take risks in hopes of having a shorter travel time when faced with two unknown travel times, they may still be weakly risk-averse and prefer average travel time to travel time prospect.

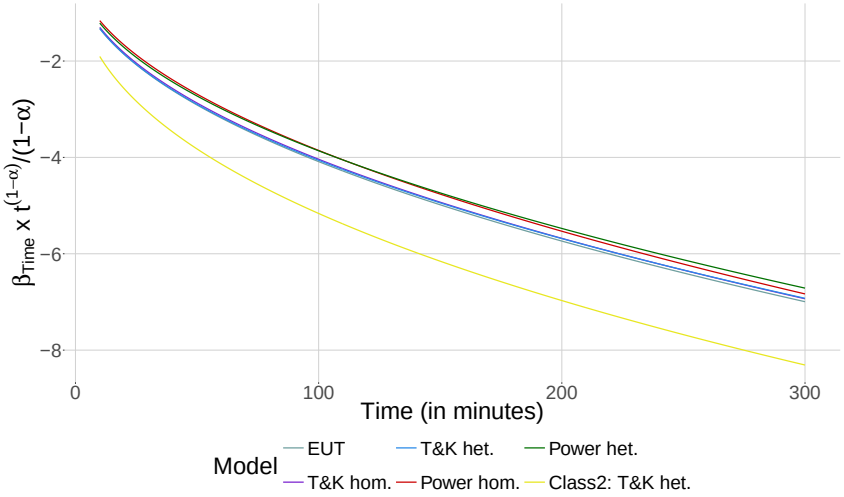
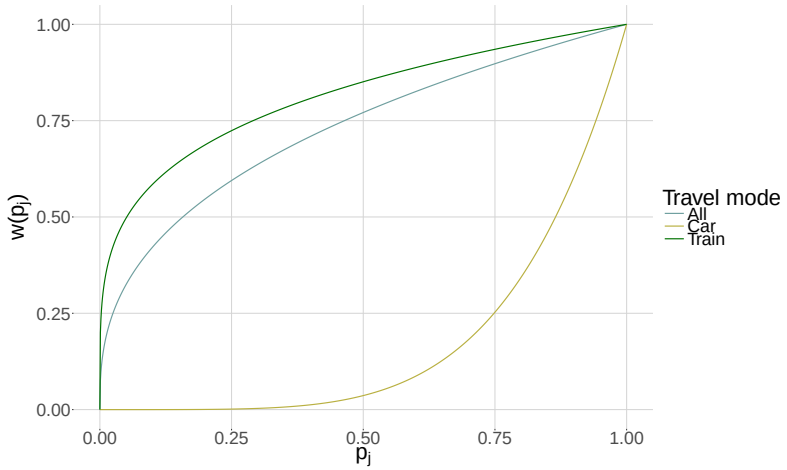


FIGURE 7.2 – Desutility (in €) associated with travel time for the six models taking into account a risk aversion parameter.
Notes: hom. = homogeneous weighting function across modes; het. = heterogeneous weighting function across modes

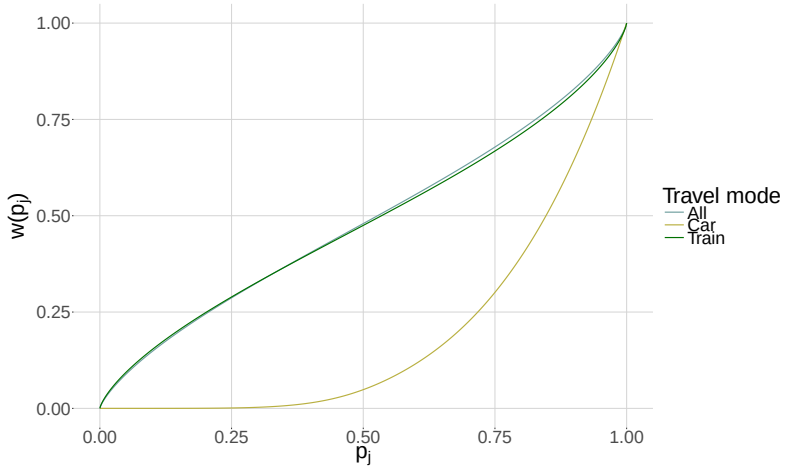
All δ parameters estimated in Model 3 to 7 are significantly different from one. For Model 3, the coefficient δ is lower than one, indicating that probabilities are overall overweighted (see Figure 7.3). People perceive delays as more probable than they are. Considering a power probability weighting function, de Lapparent and Ben-Akiva (2014) find a similar result for commuters.

However, Model 4 reveals a strong heterogeneity across travel modes: probabilities are not perceived the same way for train and for car trips. For car trips, probabilities are underweighted, which supports Hypothesis 1 and, for train trips, probabilities are overweighted, which supports Hypothesis 2.

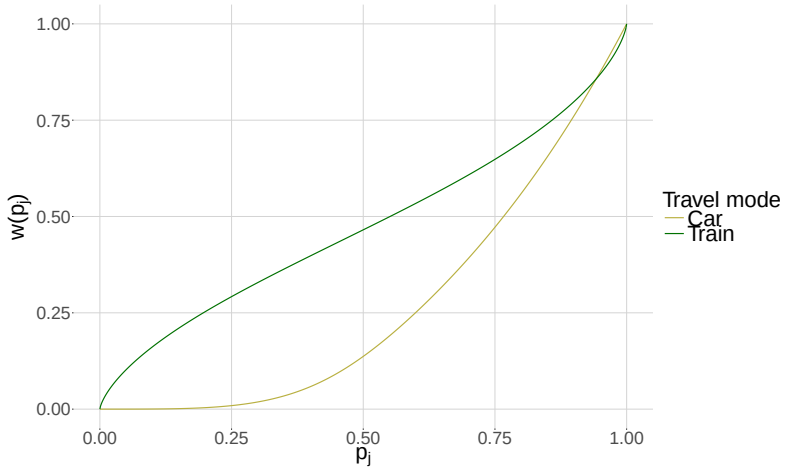
Model 5 makes it possible to differentiate low and high probability weightings. Since the coefficient $\delta = 0.783$ is lower than one, low probabilities are overweighted whereas high probabilities are



(a) Nested logit, power transformation of the probability weighting function (Models 3 and 4)



(b) Nested logit, T&K transformation of the probability weighting function (Models 5 and 6)



(c) Latent-class model, Class2, T&K transformation of the probability weighting function (Model 7)

FIGURE 7.3 – Probability weighting functions for the three models with heterogeneous weights across modes. Notes: p_j is the probability of delay associated with the travel mode j ($j \in \{\text{Car}, \text{Train}\}$); $w(p_j)$ is the weighted probability.

underweighted, with a threshold [12] $p^\star = 0.418$. The results for train trips are consistent with the theory and with previous findings considering a T&K probability weighting function. In the loss domain, Tversky and Kahneman (1992) estimate $\delta = 0.69$. When applied to the transport field, similar values are elicited. For example, Jou and Chen (2013) find that δ is comprised between 0.36 and 0.66 depending on the trip purpose and the traffic usually encountered on the freeway. Razo and Gao (2013) find a value of 0.55. As in Model 5, these references exhibit δ coefficients different from one, which shows that EUT does not apply. In the literature, two exceptions are identified. Firstly, de Lapparent and Ben-Akiva (2014) show that for leisure purposed, EUT applies. Secondly, Hensher and Li (2012) estimate $\delta = 2.126$ with a multinomial logit and $\delta = 5.311$ with a mixed logit model. This result is similar to the one found for car trips in Model 6.

Indeed, Model 6 confirms and refines the heterogeneity unveiled by Model 4. For car trips, there is a clear underweighting of probabilities ($\delta^{Train} = 5.17$). For train trips, the same result is found as in the general case but with a somewhat lower coefficient ($\delta^{Car} = 0.762$) and, therefore, a lower threshold: $p^\star = 0.409$. In the database, regular travel times are associated with high probabilities (0.8 to 0.95), whereas delays are associated with low probabilities (0.05 to 0.2). This means that the probabilities of delay are overweighted, placing the train mode at a disadvantage in regard to reliability perception.

Model 7 further investigates heterogeneity by defining two classes that behave differently. The probability of belonging to the first class, in which travel time is considered as equal to its expected value, is higher for train users than for car users. Hypothesis 3 is therefore supported by the data: the distinction in the weighting of probabilities is more accurate for car users than for train users who are more likely to consider objective probabilities when making their choice. Train users evaluate travel time prospect accounting only for the corresponding expected travel time. Class 2 accounts for 60% of the sample, which implies that more than half of respondents do not make their decision according to EUT. In class 2, age and gender do not impact mode choice, since the associated coefficients are not significantly different from zero. The coefficients associated with the travel time variables are not significant, which means that, overall, car users are insensitive to travel time. An explanation relies on the nature of the data: car users tend to stick to their actual mode choice, even if the travel time by train is considerably shorter. For class 2, the probability weighting function indicates that probabilities under 0.394 are overweighted and probabilities above 0.394 are underweighted.

In conclusion, the results indicate that individuals, and especially car users, are risk-seekers for car trips. Indeed, the utility function is convex and the probability function is concave. If such a result

12. The value of the threshold is obtained by resolving the equation $w(p_j) = p_j$, for $0 < p_j < 1$

might seem surprising, it is consistent with Tversky and Kahneman (1992) finding that decision makers are risk-averse in gains but risk-seekers in losses. People would rather choose a travel time that has a 50% chance of being 60 minutes and a 50% chance of being 120 minutes than a travel time which has a 100% chance of being 90 minutes. With regard to uncertain travel time, de Palma and Picard (2005) also find that some travelers (slightly less than one third of their sample) are risk seekers. Concerning train trips, further work is needed to compare the curvatures of the utility and the probability weighting functions.

7.5.3 Value of time

Travel times in a range from 10 to 300 minutes are simulated. Overall, the estimated VOTs are consistent with previous works (Quinet et al., 2014; Wardman et al., 2012; Shires and De Jong, 2009).

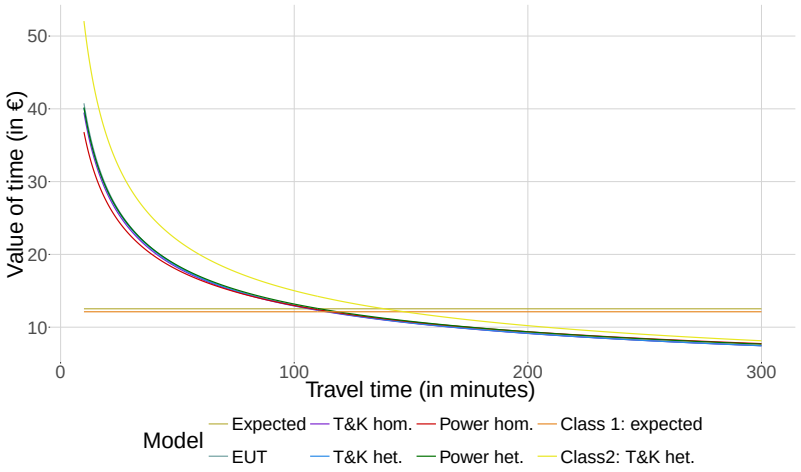
For the nested logit models taking into account risk aversion, the mean VOT for car trips is almost €3 higher than the mean VOT for train trips (see Table 7.3, lines two to six). Consistent with the finding that the first class of the latent class model is favored by train users, the value for train trips is much lower (€12.1/hour) than the VOT for car trips (€30.4/hour). Conversely, in the second class, the two values are similar (around €14.5/hour).

For the models including a risk aversion parameter, the diminishing sensitivity of utility to travel time implies a decreasing VOT (see Figure 7.4). VOTs range from €53/hour for short travel times of 10 minutes to €7.4/hour for long travel times of five hours. This result is intuitive: people are ready to pay more to travel for 10 minutes instead of 70 minutes than to travel for 5 hours instead of 6 hours.

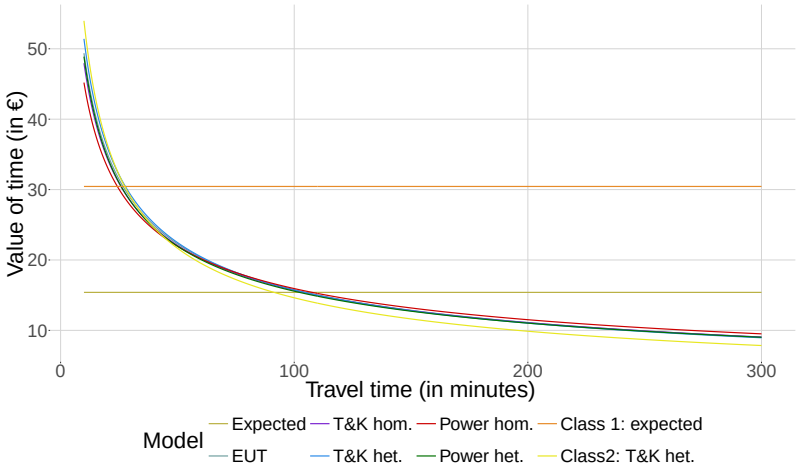
TABLE 7.3 – Value of time (in €/hour)

Model	Value of time (train)	Value of time (car)
Expected	12.53	13.40
EUT	12.68	15.36
Power hom.	12.56	15.43
Power het.	12.83	15.20
K&T hom.	12.56	15.25
K&T het.	12.63	15.41
Class 1 Expected	12.12	30.44
Class 2 K&T het.	14.72	14.45

Notes: hom.=homogeneous; het.=heterogeneous; EUT: Expected Utility Theory; Power: Power probability weighting function; T&K: Tversky and Kahneman probability weighting function. The delay probability is set to 0.1 and the delay time is set to 10 minutes.



(a) Value of time: train



(b) Value of time: car

FIGURE 7.4 – Value of time associated with the train mode (a) and the car mode (b) for the seven estimated models (including two classes for the latent class model).
Notes: hom. = homogeneous weighting function across modes; het. = heterogeneous weighting function across modes. The delay probability is set to 0.1 and the delay time is set to 10 minutes.

7.6 Conclusion

Using a choice experiment with choices among train and car modes, this paper investigates the intra- and inter-individual heterogeneity in behavior toward risk. The results highlight the difference of perceptions of probabilities across modes: delay probabilities associated with car trips are under-evaluated, whereas they are over-evaluated for train trips, at least for small probabilities such as those observed in the Rhône-Alpes region. The analysis of inter-individual heterogeneity uncovers two attitudes toward risk and travel time. Some individuals, mainly train users, are risk-neutral and consider the objective expected VOT. Other individuals, mainly car users, are risk-seekers for car trips but do not seem to be sensitive to changes in travel time. This behavior probably results from the use of stated preferences data and needs to be further investigated. The VOTs are consistent with the literature. This work also shows that the VOTs decrease with travel time and are higher for car trips than for train trips, especially for train users.

There are two key directions in which future research might build on the work presented here. Firstly, empirical investigation could consist in a more in-depth investigation of heterogeneity by means of alternative models, such as mixed logit to study the distribution of the relative risk aversion parameter, or the integration of latent variables, such as perceptions of time use or feelings during travel by public transport modes. Using the same method as in this paper, it would also be interesting to test the robustness of the results on other datasets. Secondly, only selective features of cumulative prospect theory have been introduced in the models. Since, the stated preferences survey is personalized with the characteristics of an actual trip, a full integration of cumulative prospect theory (Tversky and Kahneman, 1992) might be considered. The current travel time may be considered as the reference, and the different travel times presented in the choice questions may be framed as gains and losses to study reference dependence and loss aversion.

CHAPTER 8

Reference models: an economic and behavioral perspective[1]

This chapter is organized in three sections. The first section is an introduction to reference models from a statistical and discrete choice perspective. The second section is an empirical application of reference models to analyze mode choice and derive economic outputs. The third section is a work in progress and presents only preliminary ideas on the theoretical economic and behavioral interpretation of reference models. A latent-class specification of reference models is also proposed to avoid having to choose a specific alternative as reference for the whole sample.

8.1 An introduction to reference models

Reference models are a generalization of the multinomial logit model and have been developed by Peyhardi et al. (2015). Unlike previous generalizations of the multinomial logit model (nested logit, mixed logit, multinomial probit and so on), all based on a different specification of the error term in a random utility model framework, reference models are defined as a new family of Generalized Linear Models (GLM; see Appendix H) designed to analyze nominal response variables. By comparison to the multinomial logit model, they are defined by changing the link function. Their ability to model nominal response variables makes them good candidates to analyze discrete choices.

8.1.1 Reference models: a statistical approach

Assume a choice set composed of J alternatives. To simplify, individual subscripts are omitted. The probability of choosing alternative j ($j = 1, \cdots, J$) is denoted π_j and is explained by p explanatory

1. Section 8.2 is a paper co-authored with Iragaël Joly and Jean Peyhardi (University of Montpellier), which will be submitted to the "Review of Economics and Statistics". It has been presented at the WCTR in 2016 and at the 2017 Transportation Research Board Annual Meeting (see Bouscasse et al., 2017). For the sake of simplicity, notations throughout Chapter 8 are the same as those of this paper, in particular, with no bold symbol for vectors and with alternative-specific variables denoted by ω instead of y.

variables. π is the J vector $(\pi_1, \ldots, \pi_J)^T$. The design matrix[2] Z contains the structure of the p explanatory variables. In the general case, it is a $(J-1) \times p(J-1)$ matrix. β is a $p(J-1)$ vector of parameters. The $(J-1)$ linear predictor η is defined as

$$\eta = Z\beta. \tag{8.1}$$

The canonical link function for nominal response variables is the multinomial logit function (see Appendix H):

$$\begin{aligned} g \quad &: \quad \mathcal{M} \longrightarrow \mathbb{R}^{J-1} \\ &\quad\quad \pi \longmapsto \eta, \\ &\text{with} \\ g_j(\pi) &= \log\left(\frac{\pi_j}{1-\sum_{i=1}^{J-1}\pi_i}\right), \forall j \neq J, \\ \text{and } \mathcal{M} &= \{\pi \in]0,1[^J \sum_{j=1}^{J}\pi_j = 1\}. \end{aligned} \tag{8.2}$$

For instance, the multinomial logit model and the mixed logit model are defined using the canonical link function. Using the equality $\pi_J = 1 - \sum_{j=1}^{J-1}\pi_j$, for all $j \neq J$, $g_j(\pi)$ can be written as a composed function

$$\begin{aligned} g_j(\pi) &= \log\left(\frac{\pi_j}{\pi_J}\right), \\ &= \log\left(\frac{\pi_j/(\pi_j+\pi_J)}{\pi_J/(\pi_j+\pi_J)}\right) \\ &= \log\left(\frac{r_j(\pi)}{1-r_j(\pi)}\right) \\ &= F^{-1} \circ r_j(\pi) \end{aligned} \tag{8.3}$$

with F the Cumulative Distribution Function (cdf) of the logistic distribution and r_j a ratio of probabilities. F and r_j are respectively defined in Equations (8.4) and (8.5):

$$F(x) = \frac{1}{1+\exp(-x)} \quad F^{-1}(y) = \log\left(\frac{y}{1-y}\right) \tag{8.4}$$

$$r_j(\pi) = \frac{\pi_j}{\pi_i + \pi_J}, \quad \forall j \neq J. \tag{8.5}$$

Equation (8.3) assumes that F is the logistic cdf. However, other continuous and strictly increasing cdfs, such as the normal, Laplace, Cauchy, Gumbel or Student cdfs, can be chosen to define a new

2. See Section H.2.2.2, p.364 in Appendix H and Remark 1, p.236.

family of models, the reference models. The name "reference" is related to the ratio's dependence on the probability of the reference alternative J.

More generally, three other ratios can be chosen, which also define families of GLMs adapted to the analysis of qualitative response variables (Peyhardi et al., 2015):

- the cumulative ratio: $r_j(\pi) = \pi_1 + \cdots + \pi_J, j \neq J$. It is used in the proportional hazard logit model, which provides hazard in discrete time;
- the sequential ratio: $r_j(\pi) = \frac{\pi_j}{\pi_j + \cdots + \pi_J}$. It is used in the sequential logit, a model which proceeds as if decisions were made in a sequence of stages, such as the highest achieved level of education;
- the adjacent ratio: $r_j(\pi) = \frac{\pi_j}{\pi_j + \pi_{j+1}}$. It is used in the adjacent logit, which models the probability of the transition from one category to another.

The cumulative, sequential and adjacent ratios are designed to model ordinal response variables, whereas the reference ratio is appropriate to model nominal response variables, such as mode choice. Indeed, there is no natural order between mode alternatives such as train, car or coach.

These models can be characterized by a triplet (r, F, Z) in which r defines the ratio, F a cdf, and Z a design matrix. The multinomial logit model is defined by the triplet (reference, logistic, Z) in which Z contains individual and alternative-specific variables. The family of reference models is defined by the triplet (reference, F, Z).

8.1.2 Reference models: a discrete choice approach

In Section 8.1.1, reference models are derived from the canonical link function of GLM models designed for qualitative response variables. Reference models can be similarly derived with a discrete choice approach. Discrete choice models are based on the assumption that each individual maximizes her utility

$$U_j = \eta_j^\star + \varepsilon_j, \tag{8.6}$$

where ε_j is the error term and $\eta_j^\star$ is the deterministic part of the utility, defined as

$$\eta_j^\star = \alpha_j^\star + x^T \delta_j^\star + \omega_j^T \gamma, \ \forall j = 1, \cdots, J, \tag{8.7}$$

with

- $\alpha_j^\star$ the alternative-specific constant for alternative j;
- x the k vector containing the individual characteristics;

- $\delta_j^\star$ the k vector with the coefficients related to each individual characteristic. Each alternative j is associated with a different slope;
- ω_j the l vector containing the alternative-specific attributes of alternative j;
- γ the l vector related to each attribute. To simplify, the hypothesis is made that the slopes do not vary according to the alternative.
- $p = 1 + k + l$.

The probability of choosing option j is

$$\pi_j = \Pr(\max U_j \geq U_k,\ \forall k \neq j), \forall j = 1, \cdots, J. \tag{8.8}$$

In the multinomial logit model, the error term ε_j is i.i.d. distributed according to a G.E.V. type 1 distribution. The probability function of the multinomial logit model is therefore

$$\pi_j = \frac{\exp(\eta_j^\star)}{\sum_{i=1}^{J} \exp(\eta_i^\star)},\ \forall j = 1, \cdots, J. \tag{8.9}$$

Assume that J is the reference alternative and, for identification purposes, consider η_j, defined as

$$\begin{aligned} \eta_j &= \eta_j^\star - \eta_J^\star \\ &= \underbrace{(\alpha_j^\star - \alpha_J^\star)}_{\alpha_j} + x^T \underbrace{(\delta_j^\star - \delta_J^\star)}_{\delta_j} + \underbrace{(\omega_j^{*T} - \omega_J^{*T})}_{\tilde{\omega_j^T}} \gamma,\ \forall j \neq J. \end{aligned} \tag{8.10}$$

Therefore, Equation (8.9) can be rewritten

$$\pi_j = \begin{cases} \frac{\exp(\eta_j)}{1 + \sum_{i=1}^{J-1} \exp(\eta_i)} & ,\ \forall j \neq J, \\ 1 - \sum_{1}^{J-1} \pi_j & ,\ \text{for } j = J. \end{cases} \tag{8.11}$$

Remark 1 *In this particular case, the design matrix Z is*

$$Z = \begin{pmatrix} 1 & & x^T & & & \tilde{\omega}_1^T \\ & \ddots & & \ddots & & \vdots \\ & & 1 & & x^T & \tilde{\omega}_{J-1}^T \end{pmatrix}.$$

The vector of parameters is $\beta^T = (\alpha_1, \cdots, \alpha_{J-1}, \delta_1^T, \cdots, \delta_{J-1}^T, \gamma^T)$.

Using the expression of π_j, the ratio $r_j(\pi)$ can be written

$$\begin{aligned} r_j(\pi) &= \frac{\pi_j}{\pi_j + \pi_J} \\ &= \frac{\pi_j}{\pi_j + (1 - \sum_{k=1}^{J-1} \pi_k)} \\ &= \frac{\exp(\eta_j)}{1 + \exp(\eta_j)} \\ &= F(\eta_j), \ \forall j \neq J, \end{aligned} \tag{8.12}$$

with F the logistic cdf. As previously, F can be replaced by alternative cdfs. All regression models with a qualitative response variable identified by Tutz (2012) or Agresti (2013) share the generic equation:

$$r_j(\pi) = F(\eta_j), \ \forall j \neq J \tag{8.13}$$

with r_j the cumulative, sequential, adjacent or reference ratio, and F a continuous and strictly increasing cdf.

The probability function of a reference model can be derived from Equation (8.12). Let us express π_j as a function of π_J

$$\begin{aligned} \frac{\pi_j}{\pi_j + \pi_J} &= F(\eta_j), \ \forall j \neq J \\ \Leftrightarrow \pi_j &= \frac{F}{1 - F}(\eta_j)\pi_J. \end{aligned} \tag{8.14}$$

π_J can be isolated from Equation (8.14):

$$\begin{aligned} \sum_{j=1}^{J-1} \pi_j &= \pi_J \sum_{j=1}^{J-1} \frac{F}{1 - F}(\eta_j) \\ \Leftrightarrow 1 - \pi_J &= \pi_J \sum_{j=1}^{J-1} \frac{F}{1 - F}(\eta_j) \\ \Leftrightarrow \pi_J &= \frac{1}{1 + \sum_{j=1}^{J-1} \frac{F}{1-F}(\eta_j)}, \end{aligned} \tag{8.15}$$

and replaced by its expression

$$\pi_j = \frac{\frac{F}{1-F}(\eta_j)}{1 + \sum_{j=1}^{J-1} \frac{F}{1-F}(\eta_j)}, \ \forall j \neq J. \tag{8.16}$$

Equations (8.15) and (8.16) define the probability function of reference models.

8.2 A new family of regression models for categorical data: an application to travel mode choice

Abstract

This paper investigates the statistical and economic implications of the reference models, a new family of regression models dedicated to non-ordered alternatives. The link function of the reference models is that of the multinomial logit model replacing the logistic Cumulative Distribution Function (cdf) by other cdfs (e.g., Gumbel, Student). We derive usual economic outputs and study the IIA property. The use of alternative cdfs leads to a better data fit because of their different characteristics (symmetry/asymmetry, heavy/thin tails). The Student cdf is the only one to retrieve consistent outputs on our dataset, with high unobserved heterogeneity.

8.2.1 Introduction

Discrete choice models (DCMs) are a valuable tool for both understanding how choices are made and deriving behavioral outputs for economic analysis and valuation. DCMs have been applied in a variety of fields (transportation, energy, food consumption, marketing, and so on) with varied objectives (preferences analysis, value elicitation, estimation of demand elasticities, market share prediction, and so on). Their use is mainly applied to model consumer choice behavior but also applied to industrial economics. For example, Thisse and Billot (1995) use DCMs to analyze product differentiation and agent selection within an organization.

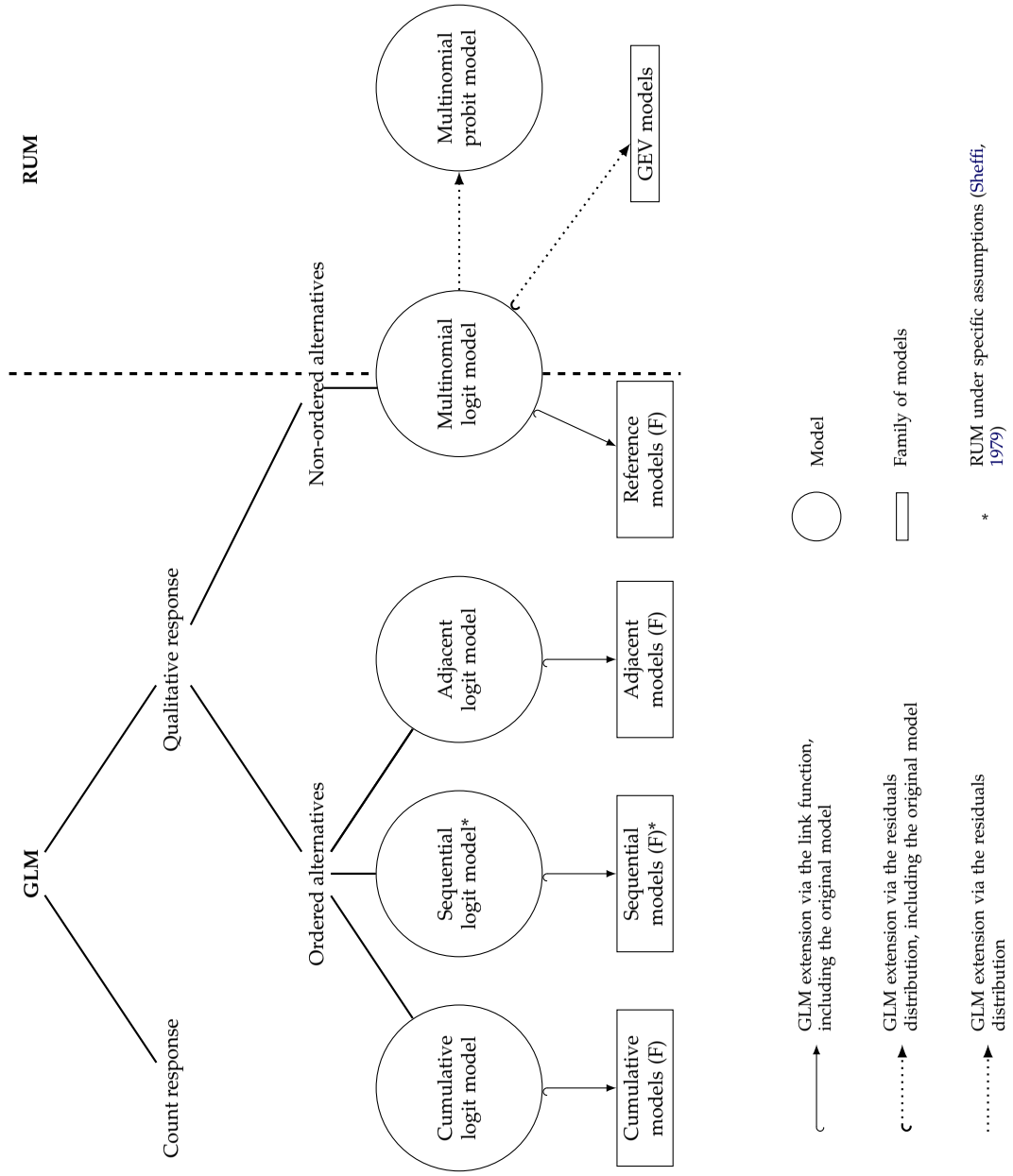
Some DCMs are derived from the Random Utility Maximization (RUM) principle such as the family of Generalized Extreme Value (GEV) models introduced by McFadden (1978), which includes the multinomial logit model. However, the family of DCMs is principally composed of generalized linear models (GLMs) for discrete response variables, also including the multinomial logit model. A non-exhaustive overview of DCMs is presented in Figure 8.1. Discrete response variables can be differentiated into count and qualitative responses. On the one hand, Poisson regression models are classically used for count data (Cameron and Trivedi, 1986). On the other hand, different models are used for qualitative choices according to the ordering assumption among the alternatives. Regression models appropriate for ordered alternatives include the cumulative (McCullagh, 1980), sequential (Tutz, 1991) and adjacent (Goodman, 1983) logit model. They have been used, for instance, to explore the dynamic of credit rating matrices (Feng et al., 2008), to explain employment stability (Kahn and

Morimune, 1979) or for partial credit scoring (Masters, 1982). For these three kinds of ordered logit models some extensions have been introduced using a cumulative distribution function (cdf) other than the logistic one (Amemiya, 1981; Fullerton, 2009). The best known of them is the ordered probit model, which is an extension of the cumulative logit model, defined with the normal cdf (McKelvey and Zavoina, 1975).

In the case of non-ordered alternatives, such extensions have only recently been proposed for the multinomial logit model leading to the family of reference models (Peyhardi et al., 2015). It should be noted that the multinomial probit model is not such an extension, despite its name. It is defined as a RUM model with multivariate Gaussian residual distribution whereas the cumulative, sequential and adjacent models do not strictly satisfy the RUM principle (Small, 1987) and are viewed as GLMs. In the same way, reference models take advantage of the GLM framework. The simple Newton-Raphson or Fisher scoring algorithm can be used, implying a fast estimation procedure compared to the complex estimation of the multinomial probit model (see Peyhardi et al. (2015) for details). In the GLM framework, the link function relates the probabilities of alternatives to the linear predictors (depending on individual and alternative-specific variables). The link function of reference models is the link function of the multinomial logit model, except that the logistic cdf is replaced by other cdfs F, such as the normal, the Gompertz or the Student cdfs. Adjusting the tail-weight and asymmetry of distributions may markedly improve the model fit (Peyhardi et al., 2015) and may have behavioral and economic implications.

The multinomial logit model has been widely used to perform economic analyses (Small and Rosen, 1981), design public policies and adapt marketing strategies (Baltas and Doyle, 2001) in a variety of fields (transportation, energy, food consumption, and so on). The use of reference models for discrete choice modeling may therefore have strong implications if they provide results that are contrasted with the multinomial logit in terms of data fit, preferences analysis, value elicitation or demand elasticities. The aim of this article is to derive economic outputs (preferences structure, marginal effects, willingness to pay, and elasticities) from the reference models and to empirically compare these new qualitative choice models with the multinomial logit model. To evaluate the added value reference models offer, in terms of statistical, behavioral and economic results, the empirical application uses a real and recent Stated Preferences (SP) survey that explores the determinants of travel mode choice, including comfort, in an intercity context. Mode choice is particularly appropriate for such an application since it is a well documented decision process (see De Witte et al., 2013, for a literature review) which has already been the support for the development of DCMs.

FIGURE 8.1 – Reference models position among discrete choice models



The paper is organized as follows. Section 8.2.2 presents the recently introduced family of reference models in the context of DCM and derives economic outputs. Section 8.2.3 presents the data and the method used to compare the estimation results of the different models and discusses the results. Section 8.2.4 concludes.

8.2.2 Generalized linear models for non-ordered choices

8.2.2.1 Logit models

Let $\pi_j^{\{n\}}$ denote the probability that individual n chooses the alternative j among the alternatives $1, \ldots, J$. Let $x^{\{n\}}$ and $\omega_j^{\{n\}}$ respectively denote the vector of individual and alternative-specific variables (for the alternative j) associated with the individual n. For the sake of simplicity, the individual index n will be suppressed in the following without loss of generality. The logit model is summarized by the $J-1$ equations

$$\pi_j = \frac{\exp(\eta_j)}{1 + \sum_{k=1}^{J-1} \exp(\eta_k)}, \tag{8.17}$$

for $j = 1, \ldots, J-1$. Depending on the form of the linear predictors η_j, we obtain different logit models:

- $\eta_j^{(1)} = \alpha_j + x^T \delta_j$. Individual variables x are used with $J-1$ different slopes δ_j.
- $\eta_j^{(2)} = \alpha_j + \tilde{\omega}_j^T \gamma$ where $\tilde{\omega}_j = \omega_j - \omega_J$. Alternative-specific variables ω_j are used with common slope δ. This is the conditional logit model introduced by McFadden (1973).
- $\eta_j^{(3)} = \alpha_j + x^T \delta_j + \tilde{\omega}_j^T \gamma$. This is a combination of the two previous parameterizations leading to the widely used multinomial logit model.

8.2.2.2 Reference models

All the classical regression models for categorical responses described by Tutz (2012) and Agresti (2013) share the generic equations (Peyhardi et al., 2015):

$$r_j(\pi) = F(\eta_j) \tag{8.18}$$

for $j = 1, \ldots, J-1$, where r is a diffeomorphism from the simplex (corner of hypercube) to an open subset of the hypercube, π is the vector of probabilities $(\pi_1, \ldots, \pi_{J-1})^T$ and F is a continuous and

strictly increasing cdf. In this framework, the Equations (8.17) of the logit models are equivalent to

$$\frac{\pi_j}{\pi_j + \pi_J} = \frac{\exp(\eta_j)}{1 + \exp(\eta_j)},$$

for $j = 1, \ldots, J-1$. We recognize the logistic cdf (on the right) and the recently introduced reference ratio (on the left). The alternative J is considered as the reference alternative and thus $\pi_J = 1 - \sum_{j=1}^{J-1} \pi_j$. Moreover, the vector of predictors $\eta = (\eta_1, \ldots, \eta_{J-1})^T$ can be expressed as the product of the design matrix Z and the vector of parameters. The three previous parametrizations of linear predictors can thus be summarized by the three following design matrices:

$$Z^{(1)} = \begin{pmatrix} 1 & & & x^T & & \\ & \ddots & & & \ddots & \\ & & 1 & & & x^T \end{pmatrix},$$

$$Z^{(2)} = \begin{pmatrix} 1 & & & \tilde{\omega}_1^T \\ & \ddots & & \vdots \\ & & 1 & \tilde{\omega}_{J-1}^T \end{pmatrix},$$

$$Z^{(3)} = \begin{pmatrix} 1 & & & x^T & & & \tilde{\omega}_1^T \\ & \ddots & & & \ddots & & \vdots \\ & & 1 & & & x^T & \tilde{\omega}_{J-1}^T \end{pmatrix}.$$

Finally, the three logit models are fully specified by the (reference, logistic, $Z^{(i)}$) triplets with $i = 1, 2, 3$.

Other models can be derived from the Equation (8.18) by choosing alternative ratios $(r_j)_{j \neq J}$ and/or alternative cdfs F. So far, four ratios have been highlighted: the reference, adjacent, cumulative and sequential ratios. The last three are appropriate for ordered alternatives, since the models generated are generally not invariant under a permutation of alternatives. On the contrary, the reference models are invariant under all permutations that fix the reference alternative and are thus appropriate for non-ordered alternatives (see Peyhardi et al., 2015 for details about invariance properties). Although the reference ratio is mandatory for non-ordered choices, the logistic cdf is not. The cdf F only needs to be strictly increasing for the interpretation of parameters and differentiable for their estimation. The set of such cdfs is denoted by $\mathfrak{F}$. In application, this set will be restricted to the logistic, normal, Laplace, Cauchy, Gumbel, Gompertz and Student cdfs (i.e. all the most usual distributions of $\mathfrak{F}$) and will be denoted by $\mathfrak{F}^0$. The characteristics of these distributions are

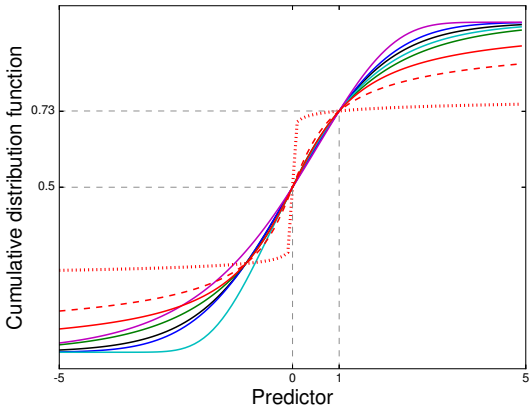
summarized in Table 8.1. Moreover, $\mathfrak{F}^{0,*}$ will denote the same set, excluding the logistic cdf. Cdfs differ in their symmetry (symmetric or asymmetric distributions) and in their tails (heavy or thin tails), each allowing a different fit of observed probabilities according to the predictor values. For example, heavy-tailed distributions, such as the Cauchy or Student cdfs, permit an abrupt transition of the response behavior around a threshold value of the predictor. The asymmetry of the Gumbel and Gompertz distributions allows an asymmetry in the response behavior around the switching point (see Figure 8.2 for a representation of cdfs $F \in \mathfrak{F}^0$). The selection of the appropriated cdf F may markedly improve the model fit. Finally, we propose to use the family of reference models specified by the triplets (reference, F, $Z^{(i)})_{F\in\mathfrak{F}^0}$ with $i = 1, 2, 3$, containing the usual logit models.

TABLE 8.1 – Characteristics of the cumulative distribution functions $F \in \mathfrak{F}^0$

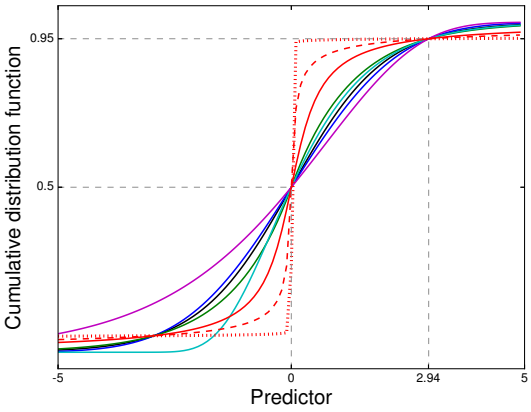
Name	$\boldsymbol{F(x)}$	**Shape parameter**	**Symmetric**	$\boldsymbol{(m,s)}$ **such that** $\boldsymbol{F_{m,s} \in \mathfrak{F}_{q_{0.95}}}$	
logistic	$\frac{\exp(x)}{1+\exp(x)}$	No	Yes	$m = 0$	$s = 1$
normal	$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{t^2}{2}\right)dt$	No	Yes	$m = 0$	$s = 1.79$
Laplace	$\begin{cases}\frac{1}{2}\exp(x) & \text{if } x < 0\\ 1-\frac{1}{2}\exp(-x) & \text{if } x \geq 0\end{cases}$	No	Yes	$m = 0$	$s = 1.279$
Cauchy	$\frac{1}{2}+\frac{1}{\pi}\arctan(x)$	No	Yes	$m = 0$	$s = 0.466$
Gumbel	$\exp\{-\exp(-x)\}$	No	No	$m = -0.414$	$s = 1.131$
Gompertz	$1-\exp\{-\exp(x)\}$	No	No	$m = 0.737$	$s = 2.012$
Student	$\frac{1}{2}+x\Gamma\left(\frac{1+\nu}{2}\right)\frac{{}_2F_1\left(\frac{1}{2},\frac{1+\nu}{2},\frac{3}{2},-\frac{x^2}{\nu}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)}$	ν (df)	Yes	$m = 0$	$s = 0.0123$ ($\nu = 0.35$)

8.2.2.3 Normalization of parameter estimates via the cdf F

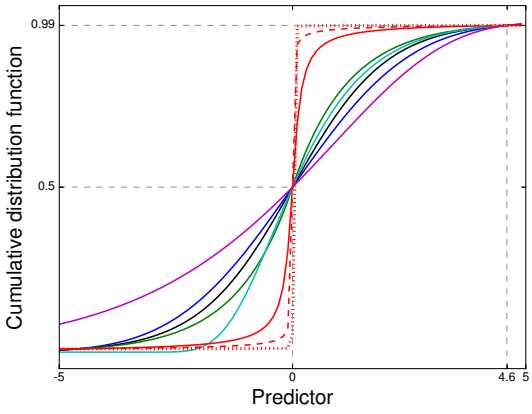
As noted by Tutz (1991), "distribution functions generate the same model if they are connected by a linear transformation". If the connection is made through a location parameter m and a scale parameter



(a) $F \in \mathfrak{F}_{q_{0.73}}$



(b) $F \in \mathfrak{F}_{q_{0.95}}$



(c) $F \in \mathfrak{F}_{q_{0.99}}$

FIGURE 8.2 – Normalized cdf $F \in \mathfrak{F}_{q_p}$ with scales (A) $q_{0.73} = 1$, (B) $q_{0.95} = 2.94$ and (C) $q_{0.99} = 4.6$ using the logistic (black), normal (blue), Laplace (green), Gumbel (cyan), Gompertz (magenta), Cauchy (red), Student$_{\nu=0.5}$ (dashed red line) and Student$_{\nu=0.05}$ (dotted red line) cdfs.

s such that $F_{m,s}(z) = F\{(z-m)/s\}$, we have

$$F_{m,s}(\eta_j') = F\left(\frac{\eta_j' - m}{s}\right) = F\left(\frac{\alpha_j' - m}{s} + x^T\frac{\delta_j'}{s} + (\omega_j - \omega_J)^T\frac{\gamma'}{s}\right),$$

for $j = 1, \ldots, J-1$. An equivalent model is obtained using the reparametrization $\alpha_j' = s\alpha_j + m$, $\delta_j' = s\delta_j$ and $\gamma' = s\gamma$. Therefore, for each distribution, only one representative element F_{m_0,s_0} has to be chosen among the class of cdfs $(F_{m,s})_{m,s>0}$. In the context of binary regression, Tutz (2012) proposed the transformation of F so that the mean and the variance according to different distributions are the same, using the logistic cdf as the reference (mean and variance of the logistic cdf are 0 and $\pi^2/3$). However, this method cannot be used with the Cauchy cdf since the mean and variance do not exist. More generally, the mean and variance of the Student distribution do not exist when $\nu < 1$ and $\nu < 2$ respectively. Therefore, we propose to normalize the different reference models via the location and scale parameters of the cdf F. To do this, two real points a and b are chosen such that all cdfs in $\mathfrak{F}$ have the same values $F(a)$ and $F(b)$. The logistic cdf is naturally proposed as the reference cdf since it corresponds to the canonical regression model.

The first condition $F(0) = 1/2$, which states that the median is null, is necessary to keep the alternative-specific constant α_j interpretable. Assume that variables x and ω have the same effect on alternative j with respect to the reference alternative J, i.e. $\delta_j = 0$ and $\omega_j = \omega_J$. Therefore, the ratio of probabilities only depends on the alternative-specific constant $\pi_j/\pi_J = F(\alpha_j)/\{1 - F(\alpha_j)\}$ and the null hypothesis $H_0 : \alpha_j = 0$ is thus equivalent to the equality between probabilities π_j and π_J. Testing the significance of the alternative-specific constant α_j is therefore equivalent to testing the significance of the difference between the probabilities π_j and π_J. Note that the condition $F(0) = 1/2$ is already satisfied for symmetric distributions (e.g. logistic, normal, Laplace, Cauchy) and has to be imposed for non-symmetric distributions (e.g. Gumbel and Gompertz).

For the second condition, $F(b) = e^b/(1+e^b)$, the choice of b is more debatable. Using $b = 1$ seems natural but other values could be used to calibrate the different cdfs F, such as the quantile of the logistic distribution $b = q_p$ for some probability $p \neq 1/2$. For each probability $p \neq 1/2$, we obtain the normalized space $\mathfrak{F}_{q_p} = \{F \in \mathfrak{F} : F(0) = 1/2,\ F(q_p) = p\}$. The logistic, normal, Laplace, Cauchy, Gumbel and Gompertz cdfs $F \in \mathfrak{F}_{q_p}$ are represented in Figure 8.2 for the three cases $q_{0.73} = 1$, $q_{0.95} \simeq 2.94$ and

$q_{0.99} \simeq 4.6$. In terms of location and scale parameters, we have $F_{m_0,s_0} \in \mathfrak{F}_{q_p}$ if

$$\left\{ \begin{array}{rcl} m_0 & = & -\dfrac{F^{-1}(1/2) \cdot q_p}{F^{-1}(p) - F^{-1}(1/2)}, \\ s_0 & = & \dfrac{q_p}{F^{-1}(p) - F^{-1}(1/2)}. \end{array} \right. \tag{8.19}$$

After transformation, the new parameters are $\alpha'_j = (\alpha_j - m_0)/s_0$, $\delta'_j = \delta_j/s_0$ for $j = 1, \ldots, J-1$ and $\gamma' = \gamma/s_0$. The location m_0 implies a translation only of alternative-specific parameters whereas the scale s_0 impacts all the parameters. It should therefore be borne in mind that, even if the parameters are more interpretable after normalization, the best way to compare results obtained with different models is to compare ratios of slopes which stay the same for all normalizations (e.g., $\gamma'_{\text{time}}/\gamma'_{\text{cost}} = \gamma_{\text{time}}/\gamma_{\text{cost}}$). Moreover, if the distribution F is symmetric, then $m_0 = 0$ for all normalizations $\left(\mathfrak{F}_{q_p}\right)_{p \neq 1/2}$ and thus the ratios of alternative-specific constants also stay the same (i.e. $\alpha'_j/\alpha'_k = \alpha_j/\alpha_k$ for different alternatives j and k).

8.2.2.4 The case of the Student distribution

We have seen that a change in location and/or scale parameters does not impact the loglikelihood of a reference model. Therefore, only one representative element F_{m_0,s_0} is chosen for each distribution: the logistic, normal, Laplace, Cauchy, Gumbel and Gompertz distributions. It is quite different for the Student distribution because of its additional parameter: the degree of freedom (df). Indeed, Student distributions with different degrees of freedom are not connected by a linear transformation and thus lead to different loglikelihood maxima.

Since the df has to be estimated, we are interested in the concavity of the loglikelihood according to the parameter $\nu \in \mathbb{R}_+^*$. This concavity is not theoretically demonstrated but is empirically verified. The proposed method is to estimate the (reference, Student$_\nu$, Z) models among a grid of df values and to choose the value $\nu^\star$ that maximizes the loglikelihood. We propose to use the following grid: $\nu = 0.05 \times i$ for $i = 1, \ldots, 40$ (small value between 0 and 2) and $\nu = 3, \ldots, 20$. The results clearly favor the assumption of loglikelihood concavity (see Figure 8.3). Knowing that the Student distribution converges toward the Normal distribution when $\nu \to \infty$, we note that the maximum loglikelihood value obtained with the (reference, Student$_\nu$, $Z^{(i)}$) model converges toward those obtained with the (reference, Normal, $Z^{(i)}$) model when $\nu \to \infty$. It could also be noted that the closer $\nu^\star$ is to 0, the higher the gain in loglikelihood obtained with the Student$_{\nu^\star}$ distribution in comparison to the normal and logistic distributions; see Figure 8.3b and 8.3c to compare the case $\nu^\star = 0.35$ and $\nu^\star = 1.9$. Normal

and logistic distribution are unable to capture the abrupt change in probability obtained thanks to the heavy tails of the Student$_\nu$ distribution for small values of ν.

It should be recalled that the Cauchy and Student with $\nu = 1$ are the same distribution. Therefore, the (reference, Student$_{\nu^\star}$, Z) model will always obtain a higher loglikelihood than the (reference, Cauchy, Z) model. However, this latter model will be used because it is more parsimonious.

8.2.2.5 Independence of irrelevant alternatives

It should be borne in mind that, unlike the logit models, the (reference, F, $Z)_{F\in\mathfrak{F}^{0,*}}$ are not invariant under the transposition of the reference alternative. Thus, the choice of the reference alternative will impact the parameter estimation and also the different economic outputs, such as the elasticities for instance. This dependence on the reference alternative allows us to partially relax the IIA property. The IIA property holds for the $J-1$ ratios of probabilities π_j/π_J since

$$\frac{\pi_j}{\pi_J} = \frac{F}{1-F}(\eta_j)$$

for all $j \neq J$. However, for two non-reference alternatives $j \neq J$ and $k \neq J$, we have

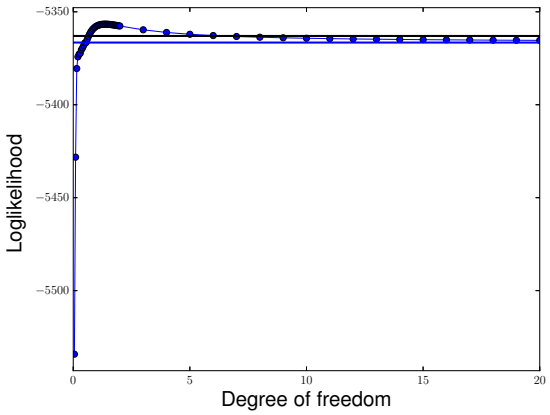
$$\frac{\pi_j}{\pi_k} = \frac{F}{1-F}(\eta_j)\frac{1-F}{F}(\eta_k).$$

With designs $Z^{(2)}$ or $Z^{(3)}$, the predictors η_j and η_k are dependent on alternative-specific variables ω_J of the reference alternative. This dependence disappears only if F is the logistic cdf noting that $F/(1-F)$ is the exponential function and thanks to the exponential property $\exp(\eta_j)/\exp(\eta_k) = \exp(\eta_j - \eta_k)$. If F is not the logistic cdf, then the IIA property holds for the $J-1$ ratios of probabilities π_j/π_J and does not hold for the $\binom{J-1}{2}$ ratios of probabilities π_j/π_k.

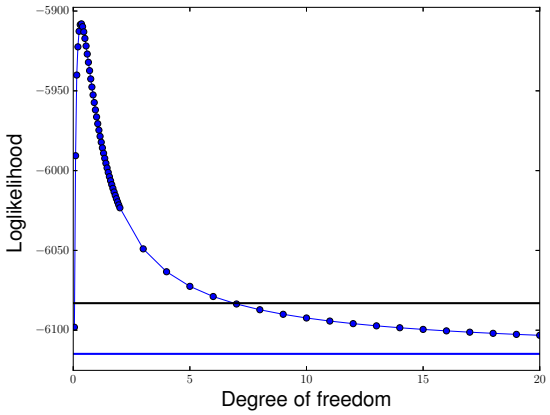
8.2.2.6 Economic outputs of reference models

Reference models can be used to discuss behavioral hypotheses and elicit economic outputs. To set the theoretical framework, we first discuss the interpretation of the parameter estimates before computing willingness to pay, marginal effects and elasticities formulas. The parameter estimates of reference models can be easily interpreted since

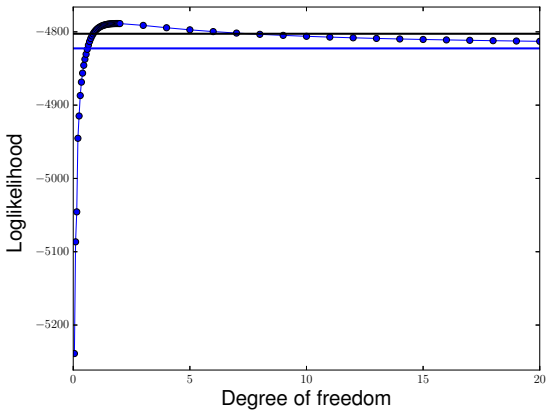
$$\frac{\pi_j}{\pi_J} = \frac{F}{1-F}(\eta_j)$$



(a) (reference, Student$_\nu$, $Z^{(1)}$) models



(b) (reference, Student$_\nu$, $Z^{(2)}$) models



(c) (reference, Student$_\nu$, $Z^{(3)}$) models

FIGURE 8.3 – Loglikelihood of (reference, Student$_\nu$, $Z^{(i)}$) models for $\nu \in (0, 20]$ (blue points), (reference, normal, $Z^{(i)}$) model (blue horizontal line) and (reference, logistic, $Z^{(i)}$) model (dark horizontal line) with (A) $i = 1$, (B) i=2 and (C) $i = 3$.

is strictly increasing with respect to η_j when $F \in \mathfrak{F}$. For instance, if the slope parameter of the alternatives cost is significantly negative, then the probability of choosing alternative j over alternative J is decreasing when its cost is increasing.

For a RUM model, the Willingness To Pay (WTP) for a continuous attribute ω (e.g. the time) is the change in cost that keeps utilities (deterministic part) unchanged after increasing the attribute value ω by one unit. It can be shown that no change in utilities implies no change in choice probabilities. The WTP can therefore be defined as the change in cost that keeps the choice probabilities unchanged after increasing ω by one unit. This is the value Δc_{j_0} such that $\pi'_j = \pi_j$ for all alternatives $j = 1, \ldots, J$ after the transformation $\omega'_{j_0} = \omega_{j_0} + 1$ for the specific alternative j_0 (the cost is transformed as $c'_{j_0} = c_{j_0} + \Delta c_{j_0}$). As the link function of a reference model is a surjective map from the simplex to $\mathbb{R}^{J-1}$, for any cdf $F \in \mathfrak{F}$, the equalities between the J probabilities $\pi'_j = \pi_j$ are equivalent to the equalities between the $J-1$ predictors $\eta'_j = \eta_j$. If $j_0 \neq J$, only the predictor η'_{j_0} is impacted by the change $\omega'_{j_0} = t_{j_0} + 1$ and we have (the linear predictor contains the interaction between time and the binary variable comfort)

$$\begin{aligned}
\eta'_{j_0} &= \eta_{j_0} \\
\gamma^{\omega} t'_{j_0} + \gamma^{\text{cost}} c'_{j_0} &= \gamma^{\omega} t_{j_0} + \gamma^{\text{cost}} c_{j_0} \\
\gamma^{\omega} + \gamma^{\text{cost}} \Delta c_{j_0} &= 0
\end{aligned}$$

If $j_0 = J$ all the $J-1$ predictors are identically impacted by the change ω'_{j_0} and we obtain the same result. Finally, the WTP is given by

$$\text{WTP} = -\frac{\gamma^{\omega}}{\gamma^{\text{cost}}}.$$

The WTP formula obtained for reference models is the same as for the multinomial logit model. Details about formulas for Value Of Time (VOT) and Time Equivalent Comfort (TEC) are given in Appendix I, for the general case of interaction between time and comfort.

Marginal effects measure the extent to which probabilities change in response to a change in the variable value. Marginal effects for reference models are described in the following Property and the details about their computation are given in Appendix I.

Property 1 *The marginal effect on the probability of choosing the alternative i in the continuous alternative-specific variable ω_j (viewed as a scalar, i.e. only the time value for instance) of alternative j is defined*

by $\text{me}_i(\omega_j) = \partial\pi_i/\partial\omega_j$*. For a reference model, we have*

$$\text{me}_i(\omega_j) = \gamma\pi_i \begin{cases} d_j(1-\pi_j) & , \ i=j, \ j\neq J, \\ -d_j\pi_j & , \ i\neq j, \ j\neq J, \\ \sum\limits_{k\neq J} d_k\pi_k & , \ i=j, \ j=J, \\ \sum\limits_{k\neq J} d_k\pi_k - d_i & , \ i\neq J, \ j=J, \end{cases}$$

where $d_i = f(\eta_i)/[F(\eta_i)\{1-F(\eta_i)\}]$ *for* $i = 1,\ldots,J-1$ *and* $f = F'$ *is the associated probability density function.*

Elasticities are an alternative measure of the response of probabilities to a change in the variable value. Their advantage is that they are normalized for the variables' units and are easily computed since

$$\text{e}_i(\omega_j) = \frac{\partial \ln \pi_i}{\partial \ln \omega_j} = \frac{\omega_j}{\pi_i}\text{me}_i(\omega_j), \tag{8.20}$$

for all $i,j \in \{1,\ldots,J\}$. The elasticities are thus given by

$$\text{e}_i(\omega_j) = \gamma\omega_j \begin{cases} d_j(1-\pi_j) & , \ i=j, \ j\neq J, \\ -d_j\pi_j & , \ i\neq j, \ j\neq J, \\ \sum\limits_{k\neq J} d_k\pi_k & , \ i=j, \ j=J, \\ \sum\limits_{k\neq J} d_k\pi_k - d_i & , \ i\neq J, \ j=J. \end{cases}$$

If $i=j$ then $\text{e}_i(\omega_j)$ is an own-elasticity, whereas if $i\neq j$ then $\text{e}_i(\omega_j)$ is a cross-elasticity. It can be seen that two cross-elasticities $e_i(\omega_j)$ and $e_{i'}(\omega_j)$ are equal when $j\neq J$ but are different when $j=J$. This is a consequence of the IIA property that does not hold for all pairs of alternatives. The change of value in Ω_J does not impact the probability of choosing other alternatives in the same way.

The special case of the logistic cdf $F(\eta) = e^\eta/(1+e^\eta)$ is presented here. It should be noted first that $f = F(1-F)$ in this case and thus $d_j = 1$ for all $j\neq J$. For alternative-specific variables, we obtain

$$\text{me}_i(\omega_j) = \gamma\pi_i\left[\mathbf{1}_{(j\neq J)}\left\{\mathbf{1}_{(i=j)} - \pi_j\right\} - \mathbf{1}_{(j=J)}\left\{\mathbf{1}_{(i\neq j)} - 1 + \pi_J\right\}\right],$$

and thus $\text{me}_i(\omega_j) = \gamma\pi_i\left\{\mathbf{1}_{(i=j)} - \pi_j\right\}$ which is the usual result obtained for the conditional and multinomial logit models (see Greene, 2003, p. 723). The elasticities are given by $\text{e}_i(\omega_j) = \gamma\omega_j\left\{\mathbf{1}_{(i=j)} - \pi_j\right\}$. The equality between cross elasticities $\text{e}_i(\omega_j)$ and $\text{e}_{i'}(\omega_j)$ is here true for all

$j = 1, \ldots, J$, as expected.

Recall that marginal effects and elasticities computation have been presented only for one individual. The aggregated elasticities are computed as follow

$$\mathrm{e}_i(\omega_j) = \frac{1}{N} \sum_{n=1}^{N} \mathrm{e}_i^{\{n\}}(\omega_j),$$

where n denotes the individual index and N the number of individuals. Owing to the IIA property, cross-elasticities have little behavioral value. To get around this issue, Louviere et al. (2000) discussed the method of sample enumeration to compute the weighted elasticities

$$\mathrm{we}_i(\omega_j) = \frac{1}{\sum_{n=1}^{N} \pi_i^{\{n\}}} \sum_{n=1}^{N} \pi_i^{\{n\}} \mathrm{e}_i^{\{n\}}(\omega_j).$$

Note that the estimation of VOT, TEC, probabilities, marginal effects and elasticities is invariant under the normalization of parameters (see Appendix I for details).

8.2.3 Application

The source code (written in *C++* and *Python*) used for the inference of reference model parameters and economic outputs is available on GitHub (*https://github.com/StatisKit*) and binaries can be installed using the Conda package management system (*http://conda.pydata.org*). Refer to the documentation for more information (*http://statiskit.rtfd.io*). For the specific case of (reference, logistic, $\mathrm{Z}^{(i)}$) models, estimation results have been compared and validated with the one provided by the package "mlogit" in R.

8.2.3.1 Data

Between 2012 and 2015, a large travel survey was conducted among 37,479 inhabitants living in the Rhône-Alpes region (France) with geographic stratified sampling. Those who had recently traveled by train were also asked to answer a web SP survey. Only respondents living in the Rhône-Alpes region, aged 18 or over, having a car and a driving license and whose trip was made or could have been made by train or coach were asked to answer the choice questions, resulting in 1,215 answers. Due to the low rate of train users in the population, they were oversampled with a face-to-face survey made in regional trains, using the quota sampling method (sex, age, motive, travel time and train line). 608 answers were collected. After cleaning the data, 1,774 answers were retained.

In order to personalize the choice questions, the respondents were first asked to describe in detail (time, cost, purpose, origin and destination) a trip they had made by coach, train or car during the last month within the Rhône-Alpes region. Then, they had to choose between three travel modes: train, coach and car, one being a status quo alternative. All alternatives were described in terms of travel mode, cost, time, probability and time delay, frequency, clock-face timetable and comfort. To avoid a cognitive burden, attributes describing the journey were split into three exercises. In Exercise 3, on which we focus here, modes vary according to travel time, travel cost and comfort. Travel time is defined from origin to destination (including access time, egress time, waiting time and in-vehicle time). Travel cost includes public transport ticket or pass, gasoline, parking cost and toll. Comfort is defined as the guarantee of seating availability. It is therefore always equal to one for coach and car alternatives but sometimes equal to zero for the train alternative, if a seat is not guaranteed. In this case, the train user may have to stand during all or part of the journey. A pivot Bayesian efficient design was implemented (Rose and Bliemer, 2007). A priori weights of attributes were taken from the literature and adjusted during the pilot tests.

Figure 8.4 gives an example of a choice question. The percentage of respondents who chose the train alternative was 29.6%, while 21% chose the coach alternative and 49.4% chose the car alternative. Respondents had to answer to four choices questions in exercise 3, leading to a database with 6,373 observations since a few respondents did not answer all four questions.

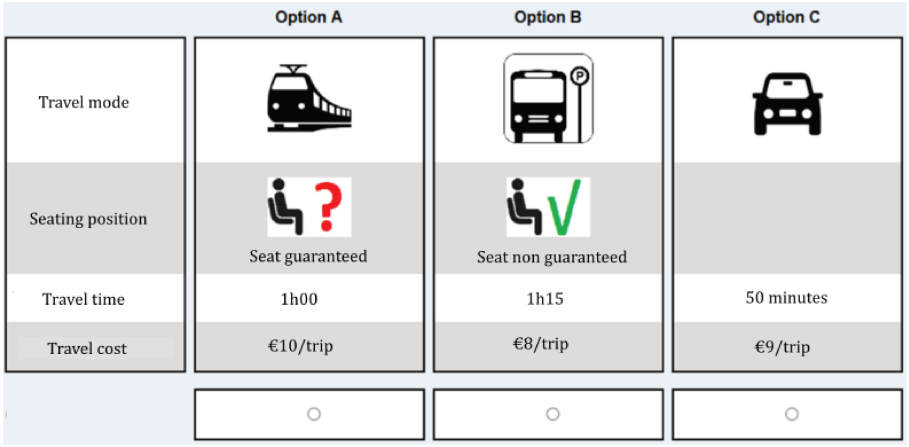


FIGURE 8.4 – Example of choice question for Exercise 3

The alternative-specific variables $(\omega_j)_{j=1,2,3}$ are the travel time, travel cost and comfort as well as the cross-variable time $\times$ comfort, which is equal to travel time when a seat is guaranteed and equal to zero otherwise. The individual variables x are related to socio-economic characteristics (age and income), spatial characteristics (% of the origin and destination municipalities in a high density area),

journey characteristics (train, coach or car user for the reference trip, frequency of the trip, and so on) and general mobility indicators (access to a car and frequency of use of alternative modes to the car). Table 8.2 displays descriptive statistics for all variables used in the models.

TABLE 8.2 – Descriptive statistics

Variable definition	**Label**	**Mean**	**S.D.**	**Min**	**Max**
Alternative-specific variables					
Travel cost by train (in euros)	Train cost	8.88	7.56	1	62
Travel cost by coach (in euros)	Coach cost	8.90	7.80	1	78
Travel cost by car (in euros)	Car cost	9 .99	8.64	1	62
Travel time by train (in minutes)	Train time	69.89	51.86	7	325
Travel time by coach (in minutes)	Coach time	70.22	53 .23	7	325
Travel time by car (in minutes)	Car time	57.24	36.89	4	330
Comfort in train (1 if a seat is guaranteed, 0 otherwise)	Comfort	0.50			
Individual variables					
Age (in years)	Age	47.57	15.76	19	90
Number of cars available per person in the household	Access_car	0.76	0.38	0.12	4.00
% of the origin municipality in high density area	Orig_dens	74.21	39.58	0	100
% of the destination municipality in high density area	Dest_dens	86.15	29.93	0	100
Monthly income above 4,000 euros (1 if yes, 0 otherwise)	Income_h	0.28			
Coach user for the reference journey (1 if yes, 0 otherwise)	Type_coach	0.03			
Car user for the reference journey (1 if yes, 0 otherwise)	Type_car	0.54			
Makes the reference trip on a regular basis (1 if yes, 0 otherwise)	Regular	0.28			
Imperative schedule at destination (1 if yes, 0 otherwise)	Imperative	0.45			
Car user who had already used public transport to make the reference trip (1 if yes, 0 otherwise)	Alt_pt	0.15			
Train or coach user who had already used a car to make the reference trip (1 if yes, 0 otherwise)	Alt_car	0.26			
Frequent use of modes other than car (1 if yes, 0 otherwise)	Freq_alt	0.41			

8.2.3.2 Method

To apply reference models, the first step is to choose the reference alternative. Train is chosen as the reference alternative because it is the second most frequently chosen mode for the reference journey (43%) and because it is the first alternative displayed in the choice questions and may thus be used as a reference by a majority of respondents.

Reference models are estimated for the three designs $Z^{(1)}$, $Z^{(2)}$ and $Z^{(3)}$. The individual variables were chosen using stepwise selection and doing numerous tests to choose their specification. Using the downgraded designs $Z^{(1)}$ or $Z^{(2)}$ instead of the full design $Z^{(3)}$ implies higher unobserved heterogeneity due to omitted variables. Such models with omitted variables, even if exaggerated for emphasis, allow us to identify the cdfs that are less sensitive to the omission of variables and that retrieve the most consistent economic outputs.

For each design $Z^{(i)}$, the degree of freedom $\nu^\star$ is first estimated for the Student$_\nu$ distribution and then the six different models (using logistic, normal, Laplace, Cauchy, Gumbel and Student$_{\nu^\star}$ cdfs [3] are estimated. For each cdf, the normalized space $\mathfrak{F}_{q_{0.95}}$ has been retained. Indeed, in comparison to the other normalized spaces, it improves the readability of results, the parameter's estimates being less dispersed; see Table 8.1 for the corresponding values of m_0 and s_0 satisfying Equations (8.19).

For each design $Z^{(i)}$, the six (reference, F, $Z^{(i)}$)$_{F\in\mathfrak{F}^0}$ models are compared in terms of goodness-of-fit indicators: loglikelihood, $\bar{\rho}^2$ as defined in Ben-Akiva and Swait (1986), AIC and BIC, regarding absolute values as well as ranking. The AIC is usually preferred to select a model for prediction whereas the BIC is preferred for explanation purposes. Both are retained since both objectives are pursued and they offer similar results. Indeed, all cdfs have the same number of parameters except the Student distribution, which requires the estimation of only one additional parameter, the degrees of freedom $\nu^\star$. To test for differences between the non-nested models for each design (therefore with the same structure of explanatory variables), we follow the method proposed by Ben-Akiva and Swait (1986), which determines the asymptotic upper bound of the probability that the difference between the $\bar{\rho}^2$ is greater than the one observed.

Provided that they are significantly different from one another, the reference models are then compared in terms of outcomes (parameter estimates and economic outputs). For each outcome, we first check the consistency of the results obtained with design $Z^{(3)}$, and in particular the best ranking model. We then discuss heterogeneity between models, first with design $Z^{(3)}$ and then with designs $Z^{(1)}$ and/or $Z^{(2)}$, and highlight, when relevant, the range of values, the key features and the difference between the best ranking model and the (reference, logistic, $Z^{(i)}$) model. To facilitate comparisons, each triplet (reference, $F^{[r]}$, Z) indicates the rank r of the model according to the ρ^2 value.

To gain further insight into the differences between cdfs, VOT and TEC are also simulated [4] using the Krinsky and Robb (1986) method.

8.2.3.3 Results

8.2.3.3.1 Goodness-of-fit indicators The degrees of freedom that maximize the likelihood obtained with (reference, Student$_\nu$, $Z^{(i)}$)$_{i=1,2,3}$ models are respectively $\nu^\star = 1.35$, $\nu^\star = 0.35$ and $\nu^\star = 1.9$ (see Figure 8.3).

With design $Z^{(1)}$, the three best models are the (reference, $F^{[1]}$ = Laplace, $Z^{(1)}$), (reference, $F^{[2]}$ =

3. The (reference, Gompertz, $Z^{(i)}$) models have also been estimated, however results are not presented here because of convergence problems (see Peyhardi, 2013, Chapter 5).
4. 10,000 draws of a multivariate normal distribution.

Student$_{1.35}$, $Z^{(1)}$) and (reference, $F^{[3]}$ = Cauchy, $Z^{(1)}$) models (see Table 8.3). According to the Ben-Akiva and Swait test, these three models are not significantly different from one another at the 5% level (see Table 8.4). However, they outperform the other three models and the (reference, $F^{[5]}$ = Gumbel, $Z^{(1)}$) is not significantly different from the (reference, $F^{[6]}$ = normal, $Z^{(1)}$) model. With design $Z^{(2)}$, the (reference, $F^{[1]}$ = Student$_{0.35}$, $Z^{(2)}$) model is the best and (reference, $F^{[2]}$ = Cauchy, $Z^{(2)}$) model follows. All models are significantly different from one another at the 5% level. With design $Z^{(3)}$, the (reference, $F^{[1]}$ = Gumbel, $Z^{(3)}$) model is the best and (reference, $F^{[2]}$ = Student$_{1.9}$, $Z^{(3)}$) and (reference, $F^{[3]}$ = Laplace, $Z^{(3)}$) models follow. As with design $Z^{(2)}$, all models are significantly different from one another at the 5% level.

TABLE 8.3 – Goodness-of-fit indicators of (reference, F, $Z^{(i)}$) models for $F \in \mathfrak{F}^0$ with $i = 1, 2, 3$

(reference, F, $Z^{(1)}$)		**logistic**	**normal**	**Laplace**	**Cauchy**	**Gumbel**	**Student$_{1.35}$**
Indicator value	Loglikelihood	-5,363	-5,367	-5,356	-5,358	-5,366	-5,357
	$\bar{\rho}^2$	0.230	0.230	0.231	0.231	0.230	0.231
	AIC	10,778	10,785	10,765	10,767	10,784	10,767
	BIC	10,954	10,961	10,941	10,943	10,959	10,950
Rank	Loglikelihood	4	6	1	3	5	2
	$\bar{\rho}^2$	4	6	1	3	5	2
	AIC	4	6	1	3	5	2
	BIC	4	6	1	2	5	3
(reference, F, $Z^{(2)}$)		**logistic**	**lormal**	**Laplace**	**Cauchy**	**Gumbel**	**Student$_{0.35}$**
Indicator value	Loglikelihood	-6,083	-6,115	-6,028	-5,966	-6,073	-5,908
	$\bar{\rho}^2$	0.130	0.126	0.138	0.147	0.132	0.155
	AIC	12,178	12,242	12,068	11,945	12,157	11,830
	BIC	12,219	12,282	12,109	11,985	12,198	11,877
Rank	Loglikelihood	5	6	3	2	4	1
	$\bar{\rho}^2$	5	6	3	2	4	1
	AIC	5	6	3	2	4	1
	BIC	5	6	3	2	4	1
(reference, F, $Z^{(3)}$)		**logistic**	**normal**	**Laplace**	**Cauchy**	**Gumbel**	**Student$_{1.9}$**
Indicator value	Loglikelihood	-4,803	-4,823	-4,791	-4,797	-4,782	-4,789
	$\bar{\rho}^2$	0.310	0.310	0.311	0.311	0.313	0.312
	AIC	9,665	9,705	9,642	9,654	9,624	9,639
	BIC	9,868	9,908	9,845	9,856	9,826	9,849
Rank	Loglikelihood	5	6	3	4	1	2
	$\bar{\rho}^2$	5	6	3	4	1	2
	AIC	5	6	3	4	1	2
	BIC	5	6	2	4	1	3

Overall, the logistic and normal cdfs perform badly whereas the Student$_{\nu^\star}$ cdf performs well (it is ranked either first or second). An interpretation is that thinner tails are needed when data are more informative about behavior.

TABLE 8.4 – Upper bound of the p-values for the Ben-Akiva and Swait test applied to the (reference, F, $Z^{(i)}$) models for $F \in \mathfrak{F}^0$ with $i = 1, 2, 3$

(reference, F,$Z^{(1)}$)	**Laplace**	**Student$_{1.35}$**	**Cauchy**	**logistic**	**Gumbel**	**normal**
Laplace	0.500					
Student$_{1.35}$	0.111	0.500				
Cauchy	0.056	0.155	0.500			
Logistic	<0.0001	<0.0001	0.001	0.500		
Gumbel	<0.0001	<0.0001	<0.0001	0.009	0.500	
Normal	<0.0001	<0.0001	<0.0001	0.004	0.115	0.500
(reference, F,$Z^{(2)}$)	**Student$_{0.35}$**	**Cauchy**	**Laplace**	**Gumbel**	**logistic**	**normal**
Student$_{0.35}$	0.500					
Cauchy	<0.0001	0.500				
Laplace	<0.0001	<0.0001	0.500			
Gumbel	<0.0001	<0.0001	<0.0001	0.500		
Logistic	<0.0001	<0.0001	<0.0001	<0.0001	0.500	
Normal	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.500
(reference, F,$Z^{(2)}$)	**Gumbel**	**Student$_{1.9}$**	**Laplace**	**Cauchy**	**logistic**	**normal**
Gumbel	0.500					
Student$_{1.9}$	<0.0001	0.500				
Laplace	<0.0001	0.026	0.500			
Cauchy	<0.0001	<0.0001	<0.0001	0.500		
Logistic	<0.0001	<0.0001	<0.0001	<0.0001	0.500	
Normal	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.500

8.2.3.3.2 Interpretation and comparison of parameter estimates The (reference, F, $Z^{(3)}$) models provide parameter estimates consistent with the theory (see Table 8.7). For the six models, the time and cost parameters are significant and negative, as expected. The comfort parameter is significant and positive. The crossed effect of time and comfort is positive and significant, indicating the propensity to accept longer travel time in comfortable conditions. When significant, the alternative-specific constants indicate that the coach and car alternatives are associated with a smaller baseline probability than the train alternative. Earning a high income (*income_h*), high motorization (*access_car*), using coach or car for the reference journey (*type_coach* or *type_car*) or having already used a car alternative for the reference journey (*alt_car*) all increase the probability of choosing the car over the train. Conversely, age, population density at destination (*dest_dens*), making the reference trip on a regular basis (*regular*), having an imperative activity at destination (*imperative*), having already used a public transport alternative for the reference journey (*alt_pt*) and frequent use of modes other than car (*freq_alt*) all decrease the probability of choosing the car over the train. For the coach, the effects are overall the same except that there are fewer individual variables discriminating the train and coach alternatives, which is intuitive since both are public transport modes. Even if the scales of parameters are partly normalized to provide comparable estimations, the direct comparison of parameter estimates is precluded. The focus is therefore only on the structure, sign and significance of the

parameter estimates.

Firstly, following Viney et al. (2005), Flynn et al. (2008) and Kerr and Sharp (2009), the structure of estimated preferences is analyzed. For each design, the order of parameter estimates is generally respected across models, with relatively stable patterns. For instance, with design $Z^{(3)}$, the car alternative-specific constant is higher than the coach alternative-specific constant across all models. However, there are some exceptions. For the car alternative, the parameter associated with the "regular" variable is lower than the parameter associated with the "freq_alt" variable with the (reference, $F^{[4]} =$ Cauchy, $Z^{(3)}$) model. The order between these two parameters is reversed with the (reference, $F^{[5]} =$ logistic, $Z^{(3)}$) models.

Secondly, when significant, parameter estimates have the same signs across the cdfs for a given design. However, depending on the cdf, parameters differ in their significance level (see Tables 8.5, 8.6 and 8.7). The choice of the cdf therefore impacts the selection of variables. For instance, with design $Z^{(1)}$ according to the (reference, $F^{[1]} =$ Laplace, $Z^{(1)}$) model, older people tend to favor the coach over the train. Whereas, according to the (reference, $F^{[4]} =$ logistic, $Z^{(1)}$) model, age does not significantly explain the differences between the coach and train probabilities. With design $Z^{(2)}$, the cross variable Time $\times$ Comfort is positive but not significant with the two best performing cdfs ($F^{[1]} =$ Student$_{0.35}$, $F^{[2]} =$ Cauchy), negative but not significant with the two middle cdfs ($F^{[3]} =$ Laplace, $F^{[4]} =$ Gumbel), and significant and negative with the two worst performing cdfs ($F^{[5]} =$ logistic and $F^{[6]} =$ normal), which raises a particular concern. The negative sign of the cross variable may be due to not capturing the anchoring bias, since the individual reference mode is omitted. This will induce inconsistent VOT and TEC with logistic and normal cdfs.

TABLE 8.5 – Parameter estimates of (reference, $F_{m,s}$, $Z^{(1)}$) models for $F_{m,s} \in \mathfrak{F}^0_{q_{0.95}}$

			logistic	**normal**	**Laplace**	**Cauchy**	**Gumbel**	**Student**$_{1.35}$
Alternative-specific constants		Coach	-0.635 *	-0.674 *	-0.492 *	-0.303 **	-0.555 **	-0.412 **
		Car	-0.765 **	-0.791 **	-0.611 **	-0.375 **	-0.554 **	-0.525 **
Individual variables	Coach	Age	0.059	0.052	0.063 *	0.036 *	0.048 .	0.047 *
		Income_h	0.302 ***	0.322 ***	0.258 ***	0.125 ***	0.218 ***	0.178 ***
		Orig_dens	0.024	0.027	-0.005	-0.008	-0.021	0.001
		Dest_dens	-0.404 **	-0.418 **	-0.358 ***	-0.174 ***	-0.278 **	-0.243 ***
		Type_coach	2.07 ***	2.232 ***	1.736 ***	0.979 ***	1.8 ***	1.291 ***
		Type_car	1.075 ***	1.193 ***	0.771 ***	0.414 ***	0.867 ***	0.592 ***
		Regular	-0.128	-0.13	-0.111	-0.056	-0.09	-0.081 .
		Imperative	-0.072	-0.09	-0.027	0.01	-0.037	-0.005
		Alt_pt	0.021	0.009	0.038	0.05	0.127	0.053
		Alt_car	0.014	0.008	0.012	0.023	0.05	0.026
		Freq_alt	-0.129	-0.138	-0.122 .	-0.044	-0.055	-0.066
		Access_car	0.23 .	0.232 .	0.201 *	0.096 .	0.191 *	0.138 *
	Car	Age	0.082 *	0.067 .	0.09 **	0.062 **	0.069 *	0.081 **
		Income_h	0.221 **	0.207 *	0.237 **	0.165 **	0.18 **	0.204 **
		Orig_dens	0.145	0.138	0.118	0.069	0.02	0.107
		Dest_dens	-0.286 *	-0.253 *	-0.315 *	-0.219 **	-0.206 *	-0.266 *
		Type_coach	1.789 ***	1.947 ***	1.496 ***	0.842 ***	1.454 ***	1.102 ***
		Type_car	2.796 ***	2.956 ***	2.499 ***	1.605 ***	2.405 ***	1.988 ***
		Regular	-0.662 ***	-0.647 ***	-0.688 ***	-0.522 ***	-0.519 ***	-0.598 ***
		Imperative	-0.321 ***	-0.326 ***	-0.292 ***	-0.167 ***	-0.244 ***	-0.224 ***
		Alt_pt	-0.533 ***	-0.548 ***	-0.516 ***	-0.323 ***	-0.214 *	-0.394 ***
		Alt_car	0.17	0.158	0.193 *	0.142 **	0.098	0.168 *
		Freq_alt	-0.544 ***	-0.557 ***	-0.522 ***	-0.293 ***	-0.387 ***	-0.38 ***
		Access_car	0.415 ***	0.386 **	0.394 ***	0.247 ***	0.405 ***	0.329 ***

Notes: *** = significant at the 0.1 % level; ** = significant at the 1 % level; * = significant at the 5 % level; . = significant at the 10 % level

TABLE 8.6 – Parameter estimates of (reference, $F_{m,s}$, $Z^{(2)}$) models for $F_{m,s} \in \mathfrak{F}^0_{q_{0.95}}$

		logistic	**normal**	**Laplace**	**Cauchy**	**Gumbel**	**Student**$_{0.35}$
Alternative-specific constants	Coach	-0.67 ***	-0.739 ***	-0.515 ***	-0.284 ***	-0.562 ***	-0.015 ***
	Car	0.176 ***	0.185 ***	0.138 ***	0.081 ***	0.185 ***	0.004 ***
Alternative-specific variables	Time	-0.012 ***	-0.011 ***	-0.013 ***	-0.01 ***	-0.012 ***	-0.001 ***
	Cost	-0.147 ***	-0.141 ***	-0.135 ***	-0.091 ***	-0.112 ***	-0.006 ***
	Comfort	0.796 ***	0.915 ***	0.554 ***	0.283 ***	0.554 ***	0.013 ***
	Time×Comfort	-0.003 **	-0.004 ***	-0.0008	0.0001	-0.0001	0.00003

Notes: *** = significant at the 0.1 % level; ** = significant at the 1 % level; * = significant at the 5 % level; . = significant at the 10 % level

TABLE 8.7 – Parameter estimates of (reference, $F_{m,s}$, $Z^{(3)}$) models for $F_{m,s} \in \mathfrak{F}^0_{q_{0.95}}$

			logistic	**normal**	**Laplace**	**Cauchy**	**Gumbel**	**Student**$_{1.9}$
Alternative-specific constants		Coach	-0.54 *	-0.77 **	-0.325	-0.193 .	-0.538 **	-0.325 .
		Car	-1.137 ***	-1.363 ***	-0.682 **	-0.162	-0.652 ***	-0.597 **
Alternative-specific variables		Time	-0.03 ***	-0.029 ***	-0.028 ***	-0.019 ***	-0.024 ***	-0.027 ***
		Cost	-0.127 ***	-0.123 ***	-0.125 ***	-0.092 ***	-0.098 ***	-0.122 ***
		Comfort	0.46 ***	0.492 ***	0.408 ***	0.23 ***	0.294 ***	0.36 ***
		Time×Comfort	0.003 **	0.003 **	0.003 **	0.002 **	0.003 ***	0.003 **
Individual variables	Coach	Age	-0.009 **	-0.009 **	-0.008 ***	-0.004 **	-0.006 **	-0.007 ***
		Income_h	0.371 ***	0.361 ***	0.316 ***	0.178 ***	0.25 ***	0.297 ***
		Orig_dens	-0.035	0.005	-0.035	-0.031	-0.109	-0.051
		Dest_dens	-0.273 *	-0.181	-0.281 *	-0.166 **	-0.16	-0.265 *
		Type_coach	2.38 ***	2.295 ***	2.237 ***	1.637 ***	2.089 ***	2.225 ***
		Type_car	1.289 ***	1.417 ***	0.97 ***	0.495 ***	1.037 ***	0.911 ***
		Regular	-0.028	0.001	-0.054	-0.037	0.007	-0.044
		Imperative	-0.04	-0.042	-0.01	0.011	-0.021	-0.013
		Alt_pt	0.026	0.011	0.038	0.073	0.137 .	0.061
		Alt_car	-0.01	-0.035	0.006	0.013	0.051	0.01
		Freq_alt	-0.213 *	-0.228 *	-0.183 *	-0.061	-0.112	-0.134 .
		Access_car	0.138	0.184	0.072	0.006	0.112	0.06
	Car	Age	-0.018 ***	-0.017 ***	-0.02 ***	-0.016 ***	-0.014 ***	-0.019 ***
		Income_h	0.346 ***	0.31 ***	0.38 ***	0.325 ***	0.257 ***	0.377 ***
		Orig_dens	0.15	0.141	0.161	0.116	-0.053	0.146
		Dest_dens	-0.057	0.037	-0.183	-0.261 **	0.019	-0.213
		Type_coach	2.34 ***	2.43 ***	2.186 ***	1.531 ***	1.768 ***	2.084 ***
		Type_car	3.701 ***	3.813 ***	3.35 ***	2.181 ***	2.97 ***	3.135 ***
		Regular	-0.469 ***	-0.408 ***	-0.569 ***	-0.528 ***	-0.37 ***	-0.564 ***
		Imperative	-0.248 **	-0.241 **	-0.235 **	-0.162 **	-0.173 **	-0.238 **
		Alt_pt	-0.276 *	-0.263 *	-0.299 **	-0.193 *	-0.103	-0.272 *
		Alt_car	0.231 .	0.254 *	0.191 .	0.102	0.141 .	0.163
		Freq_alt	-0.599 ***	-0.605 ***	-0.586 ***	-0.33 ***	-0.424 ***	-0.494 ***
		Access_car	0.442 ***	0.443 ***	0.413 ***	0.28 ***	0.415 ***	0.402 ***

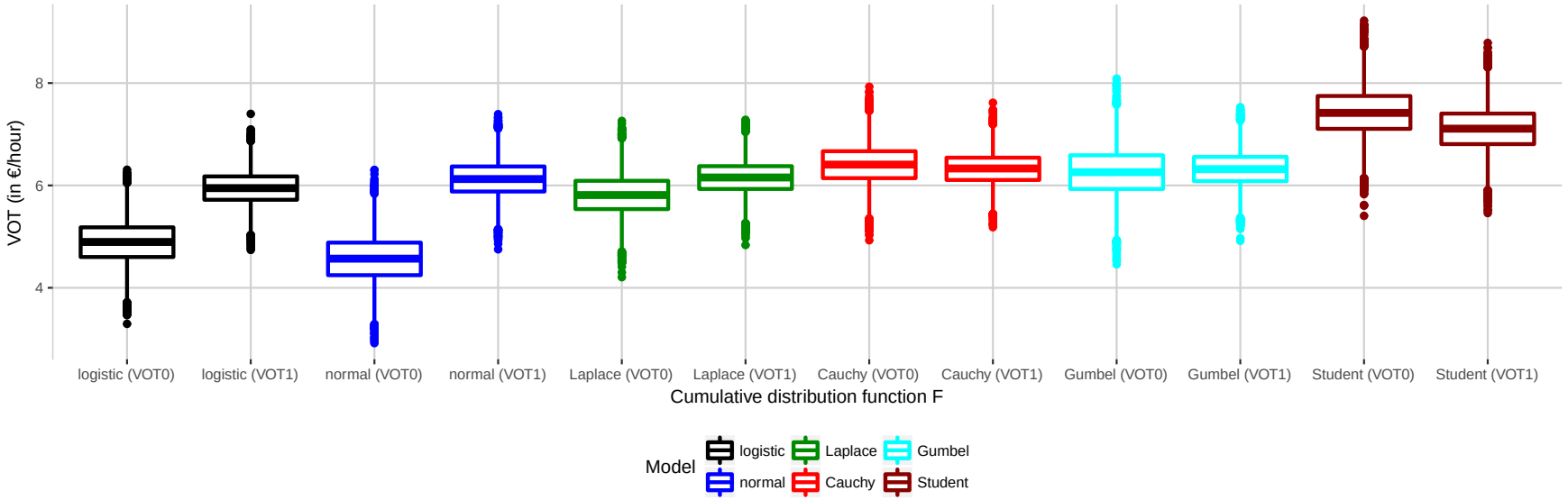
Notes: *** = significant at the 0.1 % level; ** = significant at the 1 % level; * = significant at the 5 % level; . = significant at the 10 % level

8.2.3.3.3 Value of time Because of the introduction of the cross-variable (Time×Comfort), VOT depends on the guarantee of having a seat (see Equation (I.1) in Appendix I). For the (reference, F, $Z^{(3)}$) models, VOT is consistently lower when a seat is guaranteed, since travel time can be used to rest or to work (see Table 8.8 and Figure 8.5).

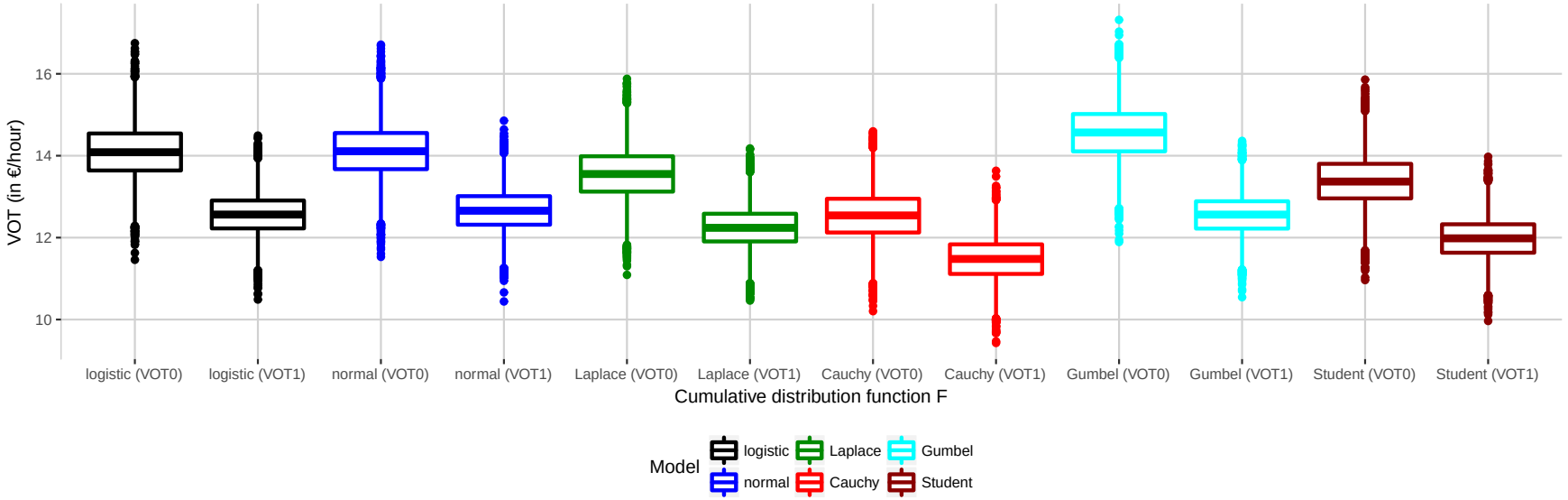
TABLE 8.8 – Value of time (VOT; in €/hour) and time equivalent of comfort (TEC; in minutes) for (reference, F, $Z^{(i)}$) models for $F \in \mathfrak{F}^0$ with $i = 2, 3$.

(reference, F, $Z^{(2)}$)	**logistic**	**normal**	**Laplace**	**Cauchy**	**Gumbel**	**Student$_{0.35}$**
VOT (comfort = 0)	4.89	4.56	5.82	6.4	6.26	7.42
VOT (comfort = 1)	5.95	6.13	6.16	6.32	6.32	7.10
TEC (time = 30 min)	60.14	75.25	40.51	29.45	47.15	19.03
TEC (time = 60 min)	53.64	64.93	38.76	29.82	46.85	20.32
TEC (time = 90 min)	47.13	54.61	37	30.19	46.54	21.60
(reference, F, $Z^{(3)}$)	**Logistic**	**Normal**	**Laplace**	**Cauchy**	**Gumbel**	**Student$_{1.9}$**
VOT (comfort = 0)	14.1	14.12	13.56	12.53	14.57	13.39
VOT (comfort = 1)	12.57	12.66	12.24	11.47	12.55	11.98
TEC(time = 30 min)	18.63	20.07	17.37	14.51	16.45	16.37
TEC (time = 60 min)	21.87	23.17	20.28	17.05	20.6	19.53
TEC (time = 90 min)	25.12	26.27	23.18	19.58	24.74	22.68

Depending on the cdf, VOT is comprised between €11.5/hour and €12.7/hour when a seat is guaranteed whereas it ranges from €12.5/hour to €14.6/hour when a seat is not guaranteed. For instance, the (reference, $F^{[1]} =$ Gumbel, $Z^{(3)}$) model elicits a VOT equal to €14.57/hour when a seat is guaranteed and €12.55/hour when a seat is not guaranteed. The ranges of VOT elicited with design $Z^{(3)}$ are in line with the literature, which proved to be very heterogeneous depending on mode, travel purpose, type of survey, and so on (Abrantes and Wardman, 2011; Wardman and Wheat, 2013). In a SP survey, Arentze and Molin (2013) find values between €14.4/hour and €17.4/hour for train travel. With revealed preferences data in Switzerland, Glerum et al. (2014) elicit VOT around CHF12/hour (around €10/hour) for train and coach. Wardman et al. (2012) make a European meta-analysis of VOT. They find that, in France, train commuters have VOT between €4.5/hour and €9.4/hour, depending on the distance traveled. Similar values are found by the French rail operator, which elicit VOT between €3.9/hour and €7.7/hour (RFF, 2013) with a SP survey. The differential between VOT according to the level of comfort is quite low in comparison to the one found in a British literature review: with a load factor of 100%, VOT is 50% higher when the user has a standing position instead of a seating position (Wardman and Whelan, 2011). However, in comparison to our result, this high proportion may be explained by the fact that the traveler has no access to a seat during the whole trip and endures additional discomfort due to a high load.



(a) (reference, F, $Z^{(2)}$) models



(b) (reference, F, $Z^{(3)}$) models

FIGURE 8.5 – Simulated value of time for (A) the (reference, F, $\boldsymbol{Z^{(2)}}$) models and (B) the (reference, F, $\boldsymbol{Z^{(3)}}$) models, with the Krinsky and Robb method with 10,000 draws

The VOT elicited with the (reference, $F^{[1]}$ = Gumbel, $Z^{(3)}$) model and (reference, $F^{[5]}$ = Logistic, $Z^{(3)}$) models are relatively similar. However, and despite overall consistency of results with design $Z^{(3)}$, some disparities are observed across (reference, F, $Z^{(3)}$) models. They are not due to scale effects since VOT is calculated as a ratio of coefficients. The minimum values are observed with the (reference, $F^{[4]}$ = Cauchy, $Z^{(3)}$) model whereas the maximum values are elicited with the (reference, $F^{[6]}$ = Normal, $Z^{(3)}$) model when a seat is guaranted and the (reference, $F^{[1]}$ = Gumbel, $Z^{(3)}$) model when a seat is not guaranteed. These results highlight the impact of choosing a specific cdf on the estimation of economic outputs, with higher confidence in the results elicited with the best performing cdf.

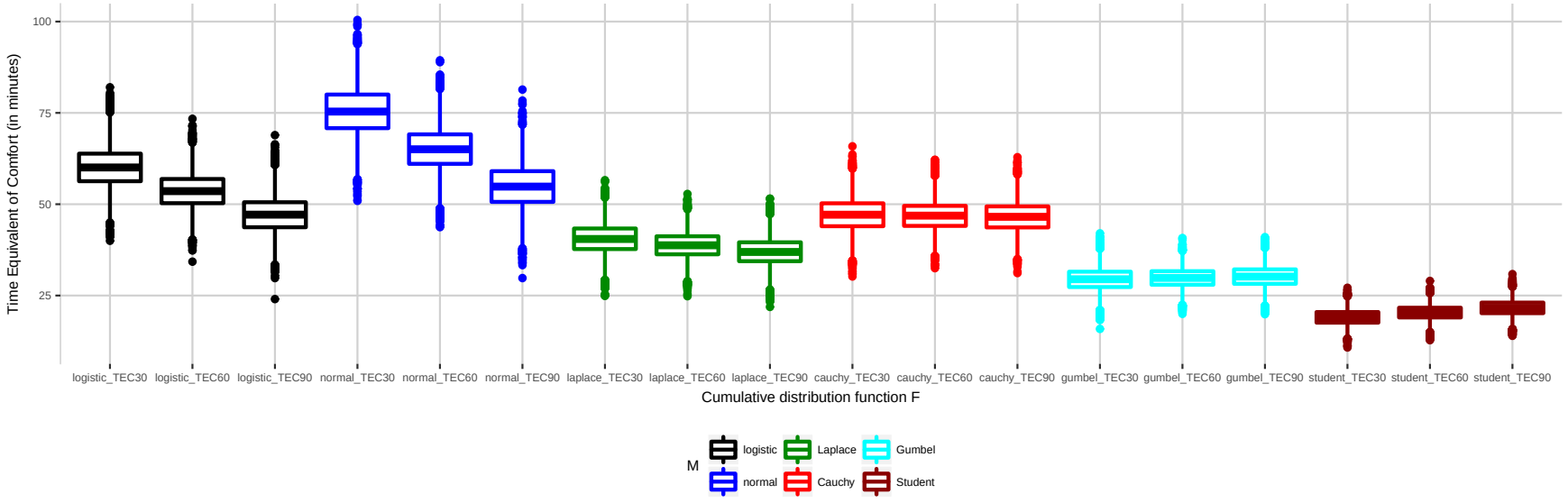
Whereas the (reference, F, $Z^{(3)}$) models provide consistent results, VOT elicited with the (reference, F, $Z^{(2)}$) models are lower and some cdfs provide inconsistent values (see Table 8.8 and Figure 8.5). Because of the negative coefficient of the cross- variable, Time $\times$ Comfort, the VOT is inconsistently lower when a seat is not guaranteed than when a seat is guaranteed, except for the two best models in terms of goodness of fit (reference, $F^{[1]}$ = Student$_{0.35}$, $Z^{(2)}$) and (reference, $F^{[2]}$ = Cauchy, $Z^{(2)}$). These two models are also the ones that elicit the VOT that are the closer to their $Z^{(3)}$ counterparts: (reference, $F^{[2]}$ = Student$_{1.9}$, $Z^{(3)}$) and (reference, $F^{[4]}$ = Cauchy, $Z^{(3)}$). The Student$_{\nu^\star}$ and Cauchy cdfs therefore seem to be more robust to unobserved heterogeneity than the other cdfs. Overall, the heterogeneity across cdfs is greater with design $Z^{(2)}$ than with design $Z^{(3)}$. VOT range from €4.6/hour to €7.4/hour if a seat is not guaranteed and from €6.0/hour and €7.1/hour if a seat is guaranteed. In particular, the best ranking model (reference, $F^{[1]}$ = Student$_{0.35}$, $Z^{(2)}$) elicits VOT equal to €7.1/hour and €7.4/hour, whereas the VOT elicited with the (reference, $F^{[5]}$ = logistic, $Z^{(2)}$) model are lower and respectively equal to €4.9/hour and €5.9/hour. We also observe overall that, with our data, the better the model in terms of goodness of fit, the higher the VOT.

8.2.3.3.4 Time equivalent of comfort According to Equation (I.2) in Appendix I, comfort is converted into minutes to obtain a time equivalent (see Table 8.8 and Figure 8.6). TEC is calculated for three travel times: 30 minutes, which is the first quartile of the sample; 60 minutes, the median, and 90 minutes, the last quartile of the sample. For (reference, F, $Z^{(3)}$) models, TEC increases when travel time increases, as expected. Depending on the cdf, TEC is comprised between 14.5 and 20.1 minutes of travel time for short travel time, 17.1 and 23.2 minutes for medium travel time and 19.6 and 26.3 minutes of travel time for long travel time. For instance, with the (reference, $F^{[1]}$ = Gumbel, $Z^{(3)}$) model, having a seat guaranteed is equivalent to 16.5 minutes of travel time for short travel time. It is equivalent to 20.6 minutes for medium travel time, and 20.6 minutes for long travel time. Although measuring the effect of crowding by seat availability is well adapted to the rail context in the

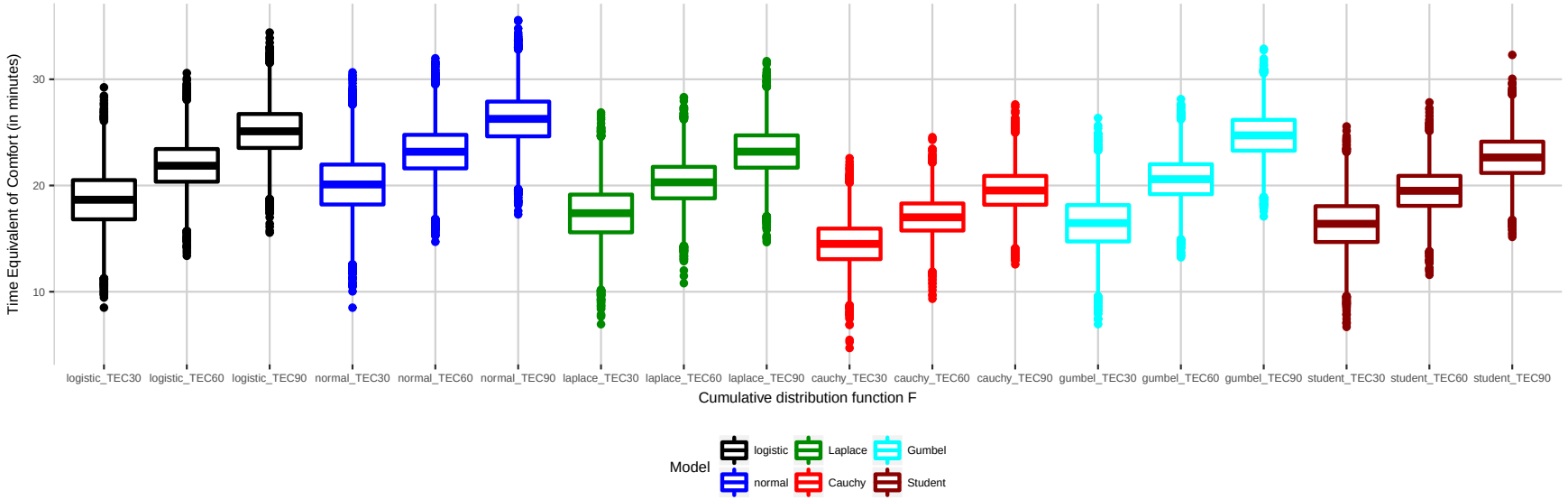
Rhône-Alpes Region, previous studies mainly adopted other measures for crowding, such as load factor or passengers per square meter (Wardman, 2014). Opportunities for comparisons are therefore limited but consistent with our results. In Richter and Keuchel (2012), the value of seat availability (during the whole trip compared to no free seat) is between 16 and 30 minutes (with a travel time between 15 and 90 minutes). In France, the rail operator (RFF, 2013) finds that users are ready to spend between 13 and 51 more minutes in the train to avoid a standing position in comparison to a seated position.

As for VOT, TEC are very close between the (reference, $F^{[1]} = \text{Gumbel}$, $Z^{(3)}$) and (reference, $F^{[5]} = \text{logistic}$, $Z^{(3)}$) models, although this does not prevent disparities across (reference, F, $Z^{(3)}$) models. The minimum and maximum values are observed with the same models as for VOT.

As expected, the TEC elicited with the (reference, F, $Z^{(2)}$) models are overall inconsistent and range from 19.0 to 75.3 minutes for short travel time, 20.3 to 65 minutes for medium travel time, and 21.6 to 54.6 minutes for long travel time. We observe that, the better the model in terms of goodness of fit, the lower the TEC. The simulated TEC are also more precise with the three best performing cdfs (see Figure 8.6a). Moreover, TEC inconsistently decrease as travel time increase except for the two best models. The best ranking model is also the only one which elicits TEC in the expected range of value. Indeed, depending on travel time, TEC elicited with the (reference, $F^{[1]} = \text{Student}_{0.35}$, $Z^{(2)}$) model are comprised between 19.0 and 21.6 minutes, whereas the (reference, $F^{[5]} = \text{Logistic}$, $Z^{(2)}$) model elicits values between 47.1 and 60.1 minutes. As for VOT, choosing the $\text{Student}_{\nu^\star}$ distribution over the logistic distribution is therefore, with our data, a means of retrieving consistent outputs.



(a) (reference, F, $Z^{(2)}$)



(b) (reference, F, $Z^{(3)}$)

FIGURE 8.6 – Simulated time equivalent of comfort for (A) the (reference, F, $\boldsymbol{Z^{(2)}}$) models and (B) the (reference, F, $\boldsymbol{Z^{(3)}}$) models

8.2.3.3.5 Elasticities For own-elasticities, we record only weighted elasticities. For cross-elasticities, both unweighted and weighted elasticities are recorded to discuss in depth the IIA property and substitution patterns.

Own-elasticities With design $Z^{(3)}$, the estimated own-elasticities are consistent with those from previous studies that take into account comfort variables (e.g., Daziano and Rizzi, 2015) (see Table 8.9 and Figure 8.7). Also consistent with the literature, the probability of choosing the train or the coach is more sensitive to its own cost and time than the probability of choosing the car (e.g. Glerum et al., 2014). For instance, with the (reference, $F^{[1]} =$ Gumbel, $Z^{(3)}$) model, a 10% increase in train cost induces a 5.1% decrease in the probability of choosing the train, whereas the same increase in car cost induces a 3.4% decrease in the probability of choosing this alternative. Moreover, a variation of 10% in travel cost has a lower impact on choice probabilities than the same variation in travel time (see De Jong and Gunn, 2001, for similar results).

With design $Z^{(3)}$, the comparison between models offers the same results as for VOT and TEC: own-elasticities are similar between the (reference, $F^{[1]} = \text{Student}_{1.9}$, $Z^{(3)}$) and (reference, $F^{[5]} =$ logistic, $Z^{(3)}$) models, but that does not prevent some disparities between models. Moreover, own-cost and -time elasticities are more homogeneous across models for the car alternative. One explanation could be that there is less unobserved heterogeneity for the car alternative, mainly chosen by car users who are more likely to stick to their reference mode than public transport users.

With design $Z^{(2)}$, the own-cost elasticities are higher than the own-time elasticities, which is inconsistent. The (reference, $F^{[1]} = \text{Student}_{0.35}$, $Z^{(2)}$) models elicit the own-cost and own-time elasticities closest to each other, and therefore reduce the effect of the misspecification. For instance, with the $\text{Student}_{0.35}$ distribution, the own-time elasticity for the train alternative is -0.98 and the own-cost elasticity -0.95, whereas they are respectively equal to -0.53 and -1.06 with the logistic distribution. Moreover, as for VOT and TEC, own-time elasticities estimated with the two best ranking models (reference, $F^{[1]} = \text{Student}_{0.35}$, $Z^{(2)}$) and (reference, $F^{[2]} =$ Cauchy, $Z^{(2)}$) are the closest to their $Z^{(3)}$ counterparts : (reference, $F^{[2]} = \text{Student}_{1.9}$, $Z^{(3)}$) and (reference, $F^{[4]} =$ Cauchy, $Z^{(3)}$). This result does not hold for the own-cost elasticities for which differences between (reference, F, $Z^{(2)}$) and (reference, F, $Z^{(3)}$) models are of the same order of magnitude whatever the cdf.

Cross-elasticities Any ratio of probabilities of two alternatives i and i' depends on the reference specific-alternative variables ω_J when F is not the logistic cdf. It implies that elasticities $\text{e}_i(\omega_J)$ and $\text{e}_{i'}(\omega_J)$ are different. This means that, if F is not the logistic cdf, a change in train travel time or cost

TABLE 8.9 – Own-time and own-cost elasticities for (reference, F, $Z^{(i)}$) models for $F \in \mathfrak{F}^0$ with $i = 2, 3$.

(reference, F, $Z^{(2)}$)	**logistic**	**normal**	**Laplace**	**Cauchy**	**Gumbel**	**Student$_{0.35}$**
Own-time elasticities						
Train	-0.535	-0.446	-0.696	-0.875	-0.631	-0.979
Coach	-0.601	-0.494	-0.82	-1.062	-0.79	-1.295
Car	-0.281	-0.24	-0.323	-0.424	-0.323	-0.462
Own-cost elasticities						
Train	-0.787	-0.708	-0.854	-0.97	-0.721	-0.948
Coach	-0.875	-0.777	-0.996	-1.169	-0.91	-1.253
Car	-0.583	-0.537	-0.612	-0.653	-0.531	-0.598
(reference, F, $Z^{(3)}$)	**logistic**	**normal**	**Laplace**	**Cauchy**	**Gumbel**	**Student$_{1.9}$**
Own-time elasticities						
Train	-1.06	-0.998	-1.107	-1.172	-1.128	-1.139
Coach	-1.37	-1.267	-1.47	-1.634	-1.468	-1.534
Car	-0.488	-0.486	-0.465	-0.398	-0.465	-0.453
Own-cost elasticities						
Train	-0.516	-0.488	-0.553	-0.619	-0.507	-0.573
Coach	-0.731	-0.67	-0.821	-0.989	-0.729	-0.866
Car	-0.364	-0.362	-0.355	-0.335	-0.339	-0.357

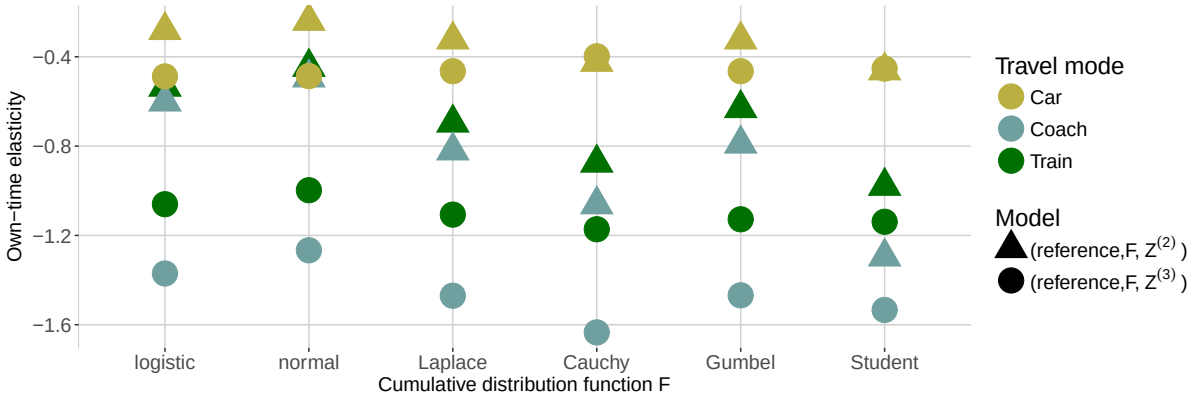


FIGURE 8.7 – Own-time elasticity for (reference, F, $Z^{(2)}$) models (triangles) and (reference, F, $Z^{(3)}$) models (circles) and three modes: car (brown), coach (blue) and train (green)

does not impact the probability of choosing the car or the coach in the same way. Conversely, a change in coach (respectively car) travel time or cost has the same effect on the probabilities of choosing the train or the car (respectively the train or the coach). For instance, with the (reference, $F^{[1]}$ = Gumbel, $Z^{(3)}$) model, when the travel time of train increases by 10%, then the probability of choosing the coach increases by 12.01% and the probability of choosing the car increases by 17.83% (see Table 8.10). Conversely, a 10% increase in coach travel time induces a 4.2% increase in the probability of choosing the train and coach modes. This higher substitution between the train and car alternatives than between the train and coach alternatives, also verified with the (reference, $F^{[6]}$ = normal, $Z^{(3)}$) model, is contrary to expectations. For the other (reference, F, $Z^{(3)}$) models, as well as the (reference, F, $Z^{(2)}$) models, we consistently observe a higher substitution between public transport modes, except for the worst ranking (reference, $F^{[6]}$ = normal, $Z^{(2)}$) model.

Once the cross-elasticities are weighted, substitutions between the two public transport modes (train and coach) are higher than substitutions between the public transport modes and car (see Table 8.11). Despite heterogeneous cross-elasticities across models, with design $Z^{(3)}$, the (reference, $F^{[1]}$ = Gumbel, $Z^{(3)}$) and (reference, $F^{[5]}$ = logistic, $Z^{(3)}$) models again provide similar results. With design $Z^{(2)}$, the (reference, $F^{[1]} = \text{Student}_{0.35}$, $Z^{(2)}$) model provides the results closest to its $Z^{(3)}$ counterpart (reference, $F^{[1]} = \text{Student}_{1.9}$, $Z^{(3)}$).

8.2.3.4 Discussion

Testing a variety of models generates more flexibility in the modeling process and allows the modeler to point out the model that best fits the data. Our results show that the choice of an appropriate cdf is important to retrieve consistent results, in particular, when the data contain high unobserved heterogeneity since some cdfs are more robust than others. In particular, the Student cdf is one of the two best cdfs for all three designs and is the more robust to unobserved heterogeneity. An explanation is that the flexibility in the degrees of freedom makes it possible to adapt the heaviness of the distribution tails to the data. It seems that the higher the unobserved heterogeneity, the lower the degrees of freedom selected. Indeed, design $Z^{(3)}$ contains the lowest unobserved heterogeneity, and we expect design $Z^{(2)}$ to contain the highest unobserved heterogeneity due to the lack of the individual variables on the type of user (“type_car” and “type_coach”). An interpretation is that thinner tails are needed when data are more informative about behavior. When unobserved heterogeneity is high, the predictor contains little information about the decision maker and the predictor is a poor basis for the detection of changes in choice probability. Distribution tails therefore

need to be heavier because the predictor only captures a raw, a gross mean behavior and fails to capture the heterogeneity in behaviors. With heavy-tailed distributions, probabilities are fairly insensitive to changes in predictor values, except around a threshold value around which the slope of the probability function is steep. The Student distribution with low degrees of freedom is able to capture the threshold value at which probability changes are abrupt but cannot capture the more subtle change in probabilities. When unobserved heterogeneity is controlled, a model with an asymmetric distribution stands out. Indeed, with design $Z^{(3)}$, the Gumbel cdf provides good results probably because of its asymmetric shape.

TABLE 8.10 – Unweighted cross-elasticities for (reference, $F_{m,s}$, $Z^{(i)}$) models for $F_{m,s} \in \mathfrak{F}_{q_{0.95}}$ with $i = 2, 3$.

(reference, F, $Z^{(2)}$)		**logistic**	**normal**	**Laplace**	**Cauchy**	**Gumbel**	**Student$_{0.35}$**
Cross-time elasticities							
Train	Coach	0,252	0,191	0,393	0,589	0,487	0,845
	Car	0,252	0,221	0,28	0,291	0,393	0,198
Coach	Train	0,174	0,139	0,251	0,34	0,218	0,447
	Car	0,174	0,139	0,251	0,34	0,218	0,447
Car	Train	0,31	0,266	0,378	0,446	0,34	0,431
	Coach	0,31	0,266	0,378	0,446	0,34	0,431
Cross-cost elasticities							
Train	Coach	0,355	0,272	0,488	0,69	0,584	0,892
	Car	0,355	0,33	0,343	0,342	0,341	0,212
Coach	Train	0,248	0,215	0,296	0,361	0,245	0,417
	Car	0,248	0,215	0,296	0,361	0,245	0,417
Car	Train	0,637	0,589	0,647	0,679	0,552	0,552
	Coach	0,637	0,589	0,647	0,679	0,552	0,552
(reference, F, $Z^{(3)}$)		**logistic**	**normal**	**Laplace**	**Cauchy**	**Gumbel**	**Student$_{1.9}$**
Cross-time elasticities							
Train	Coach	0,671	0,455	0,873	1,287	1,201	1,028
	Car	0,671	0,75	0,529	0,26	1,783	0,465
Coach	Train	0,427	0,382	0,474	0,558	0,42	0,502
	Car	0,427	0,382	0,474	0,558	0,42	0,502
Car	Train	0,746	0,801	0,663	0,497	0,631	0,613
	Coach	0,746	0,801	0,663	0,497	0,631	0,613
Cross-cost elasticities							
Train	Coach	0,293	0,158	0,465	0,83	0,553	0,57
	Car	0,293	0,32	0,242	0,139	0,578	0,223
Coach	Train	0,224	0,2	0,257	0,324	0,205	0,275
	Car	0,224	0,2	0,257	0,324	0,205	0,275
Car	Train	0,559	0,605	0,517	0,416	0,46	0,48
	Coach	0,559	0,605	0,517	0,416	0,46	0,48

TABLE 8.11 – Weighted cross-elasticities for (reference, $F_{m,s}$, $Z^{(i)}$) models for $F_{m,s} \in \mathfrak{F}_{q_{0.95}}$ with $i = 2, 3$.

(reference, F, $Z^{(2)}$)		**logistic**	**normal**	**Laplace**	**Cauchy**	**Gumbel**	**Student**$_{0.35}$
Cross-time elasticities							
Train	Coach	0.281	0.22	0.431	0.610	0.419	0.893
	Car	0.201	0.178	0.226	0.238	0.205	0.165
Coach	Train	0.194	0.155	0.271	0.362	0.24	0.458
	Car	0.139	0.115	0.194	0.251	0.181	0.338
Car	Train	0.276	0.234	0.323	0.431	0.323	0.457
	Coach	0.273	0.231	0.317	0.408	0.317	0.398
Cross-cost elasticities							
Train	Coach	0.374	0.315	0.495	0.649	0.458	0.88
	Car	0.313	0.297	0.292	0.276	0.243	0.152
Coach	Train	0.259	0.226	0.301	0.362	0.255	0.408
	Car	0.216	0.192	0.253	0.298	0.221	0.347
Car	Train	0.569	0.521	0.595	0.655	0.526	0.583
	Coach	0.57	0.519	0.597	0.642	0.527	0.527
(reference, F, $Z^{(3)}$)		**logistic**	**normal**	**Laplace**	**Cauchy**	**Gumbel**	**Student**$_{1.9}$
Cross-time elasticities							
Train	Coach	0.766	0.648	0.927	1.244	0.889	1.025
	Car	0.31	0.334	0.252	0.129	0.309	0.225
Coach	Train	0.523	0.478	0.57	0.652	0.556	0.597
	Car	0.268	0.243	0.295	0.345	0.263	0.313
Car	Train	0.405	0.395	0.395	0.372	0.425	0.399
	Coach	0.578	0.583	0.524	0.411	0.528	0.506
Cross-cost elasticities							
Train	Coach	0.317	0.252	0.431	0.672	0.369	0.491
	Car	0.175	0.191	0.139	0.061	0.151	0.124
Coach	Train	0.227	0.206	0.259	0.32	0.224	0.274
	Car	0.174	0.157	0.199	0.252	0.161	0.214
Car	Train	0.307	0.298	0.312	0.317	0.314	0.319
	Coach	0.426	0.427	0.404	0.342	0.379	0.393

8.2.4 Conclusion

Standard economic outputs (VOT, TEC, marginal effects and elasticities) have been theoretically described for the family of reference models. Reference models allow the analyst to derive economic outputs with formulas generalizing those derived with the multinomial logit model. For the particular case of the reference model using the logistic cdf, the formulas are those described in the literature.

Estimations of reference models on a real dataset describing travel mode choices show that the multinomial logit model is outperformed by other reference models. The choice of an appropriate cdf is important to retrieve consistent results, in particular, when the data contain high unobserved heterogeneity. The heterogeneous results observed between cdfs, and in particular the differences in results between the multinomial logit and the best performing reference models, have operational impacts. Firstly, depending on the chosen cdf, practitioners will not retain the same variables as leverage of public policies. Secondly, economic outputs, like willingness-to-pay indicators or elasticities, are routinely used in economic models such as cost-benefit analyses or demand forecasting models, which are, in turn, used to derive public policies. For example, VOT and time equivalent (of comfort, delay, and so on) are key components of transport models, such that differences between models cannot be ignored.

A first avenue for future research is to select the best cdf. On our dataset, the Student cdf is one of the two best cdfs for all three designs and is the more robust to unobserved heterogeneity. Using both individual and alternative specific variables, the Gumbel cdf provides good results because of its asymmetric shape. It would therefore be interesting to test the non-central Student distribution, defined with two parameters: one for the heavy/thin tails and the other for the asymmetry. In order to propose a better comparison of the different reference models, a large simulation study would be appropriate, using cross-validation to select the best cdf.

A second avenue concerns the choice of the reference alternative. Evidence from the economic literature shows that the alternatives chosen as a reference by the decision makers are certainly not homogeneous across the population. It would thus be relevant to use a mixture model in which different groups of people would have different reference alternatives. Finally, the usual extensions of the multinomial logit model, such as the mixed logit model or the nested logit model, can also be developed for reference models. The first extension is straightforward using the GLM framework. The second extension is described by Peyhardi et al. (2016) for the case of nested alternatives, and economic outputs should also be studied.

8.3 Work in progress and perspectives

A key feature of reference models is the dependence of the choice probabilities on the alternative chosen as a reference. Far from being a weakness, this dependence is supported by much experimental evidence (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991; Tversky and Simonson, 1993; Bateman et al., 1997; Herne, 1998; Swait et al., 2002; Knetsch, 2010). Hudík (2014) shows that the possibility of formally incorporating reference dependence into the theory of consumer behavior has already been explored in the late 1930s, in the context of discussions on marginal utility and ordinality/cardinality of utility. Models and theories such as (cumulative) prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992), reference-dependent theory (Tversky and Kahneman, 1991) or the model of reference-dependent preferences (Köszegi and Rabin, 2006) have been built to take into account the reference dependence of choices.

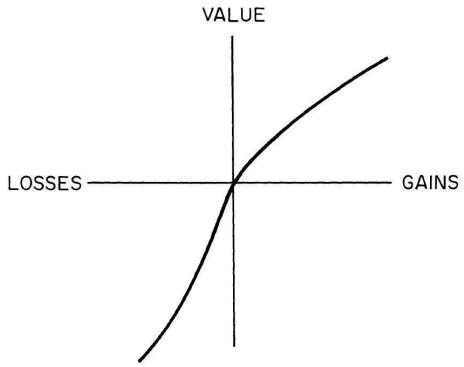
This section firstly presents the reference-dependent theory of consumer choice. Secondly, a reference-dependent utility is defined and the behavioral implications of this utility are discussed. Thirdly, a mixture of reference models and latent-class models is developed.

8.3.1 Reference-dependent theory of consumer choice

Theories of reference-dependent preferences propose that individuals evaluate outcomes as gains or losses relative to a neutral reference state. Prospect theory (Kahneman and Tversky, 1979) is the most well-known and influential account of reference-dependent preferences. It applies to choices under uncertainty and has been generalized for riskless choices by the reference-dependent theory of consumer choice (Tversky and Kahneman, 1991).

Like prospect theory, the reference-dependent theory of consumer choice assumes that decision makers first frame the outcomes of each alternative by coding them as gains and losses relative to a reference point, and then choose the outcome with the highest value. The value function has three essential features: reference dependence; diminishing sensitivity, and loss aversion. Firstly, reference dependence recognizes different value functions for gains and losses defined relative to a reference point. Secondly, diminishing sensitivity means that the marginal value of both gains and losses decreases with their size. Thirdly, loss aversion is defined by a steeper value function in the losses than in the gains domain. These characteristics give rise to an asymmetric S-shaped value function, concave above the reference point and convex below it (Figure 8.8).

As a consequence of these properties, a shift of reference induces reversals of preferences. Tversky



Source: Tversky and Kahneman (1991)

FIGURE 8.8 – An illustration of a value function which respects the three features of the reference-dependent theory

and Kahneman (1991) therefore introduce a family of preference relations indexed to a reference point. For instance, the relation $\geq_r$ defines a weak preference from reference point r. For the sake of illustration, they define a choice set $\mathcal{X} = (x, y, z \dots)$. Each option $x = (x_1, x_2) \in \mathcal{X}$ is a bundle that offers x_1 units of good 1 and x_2 units of good 2. x may also be a discrete alternative. For instance, if x is a transport mode, then x_1 may represent the travel time and x_2 the travel cost. They assume that the preference relation $\geq_r$ is complete, transitive and continuous and that each preference order is strictly monotonic. Under these assumptions, they state that each $\geq_r$ can be represented by a strictly increasing continuous reference-dependent utility function. Köszegi and Rabin (2006) also define a reference-dependent utility, which is composed of an intrinsic "consumption utility" independent of the reference level, and a gain/loss utility. In both models, the reference-dependent utility is defined as a separable function which considers for each good, or attribute, the gains or losses to a specific reference.

8.3.2 A reference-dependent utility

The probability function (Equations (8.15) and (8.16)) can be expressed in terms of a reference-dependent utility $v(\eta_j)$:

$$\pi_j = \begin{cases} \frac{\exp\left(v(\eta_j)\right)}{1+\sum_{i=1}^{J-1}\exp\left(v(\eta_i)\right)} & , \forall j \neq J \\ 1 - \sum_{j=1}^{J-1} \pi_j & , \text{for } j = J \end{cases} \tag{8.21}$$

with

$$v(\eta_j) = \begin{cases} \ln \circ \frac{F}{1-F}(\eta_j) & , \quad \forall j = 1 \neq J \\ 0 & , \quad \text{for } j = J. \end{cases} \tag{8.22}$$

In the general specification of reference models, v_j is not linear in the predictor η_j and is transformed by a function v_j. Note that if F is the logistic cdf, then $\eta_j = \ln \circ \frac{F}{1-F}(\eta_j)$.

For reference models other than the (reference, logistic, Z) model, the probability function is dependent on the alternative chosen as a reference. Indeed, if F is not the logistic cdf, then π_j is a function of x, ω_j and ω_J. For the (reference, logistic, Z) model, the choice probability does not depend on the chosen reference alternative (e.g. J or K), since

$$\begin{aligned} \pi_j &= \frac{\exp\left((\alpha_j^\star - \alpha_J^\star) + x^T(\delta_j^\star - \delta_J^\star) + (\omega_j - \omega_J)^T\gamma\right)}{\sum_{i=1}^{J} \exp\left((\alpha_i^\star - \alpha_J^\star) + x^T(\delta_i^\star - \delta_J^\star) + (\omega_i - \omega_J)^T\gamma\right)} \\ &= \frac{\exp\left((\alpha_j^\star - \alpha_K^\star) + x^T(\delta_j^\star - \delta_K^\star) + (\omega_j - \omega_K)^T\gamma\right)}{\sum_{i=1}^{J} \exp\left((\alpha_i^\star - \alpha_K^\star) + x^T(\delta_i^\star - \delta_K^\star) + (\omega_i - \omega_K)^T\gamma\right)}, \ \forall j = 1, \cdots, J. \end{aligned} \tag{8.23}$$

8.3.2.1 Behavioral implications of reference dependence

The specification of reference models implies that the probability of choosing alternative j depends on the difference in attributes between alternative j and alternative J. A behavioral interpretation is that when making a choice, decision makers first choose an alternative as a reference and then make paired comparisons as a way to simplify the decision process. The difference in predictors is processed through a function v_j to compute the utility of choosing one alternative over the others.

For example, if there are three alternatives and alternative 1 is chosen as the reference, then alternative 2 is compared to alternative 1, alternative 3 is compared to alternative 1 and alternative 2 is compared to alternative 3 in terms of differences relative to alternative 1. More generally, if J is the reference alternative, then j ($j \neq J$) is the preferred alternative if

$$\begin{aligned} \eta_j &\geq \eta_k, \ \forall k \neq j, \ k \neq J \\ \eta_j &\geq 0. \end{aligned} \tag{8.24}$$

J is the preferred alternative if

$$\eta_j \leq 0, \ \forall j \neq J. \tag{8.25}$$

Note that the preference order of the alternatives does not depend on the reference alternative, since

the reference-dependent utility function $v(.)$ is an increasing function:

$$\begin{aligned} &\eta_j \geq \eta_k \Leftrightarrow v(\eta_j) \geq v(\eta_k),\ \forall k \neq j,\ k \neq J \\ &\text{and} \\ &\eta_j \geq 0 \Leftrightarrow v(\eta_j) \geq 0. \end{aligned} \tag{8.26}$$

As in the reference-dependent theory, the level of the reference-dependent utility $v(\eta_j)$, and, therefore, the choice probabilities, depend on the attributes of the reference alternative. However, the reference-dependent theory postulates that comparisons between alternatives are made at the level of each good, or attribute, with the possibility of evaluating preference orders from different points. In reference models, comparisons between alternatives are made attribute by attribute before being processed through the reference-utility at an aggregated level. This explains why there is no reversal of preferences in the reference models. Moreover, in reference models, the reference point is a specific alternative and, therefore, depends on the choice set, while in the reference-dependent theory, the question of the reference point remains unsolved. It is generally defined by some operationalization of the *status quo*, the endowment, or the decision maker's recent expectations about outcomes (Köszegi and Rabin, 2006), for instance the price at which she expects to buy a good.

8.3.2.2 Behavioral implication of the choice of F

This section provides a comparison across the reference models to gain a deeper insight into their behavioral interpretation. Assume that the design Z is fixed. The goal is to study how the chosen cdf F impacts the shape of the reference-dependent utility function and the distribution of the probabilities.

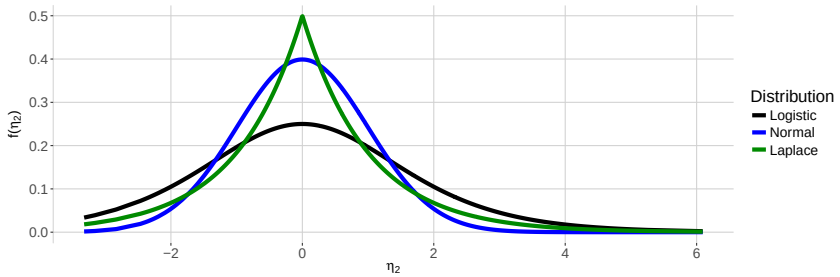
The behavior of 1,000 individuals are simulated with the following hypotheses:

- alternative-specific constants: $\alpha_2 = 1$ and $\alpha_3 = 0.5$;
- individual variables: x is normally distributed; $\delta_2 = 0.5$ and $\delta_3 = 1$;
- alternative-specific variables: ω_1, ω_2 and ω_3 are normally distributed; $\gamma = 1$.

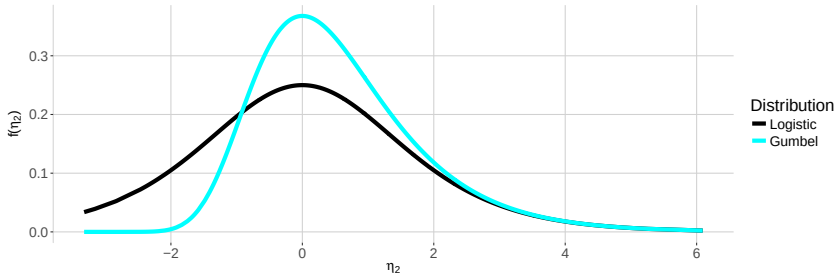
With these hypotheses, η_2 ranges between -3.6 and 5.9 and η_3 between -5.4 and 6.0. Figure 8.9 displays the logistic, normal, Laplace, Gumbel and Student density functions to highlight the main characteristics of each distribution.[5] The normal distribution has thinner tails than the logistic distribution. The Laplace distribution has heavier tails than the normal distribution with a sharp peak. The Gumbel distribution is skewed to the left and thus asymmetric. The Student distribution has

5. Scale parameters are set to 1 and location parameters to 0. For the Student distribution, the degrees of freedom has been chosen as equal to 0.35 because it is the optimal ν with the real dataset in Section 8.2.

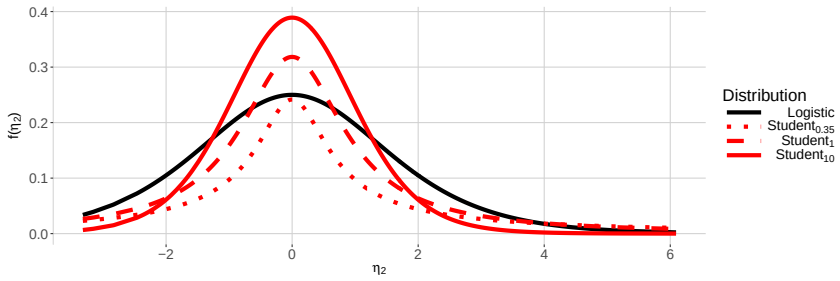
heavier tails than the normal distribution. It corresponds to the Cauchy distribution when $\nu = 1$ and the normal distribution when $\nu \to \infty$.



(a) logistic, normal, Laplace density functions



(b) logistic, Gumbel density functions



(c) Logistic, Student$_\nu$ density functions with $\nu \in \{0.35, 1, 10\}$

FIGURE 8.9 – Density functions of distributions used in reference models

Figure 8.10 displays the reference-dependent utility functions as defined in Equation (8.22) for six reference models. The levels of $v(\eta_j)$ vary depending on the chosen cdf but are not comparable and need to be normalized (see Section 8.2). Additionally, Figure 8.11 displays normalized reference-dependent utility functions estimated with the dataset used by Greene and Hensher (2003).

The shapes of the reference-dependent utility functions may be discussed regarding the three features of the reference-dependent theory of Tversky and Kahneman (1991):

1. Reference dependence: all distributions, except the logistic distribution lead to a non-linear utility function and capture the reference dependence. Indeed, the marginal reference-dependent utility depends on the level of η_j and therefore on the level of ω_J. For the

logistic distribution, $v(\eta_j)$ is linear in changes of η_j, meaning that the marginal reference-dependent utility is constant, whatever the difference between alternative $j(=2)$ and alternative $J(=1)$.

2. Diminishing sensitivity: the three distributions with the heavier tails (Laplace, Cauchy and Student$_{0.35}$) capture diminishing sensitivity since their associated reference-dependent utility is concave in the gains and convex in the losses. The marginal value of both gains and losses decreases with their size. The shape of the reference-dependent utility function defined with Gumbel distribution is also concave in gains. Due to the asymmetric shape of the Gumbel distribution, $v(\eta_j)$ is also concave in the losses until a threshold value $\tilde{\eta}_j$ and becomes convex when $\eta_j < \tilde{\eta}_j$. For the normal distribution, $v(\eta_j)$ increases with η_j. In other words, the higher the difference between alternative j and alternative J, the more the reference-dependent utility increases due to a change of one unit in the difference. This shape of the reference-dependent utility associated with the normal distribution is contrary to the theory.
3. Loss aversion: all distributions except the Gumbel distribution are symmetric and therefore depict similar behaviors in gains and losses. Concerning the reference-dependent utility with the Gumbel distribution, due to the concavity above $\tilde{\eta}_j$, the slope in losses is higher than in gains but with no discontinuity. Put differently, when alternative j is perceived as better than alternative J ($\eta_j > 0$), then the marginal reference-dependent utility decreases as η_j increases. When alternative j is perceived as worse than alternative J ($\eta_j < 0$), the higher η_j, the higher the decrease in $v(\eta_j)$ as long as $\eta_j \geq \tilde{\eta}_j$. The Gumbel distribution therefore partially depicts a loss aversion.

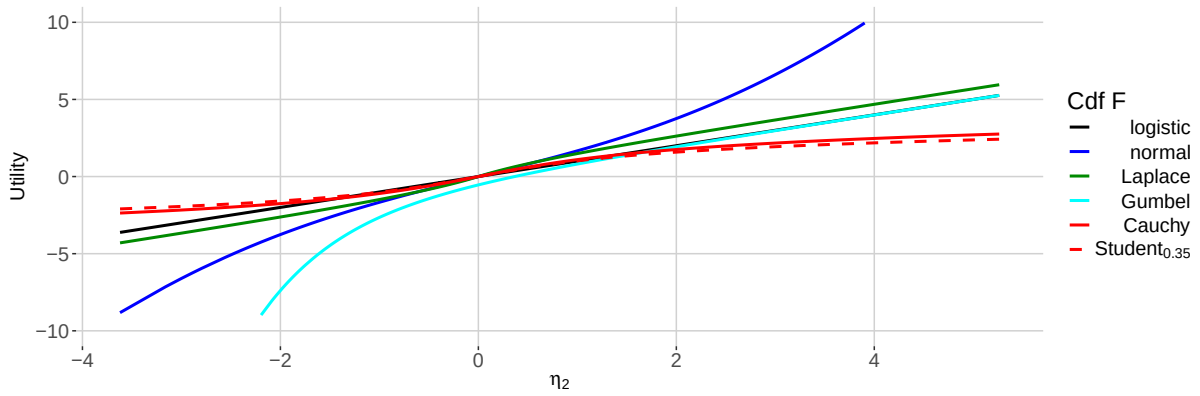


FIGURE 8.10 – Reference-dependent utility functions on simulated data where F is the logistic (black), normal (blue), Laplace (green), Gumbel (cyan), Cauchy (red), Student$_{\nu=0.45}$ (dashed red line) and Student$_{\nu=0.05}$ (dotted red line) cdf

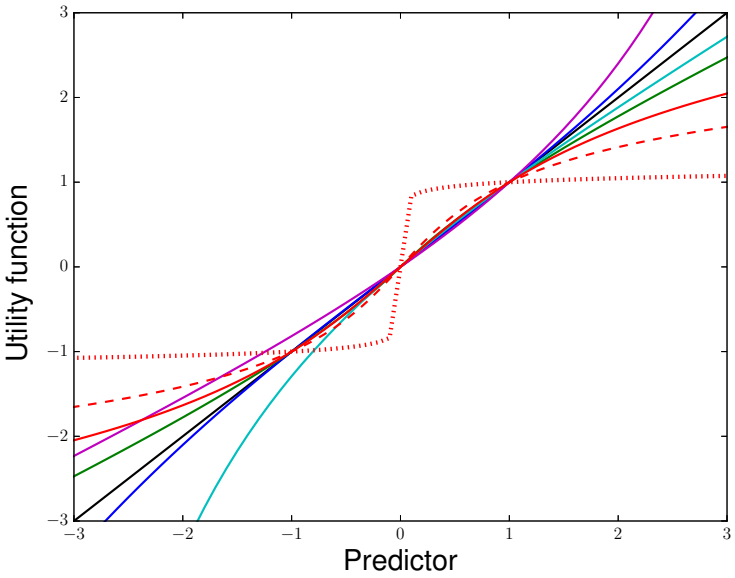


FIGURE 8.11 – Normalized reference-dependent utility functions estimated with the Greene (2003) dataset where F is the logistic (black), normal (blue), Laplace (green), Gumbel (cyan), Cauchy (red), Student$_{\nu=0.45}$ (dashed red line) and Student$_{\nu=0.05}$ (dotted red line) cdf

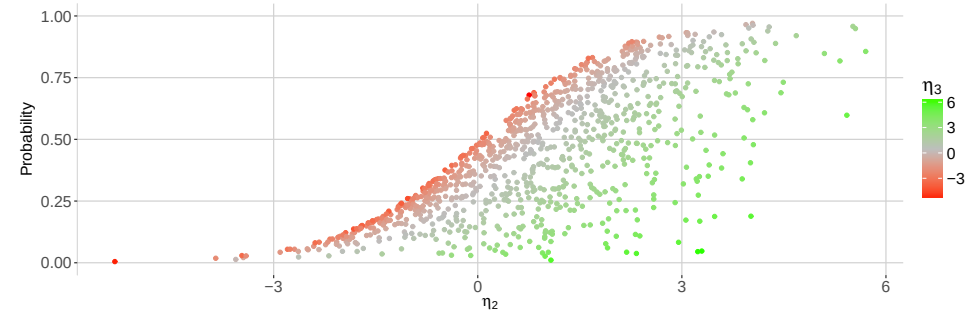
In short, none of the six distributions respects the three characteristics of the reference-dependent theory (see Table 8.12). However, the compliance of some reference models with reference dependence, diminishing sensitivity or loss-aversion highlights the behavioral value of reference models in comparison to the multinomial logit model.

TABLE 8.12 – Reference models compliance with the reference-dependent theory

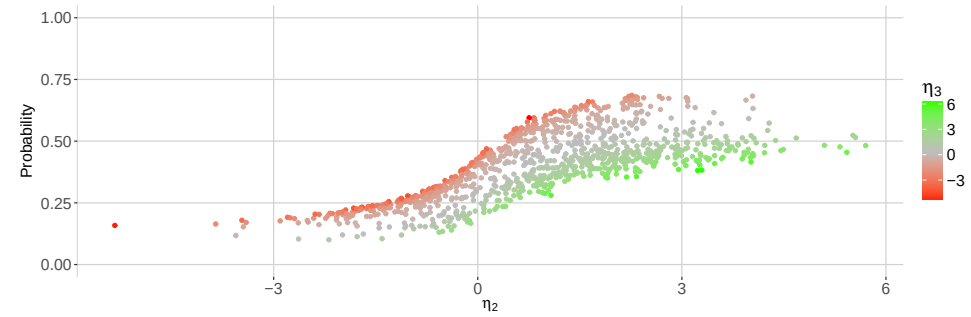
Distribution	**Reference dependence**	**Diminishing sensitivity**	**Loss-aversion**
logistic	No	No	No
normal	Yes	No	No
Laplace	Yes	Yes	No
Cauchy	Yes	Yes	No
Gumbel	Yes	Partially	Partially
Student	Yes	Yes	No

Figure 8.12 displays the probability of choosing option $j(=2)$ for the logistic, Student$_{0.35}$, Laplace and Gumbel distributions. The higher the predictor η_2, the higher the probability of choosing option 2 (π_2), and the higher the predictor η_3, the lower the probability of choosing option 2. Due to the heavy tails of the Student$_{0.35}$ distribution, π_2 takes a smaller range of values for the (reference, Student$_{0.35}$, Z) model than for the other models. For instance, with the simulated data, π_2 maximum value is 0.72, even for extreme values of η_2 and η_3. Moreover, for a given η_2, the range of values taken by π_2 depending on

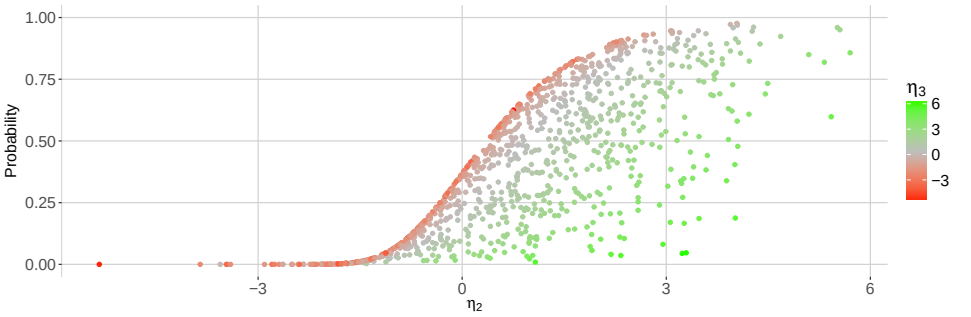
η_3 is narrower than for the other distributions. Due to the asymmetric shape of the Gumbel distribution, the difference between alternative 2 and alternative 1 has to be higher than with the other distributions to allow π_2 to take values above 0.1. The slope of π_2 for the (reference, Gumbel, Z) model is also steeper than the slope for the (reference, logistic, Z) distribution.



(a) (reference, logistic, Z) model



(b) (reference, Student$_{0.35}$, Z) model



(c) (reference, Gumbel, Z) model

FIGURE 8.12 – Probability of choosing option 2, according to η_2 and η_3

8.3.3 A mixture of reference models and latent-class models

For reference models, the probability function is not invariant under the alternative chosen as a reference. Therefore, the analyst has to choose the same reference alternative as the decision maker. However, this choice is unknown and hypotheses have to be formulated. For stated preferences

surveys, the most intuitive reference alternative is the *status quo* option (if any in the choice set) or the first displayed option. For revealed preferences surveys, the reference alternative depends on the field and the objective of the survey. If the objective is to model mode choice, the reference alternative may be the mode used most frequently, either for a specific journey or in general. In both cases, the reference alternative depends on the decision maker and is not known with certainty by the modeler. Therefore, the model has to take into account this latent heterogeneity. One way of doing this is to combine the reference models with a latent class model. In latent class models (Greene and Hensher, 2003), the utility function can be written differently for each class. The assignment to a specific class is probabilistically determined by a set of explanatory variables. In a choice set with J alternatives, each alternative $k = 1, \ldots, J$, may be chosen as a reference. If alternative k is chosen as a reference, then the deterministic part of the reference-dependent utility for alternative j is

$$v^{\{k\}}(\eta_j) = v^{\{k\}}(\alpha_j + x^T\delta_j + (\omega_j - \omega_k)\gamma),\ \forall j \neq k. \tag{8.27}$$

The probability of choosing option j conditionally on having option k as a reference is

$$\pi_j^{(k)} = \begin{cases} \frac{\exp\left(v^{\{k\}}(\eta_j)\right)}{1+\sum_{i\neq k}\exp\left(v^{\{k\}}(\eta_i)\right)} & , \quad \forall j \neq k \\ 1 - \sum_{i\neq k}\pi_i^{\{k\}} & , \quad \text{for } j = k \end{cases} \tag{8.28}$$

The choice of the reference alternative depends on the individual characteristics $x^\star$ of the decision maker. $x^\star$ may include the actual transport mode , the mobility styles or habits, the motorization, the holding of a season ticket, and so on. The probability of considering k as the reference alternative, $\tilde{p}^{\{k\}}$, is equivalent to the probability of being in class k. It is defined by a multinomial logit model:

$$\tilde{p}^{\{k\}} = \frac{\exp\left(\theta^{\{k\}T}x^\star\right)}{\sum_{i=1}^{J}\exp\left(\theta^{\{k\}T}x^\star\right)},\ \forall k = 1, \ldots, J, \tag{8.29}$$

where $\theta^{\{k\}}$ is the parameters vector associated with class $k = 1, \cdots, J$. When Equations (8.28) and (8.29) are combined, the probability of choosing j with a combined reference/latent-class model is

$$\pi_j^{\text{LC}} = \sum_{k=1}^{J}\pi_j^{\{k\}} \times \tilde{p}^{\{k\}}, \forall j = 1, \ldots, J \tag{8.30}$$

These models still need to be implemented on real and simulated datasets. This work is left for future research.

Conclusion Générale

En utilisant les données d'une enquête de préférences déclarées à l'échelle de la Région Rhône-Alpes, les travaux de la thèse mesurent l'impact, sur le choix modal, de déterminants observables, le temps et le coût de déplacement, le temps et la probabilité de retard, la fréquence de passage des trains et des cars, le cadencement des horaires et le confort, et les traduisent en consentements à payer et équivalents temps. Ces outputs économiques sont synthétisés et discutés dans l'annexe J.

Les théories comportementales permettent à la fois de mieux prendre en compte la façon dont ces déterminants influencent le choix modal et d'élargir le champ des déterminants qui peuvent être analysés. En particulier, trois enseignements de théories de psychologie et d'économie comportementale ont été introduits dans les modèles économétriques et ont servi à en discuter la portée théorique et empirique. Premièrement, le choix modal a été discuté à l'aide de variables latentes, portant notamment sur le confort et l'utilité indirecte de se déplacer. Deuxièmement, l'hétérogénéité de la prise en compte des probabilités de retard a été analysée dans un modèle d'utilité dépendante au rang. Troisièmement, les modèles de référence ont été comparés au logit multinomial en s'appuyant sur le modèle de dépendance à la référence de Tversky et Kahneman (1991). Les modèles de choix discret estimés dans la thèse, et en particulier les modèles de choix hybrides et les modèles de référence, sont un moyen de rendre opérationnels les enseignements de ces théories comportementales puisque leur estimation va au-delà de la description des processus comportementaux et permet d'expliquer et de prévoir les choix individuels et donc la demande, ce qui est essentiel en économie pour identifier des leviers d'actions publiques, simuler les effets des politiques publiques, cibler des segments d'usagers, calculer des parts de marché, des élasticités ou encore des consentements à payer.

La conclusion résume les principaux résultats et apports de la thèse, discute des implications en termes de politiques publiques, puis identifie des limites et propose de futures pistes de recherche.

Principaux résultats

Réalisation d'une enquête de préférences déclarées

La réalisation d'une enquête de préférences déclarées a permis d'améliorer la base de connaissance disponible en Région Rhône-Alpes pour l'analyse du choix modal, et plus particulièrement le choix

entre train, car et voiture. Les données disponibles avant la thèse ne permettaient pas d'observer l'effet de certains attributs, le temps et la probabilité de retard, le cadencement des horaires de train et de car, ainsi que le confort, sur le choix modal, pourtant nécessaires à la réalisation d'un modèle régional de prévision de demande de transport intégrant le train. L'enquête de préférences déclarées, en plus de s'appuyer sur une analyse qualitative préalable, comporte deux principales originalités. La première est d'être **personnalisée** pour chaque individu, tout en étant réalisée à l'échelle régionale, et non sur un unique corridor. La personnalisation des questions de choix a été permise par la constitution d'une base de données comportant les temps de parcours (en train et en voiture) pour l'ensemble des 8,6 millions de paires origine-destination pour lesquelles un déplacement en train est réalisable. La deuxième originalité de l'enquête est la **mesure de variables latentes** à l'aide d'indicateurs mesurés sur des échelles de Likert qui ont été testées par une enquête et une analyse préalables. Trois catégories d'indicateurs sont retenues : 1) les indicateurs mesurant la conscience environnementale, et plus particulièrement le nouveau paradigme environnemental (*New environmental paradigm*), la reconnaissance des problèmes liés à l'utilisation de la voiture et l'attitude vis-à-vis des politiques de transport ; 2) les indicateurs mesurant les attitudes vis-à-vis de la voiture, et plus particulièrement les motivations affectives et symboliques d'utilisation de la voiture ; 3) les indicateurs mesurant les perceptions des transports en commun, et plus particulièrement la perception de contrôle sur l'usage des transports en commun, la perception du temps dans le train et le car et les sentiments éprouvés en voyageant en train ou en car.

Revue de littérature sur les modèles de choix hybrides

L'incorporation de variables latentes mesurées à l'aide de ces indicateurs dans un modèle de choix discret, n'est pas triviale d'un point de vue économétrique. Les travaux de Ben-Akiva et al. (2002b) ont permis de poser les bases d'une telle intégration au travers des modèles de choix hybrides. La revue de littérature a identifié comment les **variables latentes** introduites dans les modèles de choix hybrides expliquent le choix modal et a mis en évidence leur **hétérogénéité, en termes de mesure et de définition**. Afin de faciliter l'analyse de l'effet des variables latentes sur le choix modal, une **grille de lecture à trois entrées** est proposée : 1) une entrée thématique (trait de personnalité, conscience environnementale, confort, sécurité, fiabilité, flexibilité, accessibilité et habitudes) ; 2) une entrée qui s'appuie sur des concepts psychologiques et distingue les variables portant sur une description intrinsèque de l'individu (ses normes, valeurs, besoins), les variables portant sur la perception que l'individu a qu'a l'individu des modes de transport (par exemple, la perception du confort dans le

train) ; 3) une entrée qui distingue les variables considérant les individus isolément et celles qui prennent en compte leur intégration dans un groupe. Par ailleurs, la revue de littérature montre que les **enseignements de la recherche en gestion et en psychologie dans la mesure des variables latentes**, notamment pour assurer leur validité interne et externe, sont relativement peu pris en compte. Un autre enseignement est que les **modèles de psychologie comportementale**, qui pourraient théoriquement être incorporés dans les modèles de choix hybrides afin de combiner à la fois une représentation plus fine du processus décisionnel et les avantages des modèles de choix discret (fondements économiques, prise en compte explicite des facteurs contextuels) sont, à quelques exceptions près, soit occultés, soit **simplifiés**. Les variables latentes sont généralement incorporées dans l'utilité sans prendre en compte les relations de médiation ou de modération décrites dans les modèles de psychologie. Un corollaire est que les modèles structurels des modèles de choix hybrides ne reposent que sur des variables socio-économiques observables et semblent faiblement contribuer à expliquer les valeurs prises par les variables latentes.

Analyse de l'hétérogénéité de la valeur du temps selon le confort objectif et perçu dans les transports en commun

Le modèle de choix hybride estimé dans le cadre de la thèse confirme la difficulté à expliquer les variables latentes uniquement à partir de variables socio-économiques. Malgré cette limite, le modèle rend compte de l'influence du confort, objectif ou perçu, sur le choix modal et la valeur du temps dans le train et le car. Ce travail repose sur le constat que se déplacer n'est pas uniquement une perte de temps mais peut permettre de pratiquer des activités, différentes selon le mode de transport et le niveau de confort qui y sont associés. Les résultats montrent que **la garantie d'une place assise dans le train, qui mesure le confort « objectif », n'explique pas le choix modal mais diminue la valeur du temps associée au train**. Par ailleurs, les personnes qui perçoivent l'usage du car et du train comme étant contraignant, notamment pour voyager avec des enfants ou des bagages, empruntent moins les transports en commun. En revanche, s'ils voyagent en train ou car, leur valeur du temps n'est pas plus élevée que celle des autres voyageurs. La **perception de contrôle sur l'usage des transports en commun a donc un effet sur le comportement *a priori* mais ne modifie pas la perception du temps une fois le choix modal réalisé**. Enfin, éprouver des sentiments positifs lors de déplacements en transports en commune, et dans une moindre mesure la perception que le temps de déplacement, en train ou en car, est un temps utile, diminuent la valeur du temps. Cependant, la simulation de scénarios montre que **des améliorations tangibles du confort (place assise garantie) génèrent un gain économique, par la**

diminution de la valeur du temps, plus important que des mesures plus « softs » visant à améliorer les perceptions latentes du confort.

Analyse de l'influence de la conscience environnementale sur les habitudes de choix modal

L'estimation de modèles d'équations structurelles permet d'approfondir l'idée selon laquelle se déplacer ne permet pas uniquement d'aller d'un point A à un point B mais remplit également d'autres fonctions qui varient selon le mode de transport. Tout comme les transports en commun peuvent générer une utilité indirecte au travers, par exemple, de l'utilisation efficace du temps de déplacement, la voiture peut remplir des fonctions affectives ou symboliques au travers, par exemple, de la vitesse de conduite ou de l'image véhiculée par la marque de la voiture conduite. L'estimation des modèles montre que **les personnes ayant une conscience environnementale élevée perçoivent l'usage des transports en commun comme plus facile, plus utile et plus plaisant** que les personnes qui n'ont pas cette motivation environnementale. De telles attitudes vis-à-vis des transports en commun **favorisent l'usage des transports en commun**. Au contraire, une **faible conscience environnementale génère des motivations affectives ou symboliques d'utilisation de la voiture**. Par contre, **la relation entre motivations non-instrumentales d'utilisation de la voiture et habitudes d'usage de la voiture n'est pas robuste**. En conclusion, la conscience environnementale influence les habitudes de choix modal et cette relation est médiée, partiellement, par les perceptions des transports en commun mais pas par les motivations d'utilisation de la voiture.

Analyse de l'hétérogénéité intra- et inter-individuelle dans la perception des probabilités de retard

Comme le confort et l'utilité indirecte de se déplacer, la fiabilité du temps de déplacement peut également être perçue de façon contrastée selon le mode de transport et selon les usagers. L'hypothèse selon laquelle les perceptions des probabilités de retard sont différentes selon le type d'usager (hétérogénéité inter-individuelle) et selon le mode de transport emprunté (hétérogénéité intra-individuelle) est testée en utilisant la théorie de l'utilité dépendante aux rangs. Les modèles estimés à l'aide d'un nested logit montrent que les **probabilités de retard de la voiture sont sous-évaluées** quelle que soit la fonction de transformation des probabilités utilisée. Les **probabilités de retard du train**, telles qu'elles sont observées en Région Rhône-Alpes, **sont pour leur part sur-évaluées**. Dans le modèle à classes latentes, ces résultats se vérifient également pour la classe à

laquelle les automobilistes ont une probabilité plus élevée d'appartenir que les usagers du train. Les **usagers du train** ont, quant à eux, une probabilité plus élevée d'appartenir à la classe qui suppose un **comportement neutre vis-à-vis du risque et pas de pondération des probabilités**. Ils auraient donc, en moyenne, une perception des probabilités de retard correspondant effectivement à ce qui leur est présenté. Par ailleurs, dans l'ensemble des modèles intégrant un coefficient d'aversion au risque, celui-ci est positif. La **fonction d'utilité** est par conséquent **convexe** et rend compte d'une **désutilité marginale du temps** et donc de valeurs du temps décroissantes avec le temps de déplacement.

Analyse des potentialités économiques et comportementales des modèles de référence

Les modèles mentionnés jusqu'à présent (logit multinomial, nested logit, modèle de choix hybride et modèle à classes latentes) ont été développés dans une perspective économétrique. Les **modèles de référence** sont une famille de modèles généralisant le logit multinomial et ont été développés dans une perspective statistique. Le chapitre qui y est consacré montre que les modèles de référence peuvent être utilisés pour calculer des outputs économiques, tels que des indicateurs de consentement à payer, les élasticités et les effets marginaux. Grâce aux modèles de référence, il est également possible de lever partiellement la propriété d'indépendance des alternatives non pertinentes, ce qui conduit à des élasticités croisées différenciées en réponse à une variation de niveau dans l'alternative de référence. L'application de cette famille de modèles à la question du choix modal montre que, exceptée la loi normale, les autres lois testées (Student, Cauchy, Laplace, Gumbel) **permettent d'obtenir une meilleure adéquation aux données que le logit multinomial**. Choisir le modèle de référence dont la loi permet la meilleure adéquation aux données a de fortes **implications**, notamment **en termes de sélection de variables et de calcul d'outputs économiques**. En particulier, grâce à ses queues de distribution épaisses, la **loi de Student** semble particulièrement bien capturer l'effet des variables explicatives sur le choix et **génère des outputs économiques**, et notamment un équivalent temps du confort, **cohérents même en présence d'une forte hétérogénéité inobservée**. Il semblerait également que, quand l'hétérogénéité inobservée est bien contrôlée, une distribution asymétrique telle que la loi de Gumbel offre les meilleurs résultats.

L'une des spécificités des modèles de référence est que les probabilités de choix dépendent de l'alternative choisie comme référence (par exemple, le train) sans pour autant que la relation d'ordre entre les alternatives ne soit modifiée. Afin de traduire les modèles de référence en termes économiques, il est nécessaire de définir une fonction d'**utilité dépendante à la référence**. Une telle

dépendance à la référence se justifie théoriquement et empiriquement et est notamment analysée en détails dans le modèle de dépendance à la référence de Tversky et Kahneman (1991). Graphiquement, les **lois de probabilité qui fournissent les meilleurs résultats empiriques sont aussi celles qui permettent de satisfaire certains principes du modèle de Tversky et Kahneman** : 1) dépendance à la référence ; 2) décroissance marginale de l'utilité dans les pertes et dans les gains ; 3) aversion aux pertes.

Implications opérationnelles et recommandations de politiques publiques

Le calcul des consentements à payer et équivalents temps pour l'amélioration de la qualité de service du train et du car (temps, cadencement, fréquence, retards) **peuvent servir à la réalisation d'analyses coûts-bénéfices**, pour évaluer des politiques ou de nouvelles infrastructures de transport, et à la réalisation d'**analyses coût-efficacité** pour comparer différents scénarios visant à augmenter l'attractivité de ces modes de transport. Les résultats soulignent le rôle du confort et de la fiabilité du temps de déplacement dans le choix modal et rappellent l'**importance d'améliorations tangibles de la qualité de service** (garantir une place assise, diminuer les temps et probabilités de retard).

L'ensemble des travaux de la thèse montre également l'intérêt d'incorporer des éléments de psychologie et d'économie comportementale dans les modèles de choix discret, en particulier en économie des transports. Les implications opérationnelles d'une telle approche se traduisent en termes d'**adaptation des modèles de prévision de la demande en transport et de politiques publiques**. Premièrement, comprendre l'importance du confort dans le choix modal peut aider à adapter les infrastructures ferroviaires, mais aussi à adapter les valeurs du temps prises en compte dans les modèles de prévision de la demande. Par ailleurs, utiliser les modèles de référence au lieu du modèle logit multinomial est un moyen d'obtenir des équivalents temps ou des consentements à payer, et donc des **fonctions de temps ou de coût généralisés, plus robustes** à intégrer dans les modèles de transport. De même, lorsque ces modèles intègrent une valeur de la fiabilité du temps de déplacement, prendre en compte des formes non linéaires, lorsqu'elles sont pertinentes, permettrait d'améliorer leurs performances. Une meilleure prévision de la demande permet à son tour de mieux adapter les investissement aux besoins et sert à l'évaluation de projets et de politiques publiques.

Deuxièmement, une meilleure compréhension des effets de la conscience environnementale ou des perceptions du temps et du confort par les modèles de choix hybrides, une meilleure intégration de

l'effet des probabilités de retard sur le choix modal par les modèles d'utilité dépendante au rang et une meilleure sélection des variables explicatives par les modèles de référence, sont toutes trois essentielles pour **identifier les leviers d'action des politiques publiques**. Par exemple, les résultats soulignent l'effet du confort dans les trains sur la valeur du temps et donc l'importance de prendre en compte cette dimension dans le cadre de la réflexion sur leur remplacement par des cars sur certaines lignes régionales. Des investissements visant à augmenter la capacité des trains, favoriser des conditions de travail efficaces (par exemple, par l'installation systématique de prises) ou encore permettre un certain isolement renforceront l'attractivité des lignes ferroviaires. Par ailleurs, le lien entre niveau de conscience environnementale et habitudes de choix modal, et la médiation de ce lien par les perceptions des transports en commun, peut inciter à développer des campagnes de communication présentant les transports en commun comme un symbole environnemental de façon à contrebalancer le symbole de statut social de la voiture (Beirão et Cabral, 2007).

Troisièmement, les enseignements de la thèse peuvent **contribuer à la mise en place de campagnes d'information et de communication** sur les différents modes afin d'encourager l'usage des transports en commun. Par exemple, communiquer sur les impacts environnementaux de la voiture, peut faciliter des changements de comportement ou, du moins, faciliter l'acceptabilité sociale de mesures coercitives. Lorsque l'offre de transports en commun est performante comparativement à la voiture, communiquer sur les temps de déplacement permettrait d'objectiver les caractéristiques de l'offre et d'informer les usagers potentiels.

Plus généralement, ce n'est qu'en comprenant mieux ce qui motive le choix d'un mode de transport et par quels mécanismes comportementaux ce choix est réalisé qu'il est possible de mieux prévoir l'effet des politiques publiques et d'induire des changements de comportement, par des politiques d'incitation économiques mais également par de nouveaux outils de politique publique qui influencent le choix des individus en rendant certaines caractéristiques de leur univers de choix plus saillantes ou en changeant leur option de référence (voir Sunstein et Reisch, 2013, pour des exemples). Ces outils sont désignés sous l'appellation « **nudge** », c'est-à-dire « n'importe quel aspect de la prise de décision qui peut affecter les décisions individuelles sans modifier les incitations économiques » (Croson et Treich, 2014, p.3). Par exemple, dans le domaine des transports, les modes de transport doux (transports en commun, vélo, marche) pourraient être ceux présentés par défaut dans les outils de calcul d'itinéraire afin que ces alternatives soient envisagées en premier par les usagers. En comprenant mieux la formation des choix par l'introduction de variables complémentaires, d'autres types de nudges, s'appuyant sur la comparaison sociale ou encore sur la perception du temps de

déplacement, peuvent être envisagés. Par exemple, puisque le temps de déplacement n'est pas seulement une question d'infrastructures et de rapidité, un nudge pourrait être de développer des outils pour aider les passagers à pratiquer des activités dans un environnement agréable, afin qu'ils investissent ce temps et le perçoivent le plus positivement possible. C'est dans cet optique que Watts et Lyons (2010) ont développé un « *travel remedy kit* » pour rendre les passagers actifs et engagés durant leur temps de déplacement.

Limites et recherches futures

Les limites des travaux présentés dans la thèse sont abordés dans cette section. Cinq séries de pistes de recherche sont proposées afin d'y répondre : 1) la nature des données de préférences déclarées ; 2) l'exploitation de l'enquête de préférences déclarées qui peut être approfondie ; 3) des propositions pour creuser certaines thématiques de l'enquête de préférences déclarées ; 4) la réplicabilité des résultats ; 5) des développements théoriques sur les modèles de référence et l'évaluation de la flexibilité sur le choix modal permise par les nouveaux systèmes d'information.

Adopter ou **s'inspirer des idées novatrices des théories comportementales n'est pas sans poser certaines difficultés** en lien les unes avec les autres. Premièrement, même si les implications opérationnelles de la thèse ont été évoquées, l'intégration simultanée de l'ensemble des éléments (variables latentes, probabilités non linéaires, dépendance à la référence) dans un modèle opérationnel de prévision de la demande semble pour l'instant irréaliste. Deuxièmement, augmenter le réalisme comportemental s'accompagne généralement d'une plus grande complexité économétrique et d'une augmentation du nombre de paramètres à estimer (par exemple, paramètres du modèles d'équations structurelles, coefficients d'aversion au risque et de pondération des probabilités) et donc d'un potentiel problème d'identification et un plus grand besoin de données. Troisièmement, les comportements de choix de déplacement sont éminemment complexes. Les théories développées en économie comportementale, qui reposent généralement sur des expériences contrôlées en laboratoire portant, par exemple, sur des choix entre des loteries ou sur des tâches mesurant des efforts réels, doivent donc être adaptées pour prendre en compte les spécificités propres au champ disciplinaire des transports. Quatrièmement, la nature des modèles comportementaux rend leur estimation plus aisée avec des données d'enquête de préférences déclarées qui ont tout de même leurs propres limites, dont la principale réside dans le caractère hypothétique du choix, malgré toutes les précautions qui ont été prises.

L'objet n'est pas ici de rentrer dans les détails de ces limites, abordées par exemple par Louviere et al. (2000), mais d'**envisager comment le choix modal pourrait être modélisé à l'aide de données alternatives aux enquêtes de préférences déclarées**, ce qui conduit à une **première série de pistes de recherche**. Premièrement, les données de l'enquête de préférences déclarées pourraient être combinées aux données de l'enquête déplacement régionale (données de préférences révélées), ce qui n'a pas pu être réalisé durant la thèse car le réseau de transports en commun n'était pas codé. Cette approche combinatoire, largement abordée dans la littérature (Hensher et al., 1999 ; Bhat et Castelar, 2002), permettrait de bénéficier des forces de chacune des deux méthodes d'enquête et notamment de s'appuyer sur des données de choix réels et de calibrer les constantes modales, tout en bénéficiant de la mesure des variables latentes. Une extension envisagée est de modéliser des choix effectués simultanément (choix modal, choix de motorisation, abonnement) avec une méthode de type *Generalized Heterogeneous Data Model* (Bhat, 2015) [6].

Deuxièmement, les déterminants du choix modal pourraient être analysés à l'aide d'une expérience de terrain dans laquelle les réactions d'un échantillon suite à la mise en place d'interventions (prêt de vélo, incitations financières, campagne d'informations...) est suivi. Par exemple, Abou-Zeid et Ben-Akiva (2012) rendent compte de deux expériences de ce type, l'une en Suisse et l'autre aux États-Unis (MIT), consistant à inciter des usagers de la voiture à utiliser temporairement les transports en commun [7]. Un tel travail pourrait également permettre d'analyser les internalités induites par le choix de la voiture, internalités qui n'ont pas du tout été abordées dans la thèse [8].

Troisièmement, une expérience incitée en laboratoire permettrait d'analyser plus précisément l'arbitrage entre temps et coût de déplacement en situation de risque. Ce type d'expérience, encore rarement réalisé en économie des transports, permet de se concentrer sur des dimensions spécifiques du choix, en contrôlant précisément les conditions dans lesquelles il est réalisé, ce qui leur permet d'avoir une forte validité interne. Quelques exemples d'application incluent Gaker et al. (2010) qui ont réalisé trois expériences portant sur l'apport d'information et les influences sociales [9], Denant-Boèmont et Petiot (2003) et Denant-Boèmont et Hammiche (2010) qui ont réalisé des expériences en laboratoire pour calculer la valeur de l'information sur le temps de déplacement en situation de congestion avec

6. Un partenariat avec un doctorant de l'université du Texas à Austin a été développé dans cet objectif.

7. Après l'intervention, aucun des Suisses n'adopte définitivement les transports en commun, 30% des Américains, en revanche, changent de mode de transport après l'intervention.

8. Une expérience de terrain sera mise en œuvre dans le cadre du contrat post-doctoral prévu à l'issue de la thèse. Ce projet, en collaboration avec des épidémiologistes et des spécialistes du comportement dans le domaine du sport devrait permettre d'aborder la question des impacts sanitaires du choix modal et ainsi d'analyser comment les décisions sous-optimales peuvent être corrigées.

9. La première expérience porte sur l'effet d'un apport d'informations environnementales personnalisées sur le choix d'un itinéraire. La deuxième porte sur l'effet des influences sociales sur le choix d'achat d'une voiture. La troisième combine informations sur la loi et la sécurité et influences sociales pour analyser leurs effets respectifs sur le comportement des piétons.

choix entre deux modes (train et voiture). S'appuyer sur de telles expériences de terrain et en laboratoire permettrait de contextualiser les enseignements de l'économie comportementale et ainsi répondre au besoin d'adaptation au contexte particulier du choix modal identifié ci-dessus.

Une deuxième série de pistes de recherche permettrait d'approfondir l'analyse des données de l'enquête de préférences déclarées. Premièrement, le modèle de médiation utilisé pour expliquer les habitudes de choix modal pourrait être inclus dans un modèle de choix hybride, afin d'analyser le rôle des habitudes sur le choix modal. Une méthode d'estimation non dépendante du nombre de variables latentes telle que celle proposée par Bhat et Dubey (2014) devra être mise en œuvre pour y parvenir.

Deuxièmement, l'analyse du choix modal avec des temps de déplacement incertains montre une fonction d'utilité convexe. Dans le cadre d'un modèle d'utilité dépendante au rang, ce résultat peut traduire une recherche de risque, si la personne est optimiste par rapport aux probabilités de pertes, ou une aversion faible au risque, si la personne est suffisamment pessimiste par rapport aux probabilités de perte. La courbure de la fonction d'utilité doit donc être comparée à celle de la fonction de probabilités, ce qui peut être évalué à l'aide de la méthode proposée par Chateauneuf et Cohen (1994) et Cohen (2015). L'application empirique nécessite encore d'être approfondie avant de conclure. De plus, l'estimation des mêmes modèles (modèle à classes latentes et nested logit) est en cours avec des méthodes de statistiques bayésiennes [10]. Les facteurs explicatifs de l'attitude par rapport au risque pourraient également être enrichis en utilisant des variables latentes telles que les sentiments éprouvés dans les transports en commun.

Troisièmement, parmi les modèles de référence, tester les performances d'une loi asymétrique et dont l'épaisseur des queues de distribution peut être contrôlée, telle que la loi de Student décentrée, semble prometteur. Par ailleurs, l'une des difficultés majeures avec les modèles de référence est de connaître l'alternative choisie comme référence par les personnes interrogées. Cette difficulté peut être levée avec un modèle combinant modèles de référence et classes latentes, dans lequel l'appartenance (probabiliste) à une classe définie conditionne l'alternative choisie comme référence. L'estimation d'un tel modèle, ainsi que celle d'un modèle de référence avec une loi de Student décentrée, devrait être prochainement réalisée.

Quatrièmement, l'analyse des données de l'enquête de préférences déclarées a montré qu'un certain nombre de personnes tendent à rester captives de leur mode de transport actuel. Une analyse plus détaillée des résultats permettrait de comprendre si cette captivité s'explique par des contraintes

10. En collaboration avec Matthieu de Lapparent (HEIG-VD, Suisse).

liées par exemple à des critères géographiques, à une congestion importante en voiture ou à une offre de transport en commun insuffisante qui rend les alternatives présentées peu réalistes [11] ou s'explique par des perceptions et attitudes négatives (ou positives) vis-à-vis des transports en commun ou de la voiture ou encore un refus ou une incapacité à répondre à l'enquête. La combinaison d'un modèle analysant les raisons qui poussent à choisir systématiquement l'alternative de référence et d'un modèle de choix discret pourrait appuyer un telle analyse.

Une troisième série de pistes de recherche serait de s'appuyer sur des données complémentaires à celle de l'enquête de préférences déclarées afin de mettre en œuvre des recommandations de la thèse ou d'approfondir l'interprétation de certains de ses résultats. Premièrement, les échéances courtes de réalisation de l'enquête de préférences déclarées et des contraintes de temps de passation du questionnaire n'ont pas permis d'introduire l'ensemble des indicateurs permettant d'estimer un modèle de choix hybride intégrant une théorie de psychologie comportementale. Construire une nouvelle enquête [12], en collaboration avec des psychologues, permettrait l'estimation d'un tel modèle de choix qui pourrait par exemple s'appuyer sur la théorie du comportement interpersonnel, le modèle transthéorique du changement (Prochaska et DiClemente, 1986) traduisant les différentes étapes nécessaires à un changement de comportement ou encore un modèle faisant intervenir à la fois des normes sociales et des facteurs externes, tel que celui proposé par Collins et Chambers (2005). Les normes sociales, et plus généralement, l'importance du groupe social dans la prise de décision ou encore la volonté de participer au bien-être collectif, ont encore été relativement peu étudiés en transport, et mériteraient un développement dans de futures recherches. En effet, l'un des apports essentiels de l'économie comportementale est l'importance du groupe social dans la prise de décision.

Deuxièmement, les contraintes de temps de passation du questionnaire et la volonté d'intégrer un grand nombre d'attributs n'ont permis de mesurer le confort « objectif » qu'au travers de la garantie ou non d'avoir une place assise. La dimension multiparamétrique du confort (place assise, confort des sièges, propreté, place disponible...) pourrait donc être prise en compte dans une enquête dédiée afin de mieux rendre compte de l'influence de ces différentes composantes sur le choix modal.

Une quatrième série de pistes de recherche concerne la réplicabilité des résultats. Différentes pistes peuvent en effet être explorées afin de tester dans quelle mesure les résultats de la thèse sont

11. Malgré la personnalisation des questions de choix, il est possible que sur certains couples origine-destination à des horaires spécifiques, le train ou la voiture soient peu réalistes.
12. Dans le cadre du contrat postdoctoral, une réponse à un appel d'offres est en cours pour financer une enquête de ce type.

généralisables. Premièrement, il faudrait analyser, à l'aide d'autres bases de données, si l'attitude face au risque et les perceptions des probabilités de retard différenciées selon le mode de transport sont robustes [13].

Deuxièmement, l'estimation de modèles et le calcul d'outputs économiques avec les modèles de référence devront également être effectués avec des données provenant de thématiques (énergie, alimentation...) et de type d'enquêtes (préférences déclarées, révélées) variées afin de tester dans quelle mesure la valeur ajoutée de ces modèles est généralisable.

Une cinquième série de pistes de recherche s'appuie sur des travaux théoriques sur les modèles de référence et un modèle d'options réelles. Premièrement, la connexion entre modèles de référence et théorie économique n'en est qu'à ses débuts et nécessitera une formalisation mathématique plus poussée, notamment pour définir une utilité dépendante à la référence et faire le lien avec la théorie de l'utilité aléatoire [14]. La théorie comportementale sur la dépendance à la référence et l'aversion au risque devra être mobilisée (Bateman et al., 1997 ; Munro et Sugden, 2003 ; Köszegi et Rabin, 2006 ; Gill et Prowse, 2012).

Deuxièmement, les entretiens réalisés durant la phase qualitative montrent qu'avec les nouveaux systèmes d'information (smartphones, sites de covoiturage...), les individus deviennent plus flexibles sur leur mode de déplacement. Un déplacement commencé en train peut par exemple se terminer en voiture grâce au covoiturage. La valeur de cette flexibilité peut être mesurée à l'aide d'un modèle d'options réelles. Ce modèle théorique a été construit [15] mais comporte encore un certain nombre de limites sur le plan mathématique. L'application numérique sera effectuée grâce aux valeurs du temps de déplacement et valeur du temps de retard calculées avec l'enquête de préférences déclarées.

13. Une base de données récemment fournie par la SNCF pourrait être utilisée. Cependant, les données portent sur des choix entre alternatives train et ne permettront pas de différencier les formes des fonctions de probabilités selon le mode de transport.
14. En collaboration avec Jean Peyhardi (Université de Montpellier).
15. En collaboration avec Oana Inoescu (GAEL, Grenoble).

Appendices

APPENDIX A

Théories de psychologie environnementale (Introduction)

De très nombreux modèles de psychologie environnementale existent. Seuls les principaux modèles utilisés pour analyser les comportements pro-environnementaux sont discutées ici. Parmi ces modèles, il est possible de distinguer ceux motivés par des intérêts personnels et se basant sur un processus décisionnel rationnel de ceux qui s'appuient sur les valeurs et les normes ou les habitudes pour expliquer les choix.

1. Parmi les **modèles « rationnels »**, la théorie du comportement planifié (*Theory of planned behavior*; Ajzen, 1991) identifie trois facteurs influençant les comportements : les attitudes [1] ; les normes subjectives, i.e. la perception de la pression sociale et la motivation à s'y plier et enfin la perception de contrôle (*perceived behavior control*) sur le comportement, i.e. la facilité ou non à mettre en œuvre le comportement. Ces trois facteurs jouent sur l'intention comportementale qui détermine le comportement proprement dit. L'intention est l'expression de la motivation à adopter le comportement Cette théorie fait donc l'hypothèse que les comportements s'expliquent uniquement par des facteurs mentaux propres à l'individu (internes) plutôt que des facteurs contextuels. Cependant, comparativement à la théorie de l'action raisonnée (*Theory of reasonned action* ; Fishbein et Ajzen, 1975 sur laquelle elle s'appuie, la théorie du comportement planifié prend en compte les contraintes telles qu'elles sont perçues par l'individu, à défaut d'une intégration explicite, grâce à la notion de perception de contrôle. Le modèle ABC (*Attitude-Behavior-Context model* ; Guagnano et al., 1995), modélise les comportements comme résultant de l'interaction entre facteurs contextuels (incitation financière, contrainte de l'offre..) et facteurs internes (attitudes, normes, croyances...). La relation

1. En psychologie, les attitudes se définissent comme la disposition à réagir de façon favorable ou défavorable à un objet particulier ou à une classe d'objets (Oskamp, 1977). Elles portent sur les croyances sur les conséquences du comportement et l'évaluation de ces conséquences.

de causalité entre les attitudes et les comportements peut être influencée par la force des facteurs contextuels facilitant ou empêchant l'adoption du comportement.

2. Parmi les **modèles s'appuyant sur des notions de normes et de valeurs**, la théorie VBN (*Value-Belief-Norm theory* ; Stern et al., 1999) a spécifiquement été développée pour expliquer des comportements pro-environnementaux. Elle se base sur la théorie de la norme altruiste (ou *Norm-activation theory* ; Schwartz, 1977) selon laquelle la réalisation d'un comportement résulte d'un sentiment d'obligation morale, qui lui-même s'explique par une prise de conscience des problèmes environnementaux et un sentiment de responsabilité vis-à-vis de ces problèmes. La VBN y ajoute la notion de croyances environnementales qui peuvent être mesurées sur l'échelle du « *New Environmental Paradigm Scale* » (Dunlap et al., 2000). Ces croyances environnementales sont elles mêmes déterminées par des valeurs personnelles. Comme la théorie du comportement planifié, la théorie VBN ne fait pas intervenir de facteurs contextuels.
3. Contrairement aux autres théories, la théorie des comportements interpersonnels (*Theory of interpersonal behavior* ; Triandis, 1977) fait intervenir la **notion d'habitudes**. Le comportement résulte de la conjonction de trois facteurs, l'intention d'adopter le comportement, l'habitude et les facteurs contextuels. Les habitudes sont le degré d'automatisme d'un comportement dans une situation donnée. Pour les choix répétés, les habitudes amènent à la reproduction quasi-inconsciente des mêmes gestes qui fonctionnent comme une « évidence intériorisée » constituant un fort facteur d'inertie (Vincent, 2008). Les cadres théoriques utilisés en économie ou en psychologie qui envisagent les choix comme étant raisonnés se révèlent dès lors insuffisants puisque le choix devient inconscient, sans intention. Ainsi, les habitudes diminuent l'importance des autres déterminants du choix.

Au sein de ces théories, quelles sont celles à privilégier pour expliquer les comportements pro-environnementaux ? Il semblerait que la théorie du comportement planifié et la théorie VBN - ou sa version antérieure, la théorie de la norme altruiste - soient les plus utilisées. Elles ont été comparées afin d'établir leur capacité explicative des comportements pro-environnementaux. Selon Kaiser et al. (2005), les composantes de la théorie du comportement planifiée (attitude, norme subjective, contrôle perçu et intention) expliquent mieux les comportements de conservation environnementale (économies d'énergie, usage de la voiture...) que celles de la théorie VBN (normes personnelles, prise de conscience, sentiment de responsabilité et croyances environnementales). La première est également meilleure pour décrire les relations structurelles entre ses composantes. Gifford et al. (2011) constatent que le choix modal implique des coûts comportementaux élevés et requiert une action plus

engageant pour l'individu que d'autres comportements pro-environnementaux (recyclage, économies d'eau et d'énergie). Dans ces conditions, la théorie du comportement planifié est plus performante. Ce résultat s'explique notamment par l'absence totale d'interactions avec les conditions externes dans la théorie VBN, alors que la théorie du comportement planifié les prend partiellement en compte au travers du contrôle perçu.

APPENDIX B

Analyse qualitative : mise en œuvre, supports et résultats (Ch.2)

B.1 Mise en œuvre opérationnelle

Pour les entretiens et les focus groups, les personnes ont été recrutées par trois canaux :

- un questionnaire en ligne qui proposait de s'inscrire pour réaliser un entretien individuel ou un focus group.
- des entretiens directs dans le train sur la ligne Lyon-Grenoble.
- une société de service (Krief études) pour compléter les focus groups.

Deux focus groups de 2h30 ont été réalisés, avec chacun 13 participants. Chaque focus group comprenait 1) des automobilistes réguliers et des usagers réguliers du train ; 2) des personnes habitant en milieu urbain et des personnes habitant en milieu rural. Les participants devaient être suffisamment différenciés selon des critères de sexe, d'âge, situation familiale et de niveau d'études. Il avait également été demandé qu'ils n'aient pas de difficulté d'expression à l'oral.

Le premier focus group a eu lieu à Grenoble (INPG-GI), le 22 avril 2014. Le deuxième focus group a eu lieu à Vaulx-en-Velin (ENTPE), le 22 mai 2014[1]. Les 11 entretiens ont eu lieu sur la ligne Lyon-Grenoble ou à domicile, entre le 25 mars et le 05 mai 2014.

Chaque focus group et chaque entretien a été retranscrit (retranscriptions disponibles sur demande) ce qui a permis la restitution « brute » des propos recueillis. Une synthèse analytique permet de tirer les enseignements de cette phase qualitative afin d'adapter le questionnaire de l'EPD.

Lors des entretiens, des focus groups et des tests de questionnaires qui ont suivis, différents formats des questions de choix, explications et supports visuels pour les attributs ont été testés. L'ensemble des supports testés est détaillé dans le rapport remis à la Région Rhône-Alpes.

1. L'animation des focus groups a été réalisé avec l'aide de Gabrielle Deyris (stagiaire) et Stéphanie Vincent (chercheuse au LAET).

B.2 Guide d'entretiens individuels

La grille d'entretiens individuels est structurée en six temps.

B.2.1 Temps 1 – Echange libre sur les déplacements et les raisons qui poussent à choisir un mode de transport

Cette étape permet de travailler sur les raisons du choix modal telles qu'elles sont spontanément décrites par le participant.

1. Pouvez-vous me décrire les déplacements que vous avez effectués cette semaine que ce soit pour le travail, pour les loisirs, pour la famille, ou pour tout autre motif ?

   Les questions suivantes ne seront posées que si la personne interrogée n'aborde pas d'elle-même ces points.
2. Pouvez-vous me décrire votre déplacement domicile-travail en particulier ?
3. Est-ce un mode de déplacement régulier ? Si non, quel est votre mode de déplacement régulier ? Si oui, depuis quand est-il régulier ? Quel était votre mode de déplacement auparavant ? Pourquoi avez-vous changé ?
4. Qu'avez-vous fait ce week-end ? Comment vous êtes-vous déplacé ? Avez-vous conduit ou pris le train ?
5. Quels sont selon vous les avantages et inconvénients du (des) mode(s) de déplacement que vous utilisez ?
6. Pouvez-vous me décrire les raisons qui vous poussent à choisir ce mode de déplacement ?
7. Avez-vous un abonnement de train, de TCU ou de péage ?

   Questions pour les personnes qui ne mentionnent pas le train
8. Avez-vous déjà pris le train ? Si oui, pour quels types de trajets prenez-vous le train ? Quels sont les atouts du train selon vous ? Si non, pourquoi ne prenez-vous pas le train ? Quels sont les limites du train selon vous ?

B.2.2 Temps 2 – Travail sur un trajet particulier et le choix modal sur ce trajet

Vous m'avez indiqué avoir réalisé un trajet entre … et …. . Pour cela, vous avez choisi d'utiliser votre voiture / le train / le bus ou le tram.

1. Pouvez-vous dessiner votre parcours sur cette feuille ? Voulez-vous y ajouter des détails ?
2. Combien payez-vous votre trajet ? Y a-t-il une prise en charge par un tiers ?

3. A quelle fréquence faites-vous ce trajet ?
4. Pendant ce déplacement, quelles activités particulières avez-vous eu ? (par exemple : lire, téléphoner, discuter)
5. En général, sur ce trajet, voyagez-vous seul ou accompagné(e) ?
6. Comment qualifieriez-vous le temps de trajet ? Vous a-t-il paru :
7. A quelles conditions est-ce que vous utiliseriez un autre moyen de transport ? Pourquoi ne prenez-vous pas le train, bus, tram, vélo pour ce trajet ? Avez-vous déjà essayé de prendre un autre mode de transport ? Que faudrait-il modifier pour que vous preniez le train ou un autre mode de transport ? A quelles conditions prendriez-vous un autre mode de transport ?
8. S'il n'était plus possible de venir en voiture/train/bus, quelle attitude adopteriez-vous ? Abandon du trajet ? Changement de trajet ? Autre mode de transport ?
9. Pour faire ce trajet, vous pourriez également prendre le train/le bus/la voiture. Pouvez-vous m'indiquer pourquoi est-ce que vous ne le prenez pas ? Si énonce des raisons qui peuvent être contrées, essayer de les contrer et de voir si d'autres raisons émergent.
10. A destination, avez-vous des contraintes horaires ? Pour le travail, est-ce que vous pointez ? Avez-vous la possibilité de faire du télé-travail ? Avez-vous des contraintes horaires le soir (enfants, loisirs, association) ?
11. Devez-vous transporter des choses quand vous allez au travail ?
12. Pour ce trajet précis, pouvez-vous me préciser les raisons qui vous ont poussé à utiliser votre voiture / le train / le bus ou le tram.

B.2.3 Temps 3 – Travail sur des adjectifs, descriptions

1. Pouvez-vous me citer trois adjectifs qui vous paraissent les plus adaptés pour qualifier ces modes de transport ? *Si répond « pratique » ou « pas pratique », demandez de préciser.*

   (a) Train
   (b) Car
   (c) Voiture
   (d) Vélo

2. Qu'évoquent pour vous les mots :

   (a) Fiabilité ? Y a-t-il des retards acceptables et d'autres qui le sont moins ?
   (b) Cadencement ?
   (c) Coût d'un trajet ? En voiture, en train, en car ?
   (d) Temps de trajet ?
   (e) Fréquence de passage des trains ?

3. Pouvez-vous décrire un train idéal ? Pouvez-vous me décrire le train idéal sur les caractéristiques suivantes ?

   (a) Temps de trajet
   (b) Equipement du train (prise)
   (c) Fréquence de passage des trains
   (d) Amplitude horaire
   (e) Fiabilité
   (f) Trains passant à heure fixe
   (g) Propreté du train
   (h) Prix du trajet

B.2.4 Temps4 – Travail sur des échelles et choix d'attributs

1. Grille d'interprétation de Kelly (voir figure B.1, page 306) : pouvez-vous m'indiquer quels sont les deux modes que vous mettriez ensemble entre le train, le tram et le bus *(ligne 1)* et quel mode vous placeriez à part ? Pour quelle(s) raison(s) ?
2. Classement : Pour un trajet en train, pouvez-vous m'indiquer ce qui est le plus important pour vous entre ... et ... *(proposition de deux caractéristiques)* ?

   (a) Un temps plus court dans le train
   (b) Une prise dans le train pour brancher mon ordinateur
   (c) Des trains passant à heures fixes
   (d) La garantie d'avoir une place assise
   (e) Plus de facilités pour se garer
   (f) Des trains plus propres
   (g) Une gare plus accessible en bus
   (h) Une fréquence plus élevée des trains
   (i) Une plus grande amplitude horaire
   (j) Une meilleure fiabilité du temps de parcours
   (k) Un arrêt plus proche de chez moi
   (l) De meilleures correspondances
   (m) Un prix moins cher
   (n) La possibilité de monter son vélo dans le train

   Une première version de cette question demandait aux personnes interrogées de noter l'importance de chacune de ces caractéristiques sur une échelle de 1 à 5 selon leur importance. Cependant, les caractéristiques n'étaient alors pas discriminées puisque toutes étaient considérées comme importantes.

3. Pouvez-vous classer ces différentes caractéristiques du train par ordre d'importance pour vous ?

   (a) Temps de trajet
   (b) Equipement du train (prise)
   (c) Fréquence de passage des trains
   (d) Amplitude horaire
   (e) Fiabilité
   (f) Trains passant à heure fixe
   (g) Propreté du train
   (h) Prix du trajet

B.2.5 Temps 5 – Questions socio-économiques

1. Quel est votre âge ?
2. Etes-vous marié ? Avez-vous des enfants ? Quel âge ont-ils ?
3. Dans quelle commune résidez-vous ? Quel est l'arrêt de transports en commun le plus proche de chez vous ?
4. Quel est votre niveau d'études ? Dans quel secteur travaillez-vous ?

B.2.6 Temps 6 – Test de présentation des attributs

Pour les attributs fiabilité, fréquence et cadencement, différentes représentation des questions de choix et plusieurs représentations graphiques / textes sont testés afin d'adapter les formulations et supports du questionnaire. Voir détails dans le rapport remis à la Région Rhône-Alpes.

B.3 Guide d'animation des focus groups

Le guide d'animation des focus groups est structuré en six temps.

B.3.1 Temps 1 : Prise de connaissance (10 minutes)

Principe : l'objectif de ce temps est de faire connaissance, de briser la glace en bougeant et en obligeant les participants à discuter entre eux. Ce temps est également l'occasion de savoir dans quelle mesure les participants connaissent le kilométrage de leur voiture principale. En préambule, un ppt permet de présenter les objectifs généraux et les règles du jeu. En revanche, les objectifs précis (méthode des choix expérimentaux, déterminants du choix modal, modèle régional) ne sont présentés qu'à la fin afin de ne pas influencer les réponses.

Déroulement :

1. Les participants se lèvent et se « classent » par ordre alphabétique de prénom (ils doivent donc discuter entre eux). Une fois l'ordre établi, chaque se présente à l'aide de son prénom et de sa ville de résidence ;
2. Dans un deuxième temps, les participants se « classent » par kilométrage annuel de leur voiture principale, puis chacun annonce le nombre de kilomètre s;
3. Le ppt avec les objectifs et règles du jeu est présenté.

B.3.2 Temps 2 : Associations spontanées autour de mots liés aux « transports » (30 minutes)

Principe : plusieurs mots sont proposés successivement aux personnes présentes. A partir de chacun des mots, chaque participant cite toutes les idées qu'il y associe spontanément. Il les inscrit sur des cartons qui sont distribués (une idée-un carton). Après 2 minutes de réflexion, chacun lance des idées et justifie ses choix (s'il le souhaite). Les idées sont inscrites au fur et à mesure sur un paper-board de la façon la plus structurée possible afin de dessiner une carte mentale associée à ce mot.

Mots proposés (dans l'ordre, selon le temps disponible) :

1. Déplacement (association de mots)
2. Transport idéal (Sur votre planète idéale, comment se déplace-t-on ? Si la personne peine : imaginez une planète idéale à tous points de vue. Comment vous déplaceriez vous ?)
3. Train/voiture (portrait chinois)

   (a) Si le train/la voiture était un animal, ce serait … ?
   (b) Couleur
   (c) Objet du quotidien
   (d) Ville
   (e) Epoque
   (f) Personnage célèbre
   (g) Territoire

4. Train idéal (association de mots – préciser qu'on parle du matériel, du service, mais aussi de la gare ou même de l'accès à la gare)

Mots complémentaires si le temps restant le permet :

1. Fréquence
2. Cadencement
3. Fiabilité
4. Confort
5. Temps de trajet
6. Coût du trajet
7. Information

B.3.3 Temps 3 : Mise en situation sur un déplacement (50 minutes)

Principe : Faire dessiner les personnes présentes et les faire associer des mots à leur déplacement en train ou déplacement domicile-travail, ainsi qu'à leur mode de déplacement. La dimension personnelle du dessin rend l'analyse de la représentation du déplacement plus riche. Elle permet d'examiner comment les personnes représentent leur trajet, les modes utilisés, les changements

éventuels, les dimensions sur lesquelles elles insistent, etc. Elle permet par ailleurs d'avoir un support pour situer les personnes rapidement.

Déroulement :

Dans un premier temps, chaque participant dessine un déplacement sur une feuille blanche (puis son mode de déplacement, dans un deuxième temps). Si les personnes peinent à dessiner, il peut leur être précisé de représenter le lieu de départ, le lieu d'arrivée et l'itinéraire. Ils doivent dessiner un déplacement datant de il y a moins d'un mois en région Rhône-Alpes et correspondant aux caractéristiques suivantes (dans l'ordre décroissant de priorité) :

1. en train ;
2. plus de 20 km ;
3. qui aurait pu être fait en train ;
4. déplacement domicile-travail.

Dans un deuxième temps, chaque personne écrit sur des cartons des mots associés au déplacement (resp. au mode de déplacement). Enfin, chaque personne affiche ses deux dessins et les décrit, ce qui permet d'analyser les mots associés aux représentations visuelles.

Pause : 15 minutes

B.3.4 Temps 4 : Mise en situation - changer de mode de transport (15 minutes)

Ce temps n'a été mis en pratique que pour le groupe 1.

Principe : A partir des dessins précédents, deux cercles sont tracés sur le paper-board correspondant aux différents usages. Par exemple, les automobilistes / cyclistes d'une part et les usagers du train / TCU d'autres part. Deux groupes physiques sont formés. Chaque groupe doit se mettre d'accord et dire à quelles conditions il changerait de mode de transport. Le fait de faire discuter les personnes à l'intérieur d'un groupe permet d'avoir des échanges sur ce qui est le plus important.

Déroulement :

1. Chaque groupe choisit et inscrit sur des cartons des raisons qui le pousserait à choisir l'autre mode de transport ;
2. Quelles sont les améliorations qu'il faudrait apporter au train / à la voiture pour que vous choisissiez ce mode de transport (3 réponses possibles) ? Les réponses doivent être qualitative (nature de la raison) et quantitatives (par exemple, si la fréquence de passage des trains doit être modifiée, quelle fréquence serait souhaitable ?) ;

3. A quel point faudrait-il que le service en train ou en voiture soit dégradé pour que vous ne choisissiez plus ce mode de transport (3 réponses possibles).

B.3.5 Temps 5 : Echelles de positionnement

Uniquement si le temps le permet.

Principe : La recherche en psychologie a permis de développer différentes échelles de valeur qui permettent d'expliquer le choix modal. L'idée est de tester ces échelles en prévision de l'EPD, à la fois dans leur compréhension mais aussi dans leur adaptation au contexte régional (les échelles n'ont pas été développées en France). Il n'est pas possible de tester toutes les échelles, le choix s'est donc porté sur celles qui permettent une discussion sur la raison poussant la personne à se positionner à tel ou tel endroit.

Déroulement : L'échelle est matérialisée au sol. Chaque personne se place là où elle le désire selon qu'elle est plus ou moins d'accord avec les affirmations suivantes :

- Ma voiture me permet d'aller là où je veux.
- Je suis en sécurité dans ma voiture.
- Je trouve qu'il est confortable de conduire.
- Ma voiture me donne une sensation de liberté.

B.3.6 Temps 6 : Questions de choix discret (30 minutes)

Des questions de choix (10 au total), présentées dans différents formats (voir les exemples du guide d'entretien individuel) sont imprimées et présentées aux participants. Pour chaque attribut (fiabilité, fréquence, cadencement), les participants doivent indiquer le choix qu'ils effectueraient, ainsi que le format qui leur semble le plus clair et le plus adapté. Dans un premier temps, chacun répond individuellement sur sa fiche puis va coller une gommette sur la fiche correspondant à ses préférences.

B.4 Exemples de résultats

Cette section résume brièvement les résultats de deux questions de l'analyse qualitative[2].

2. Les résultats complets des entretiens et focus groups, et en particulier, l'analyse des discussions sur les déclencheurs qui inciteraient à changer de mode de transport pour les automobilistes et les usagers du train, sont dans le rapport remis à la Région Rhône-Alpes.

B.4.1 Grille d'interprétation de Kelly

A l'aide de la grille d'interprétation de Kelly, il est possible de déterminer quels sont les critères de différenciation entre modes qui viennent spontanément à l'esprit des usagers lorsque trois modes de transport leur sont proposés (voir figure B.1a). Pour les modes de transport analysés (train, tram, bus, voiture, auto-partage, co-voiturage, vélo et marche à pied), la différenciation principale se fait en termes de distance, d'échelle géographique (voir figure B.1b). Viennent ensuite la motorisation du véhicule, son côté écologique, le fait qu'il soit individuel ou non et sa vitesse.

1	Train	Tram	Bus
2	Tram	Voiture	Vélo
3	Auto-partage	Vélo	Train
4	Vélo	Co-voiturage	Marche à pied
5	Voiture	Train	Co-voiturage
6	Bus	Marche à pied	Auto-partage
7	Marche à pied	Train	Voiture

(a) Triplets proposés

	Fréquence
Urbain / Interurbain Trajets longs / courts	8
Motorisé / non motorisé (effort physique)	7
Ecologique / non écologique	6
Individuel / Collectif	5
Vitesse, rythme de voyage	5
Rail / Route / Autre	3
Voiture / pas de voiture	3
Soumis ou non au trafic	2
Multimodal	2
Coût	2
Sécurisant	1
Motif (occasionnel, travail)	1
Bruyant	1
Vol	1

(b) Critères de différenciation

FIGURE B.1 – Critères de différenciation entre modes à l'aide de la grille de Kelly

Notes : La figure B.1b se lit de la façon suivante. Par exemple, huit personnes parmi les 11 interrogées différencient les modes de transport qui leurs sont présentés sur un critère géographique (urbain/interurbain ; trajets longs/courts).

B.4.2 Portrait chinois

L'exercice du portrait chinois appliqué au train permet de rendre transparente la perception que le grand public a du train. Peu de réponses permettent d'identifier explicitement des déterminants du choix modal. Il s'agit davantage d'une interprétation des réponses. En quoi la perception, l'image que le grand public a du train expliquent-elles le choix modal ? En employant différents vecteurs de représentation mentale (animal, couleur, ville, territoire, époque), trois sens humains sont traduits (vue, toucher, ouïe) ainsi qu'une représentation dans l'espace (ville, territoire) et dans le temps (époque), permettant ainsi une diversité de réponses. Par exemple, le train, quand il est comparé à un animal, évoque le mouvement et l'intensité de ce mouvement. Ainsi, les participants mentionnent parfois la vitesse (lion, guépard, lévrier), d'autres fois la lenteur (escargot, tortue). Les retards sont également mis en avant (par le truchement de l'âne) ainsi qu'une certaine inertie (éléphant).

APPENDIX C

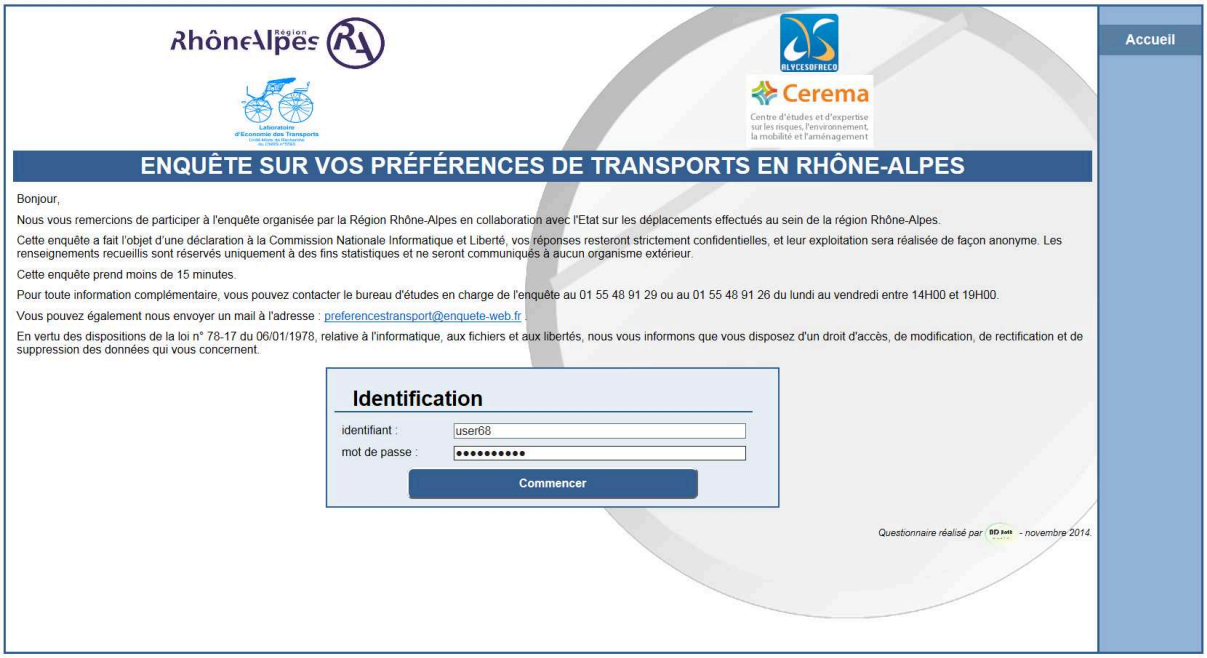
Déroulement et questionnaires de l'enquête de préférences déclarées (Ch.2)

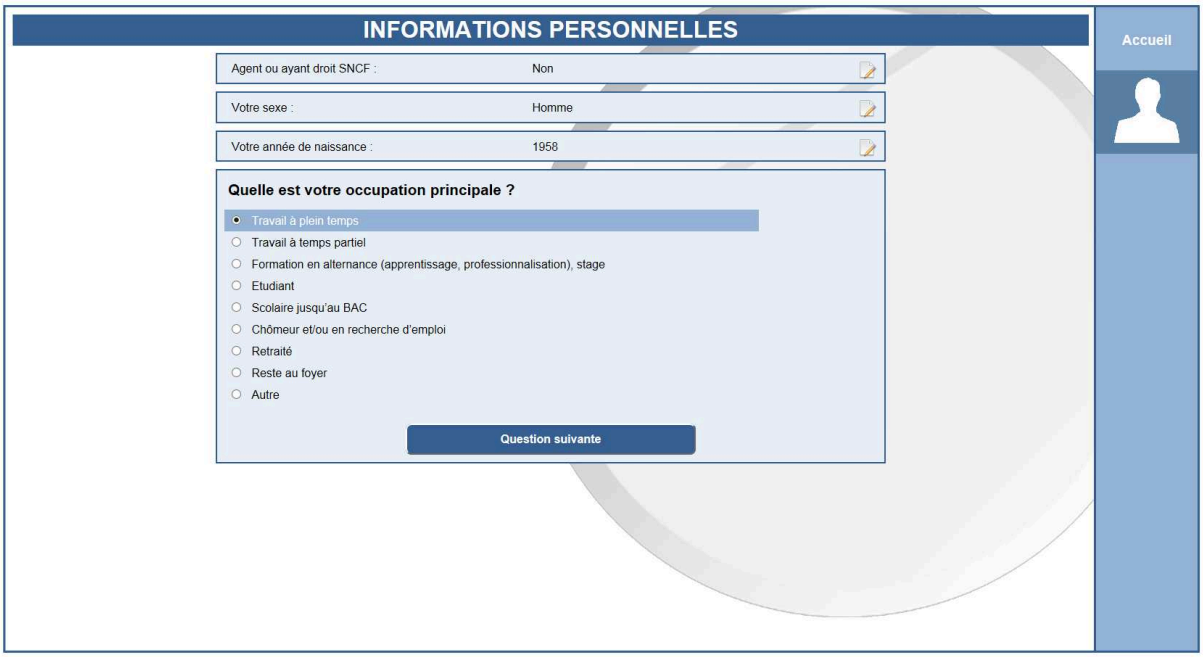
Cette annexe comprend les deux questionnaires : CAWI (section C.1) et CAPI (section C.2). La passation des questionnaires a été réalisée par l'institut de sondage AlyceSofreco. Pour le CAWI, les enquêtes ont eu lieu entre le 19 janvier et le 26 avril 2015. Pour le CAPI, elles ont eu lieu entre 19 janvier et le 09 avril 2015. Le CAPI a été réalisé en utilisant un PDA [1] et des supports papier pour expliquer clairement aux usagers les attributs utilisés dans les questions de choix ainsi que leur donner un exemple [2].

C.1 CAWI

Le document ci-dessous présente une version raccourcie du question CAWI de l'EPD, c'est-à-dire sans tous les cas de figure (types d'abonnement, types de déplacement, etc.). Une version complète est disponible sur demande. Les captures d'écran des pages suivantes se lisent par colonne (colonne de gauche puis colonne de droite).

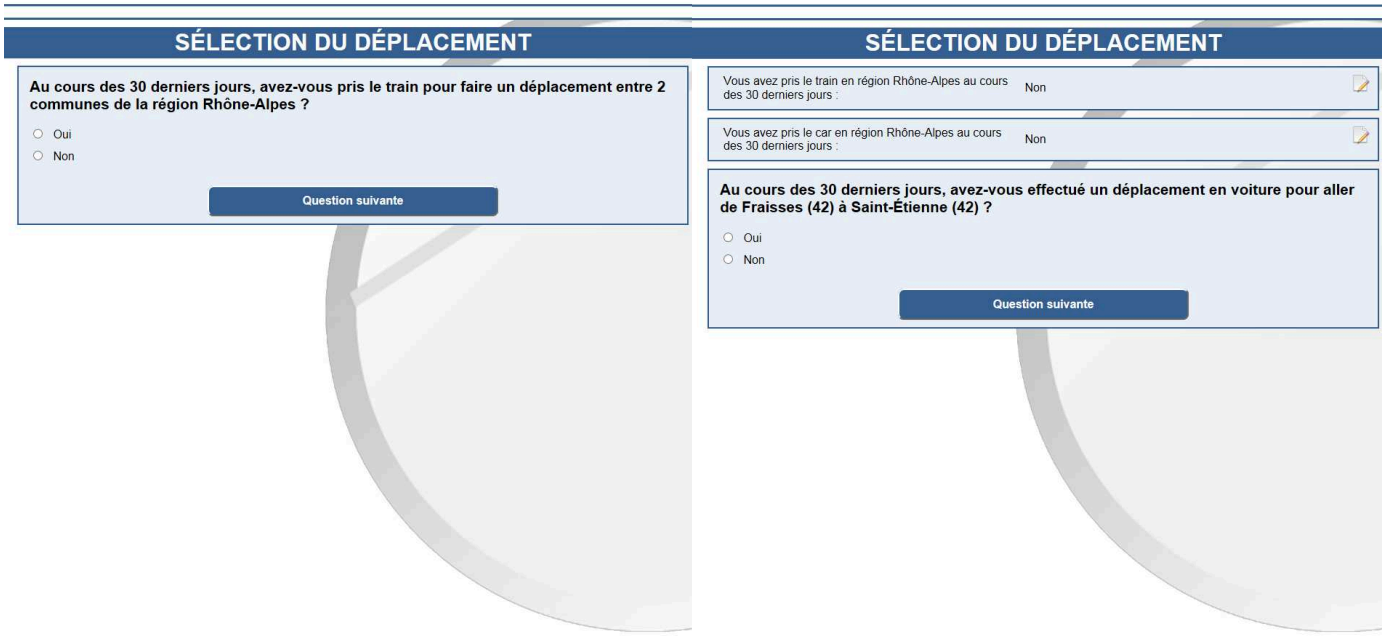
1. *Personal Digital assistant*
2. Tous les détails du déroulement de l'enquête sont dans le rapport remis à la Région Rhône-Alpes.


Accueil
ENQUÊTE SUR VOS PRÉFÉRENCES DE TRANSPORTS EN RHÔNE-ALPES
Bonjour,
Nous vous remercions de participer à l'enquête organisée par la Région Rhône-Alpes en collaboration avec l'Etat sur les déplacements effectués au sein de la région Rhône-Alpes.
Cette enquête a fait l'objet d'une déclaration à la Commission Nationale Informatique et Liberté, vos réponses resteront strictement confidentielles, et leur exploitation sera réalisée de façon anonyme. Les renseignements recueillis sont réservés uniquement à des fins statistiques et ne seront communiqués à aucun organisme extérieur.
Cette enquête prend moins de 15 minutes.
Pour toute information complémentaire, vous pouvez contacter le bureau d'études en charge de l'enquête au 01 55 48 91 29 ou au 01 55 48 91 26 du lundi au vendredi entre 14H00 et 19H00.
Vous pouvez également nous envoyer un mail à l'adresse : preferencestransport@enquete-web.fr
En vertu des dispositions de la loi n° 78-17 du 06/01/1978, relative à l'informatique, aux fichiers et aux libertés, nous vous informons que vous disposez d'un droit d'accès, de modification, de rectification et de suppression des données qui vous concernent.
Identification
identifiant :
user68
mot de passe :
Commencer
Questionnaire réalisé par - novembre 2014.


Accueil
INFORMATIONS PERSONNELLES
Agent ou ayant droit SNCF : Non
Votre sexe : Homme
Votre année de naissance : 1958
Quelle est votre occupation principale ?
Travail à plein temps
Travail à temps partiel
Formation en alternance (apprentissage, professionnalisation), stage
Etudiant
Scolaire jusqu'au BAC
Chômeur et/ou en recherche d'emploi
Retraité
Reste au foyer
Autre
Question suivante

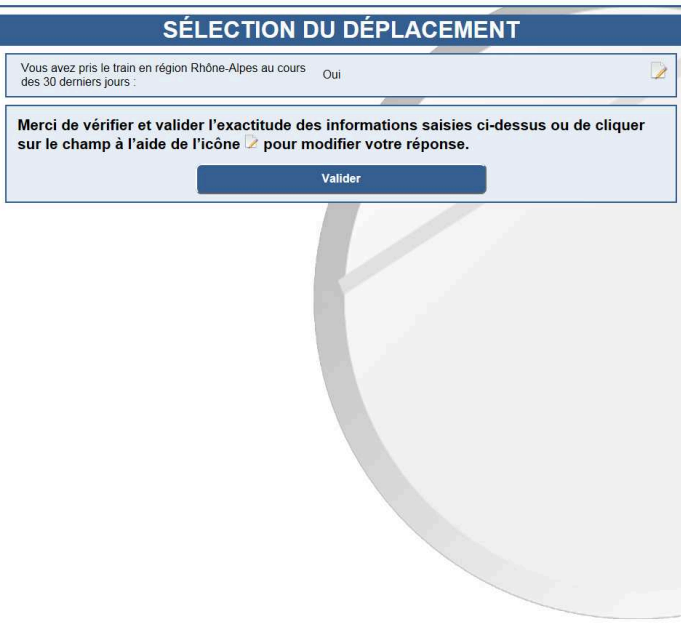
INFORMATIONS PERSONNELLES
Agent ou ayant droit SNCF : Non
Votre sexe : Homme
Votre année de naissance : 1958
Votre occupation principale : Travail à plein temps
Possession du permis : Oui
Nombre de personnes du ménage : 3 personnes
Nombre de voitures : 2 voitures
En moyenne (du lundi au vendredi), avec quelle fréquence utilisez-vous la voiture en tant que conducteur ?
Un aller-retour compte pour 2 déplacements.
Tous les jours ou presque
Deux déplacements par semaine au minimum
Deux déplacements par mois au minimum
Deux déplacements par an au minimum
Un déplacement et moins par an
Jamais
Question suivante

SÉLECTION DU DÉPLACEMENT
Vous avez pris le train en région Rhône-Alpes au cours des 30 derniers jours : Non
Au cours des 30 derniers jours, avez-vous pris le car pour faire un déplacement entre 2 communes de la région Rhône-Alpes ?
Oui
Non
Question suivante


SÉLECTION DU DÉPLACEMENT
Au cours des 30 derniers jours, avez-vous pris le train pour faire un déplacement entre 2 communes de la région Rhône-Alpes ?
Oui
Non
Question suivante

SÉLECTION DU DÉPLACEMENT
Vous avez pris le train en région Rhône-Alpes au cours des 30 derniers jours : Non
Vous avez pris le car en région Rhône-Alpes au cours des 30 derniers jours : Non
Au cours des 30 derniers jours, avez-vous effectué un déplacement en voiture pour aller de Fraisses (42) à Saint-Étienne (42) ?
Oui
Non
Question suivante

SÉLECTION DU DÉPLACEMENT
Vous avez pris le train en région Rhône-Alpes au cours des 30 derniers jours : Non
Vous avez pris le car en région Rhône-Alpes au cours des 30 derniers jours : Non
Déplacement de l'EDR au cours des 30 derniers jours : Non
Au cours des 30 derniers jours, avez-vous fait un déplacement en voiture de plus de 20 km entre 2 communes de la région Rhône-Alpes ?
Oui
Non
Question suivante


SÉLECTION DU DÉPLACEMENT
Vous avez pris le train en région Rhône-Alpes au cours des 30 derniers jours : Oui
Merci de vérifier et valider l'exactitude des informations saisies ci-dessus ou de cliquer sur le champ à l'aide de l'icône pour modifier votre réponse.
Valider

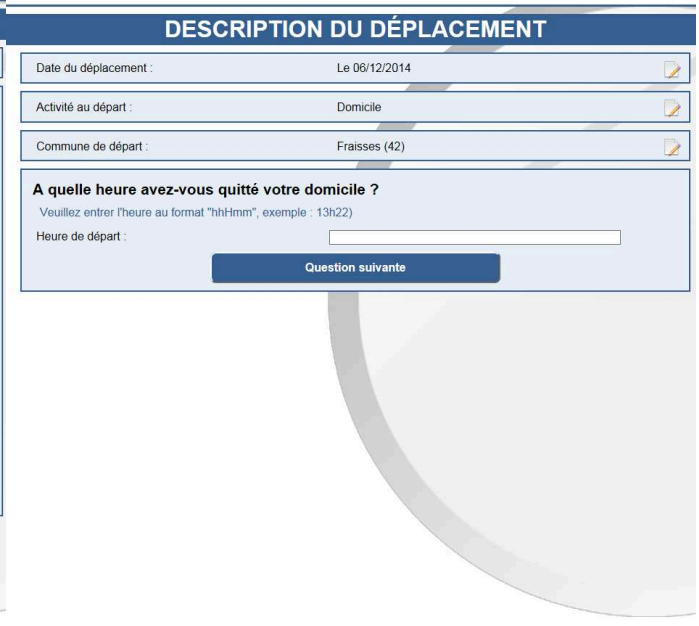
SÉLECTION DU DÉPLACEMENT
Vous avez pris le train en région Rhône-Alpes au cours des 30 derniers jours : Non
Vous avez pris le car en région Rhône-Alpes au cours des 30 derniers jours : Non
Déplacement de l'EDR au cours des 30 derniers jours : Non
Déplacement de plus de 20km au cours des 30 derniers jours : Non
Au cours des 30 derniers jours, avez-vous fait un déplacement en voiture entre 2 communes de la région Rhône-Alpes qui aurait pu être fait en train ou en car ?
Oui
Non
Je ne sais pas
Question suivante

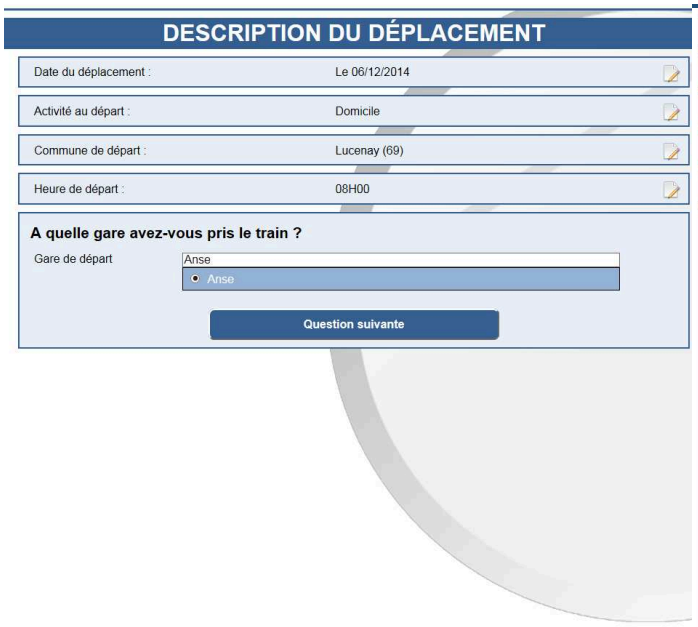
DESCRIPTION DU DÉPLACEMENT
Nous allons maintenant vous demander de décrire précisément ce déplacement que vous avez fait en train. Merci de nous indiquer :
- La date du déplacement ;
- L'activité à l'origine et à la destination du déplacement ;
- La commune de départ et d'arrivée ;
- L'ensemble des moyens de transport utilisés…
Question suivante

DESCRIPTION DU DÉPLACEMENT
A quel jour remonte ce déplacement ?
Veuillez entrer la date au format jj/mm/aaaa, par exemple 12/01/2015)
Date du déplacement :
12/01/2015
Question suivante

DESCRIPTION DU DÉPLACEMENT
Date du déplacement :
Le 06/12/2014
Activité au départ :
Domicile
Pouvez-vous nous indiquer la commune (ou l'arrondissement pour Lyon) où vous étiez avant de commencer ce déplacement ?
Commune de départ
Question suivante

DESCRIPTION DU DÉPLACEMENT
Date du déplacement :
Le 06/12/2014
Pouvez-vous nous indiquer le lieu où vous étiez au moment de votre départ ?
J'étais au domicile
J'étais au travail
J'étais sur mon lieu d'études
Collège
Lycée
Universités et grandes écoles
J'étais en train de faire des achats
Grands Magasins et leurs galeries marchandes
Hypermarchés et Supermachés
Petits et moyens commerces, marchés
J'étais en train de faire des démarches personnelles (administratif, santé, recherche d'emploi…)
J'étais en train de faire des visites ou une activité de loisir
J'étais en train de manger hors du domicile
J'étais en train d'aller chercher / déposer quelqu'un
Autre, précisez :
Question suivante

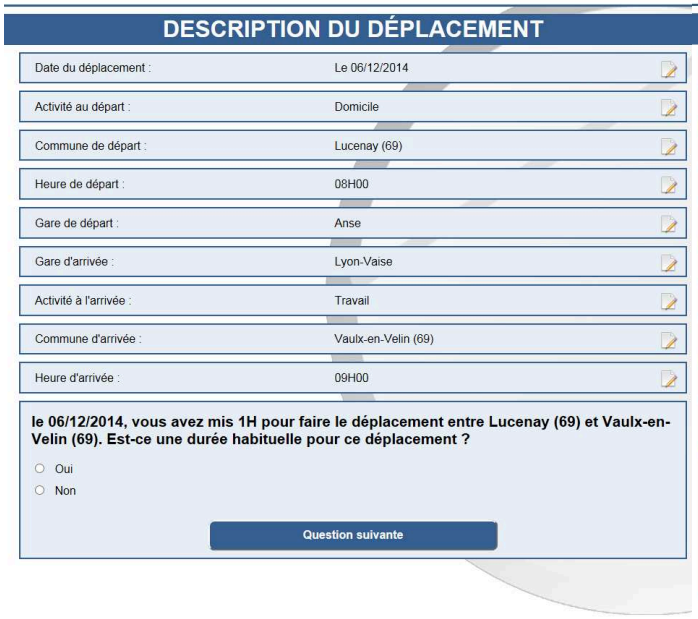

DESCRIPTION DU DÉPLACEMENT
Date du déplacement :
Le 06/12/2014
Activité au départ :
Domicile
Commune de départ :
Fraisses (42)
A quelle heure avez-vous quitté votre domicile ?
Veuillez entrer l'heure au format "hhHmm", exemple : 13h22)
Heure de départ :
Question suivante


DESCRIPTION DU DÉPLACEMENT
Date du déplacement : Le 06/12/2014
Activité au départ : Domicile
Commune de départ : Lucenay (69)
Heure de départ : 08H00
A quelle gare avez-vous pris le train ?
Gare de départ
Anse
Anse
Question suivante

Activité au départ : Domicile
Commune de départ : Lucenay (69)
Heure de départ : 08H00
Gare de départ : Anse
Gare d'arrivée : Lyon-Vaise
Pouvez-vous nous indiquer où vous êtes allé(e) ?
Je suis allé(e) à mon domicile
Je suis allé(e) au travail
Je suis allé(e) sur mon lieu d'études
Collège
Lycée
Universités et grandes écoles
Je suis allé(e) faire des achats
Grands Magasins et leurs galeries marchandes
Hypermarchés et Supermachés
Petits et moyens commerces, marchés
Je suis allé(e) faire des démarches personnelles (administratif, santé, recherche d'emplo
Je suis allé(e) faire des visites ou une activité de loisir
Je suis allé(e) manger hors du domicile
Je suis allé(e) chercher / déposer quelqu'un
Autre
Question suivante

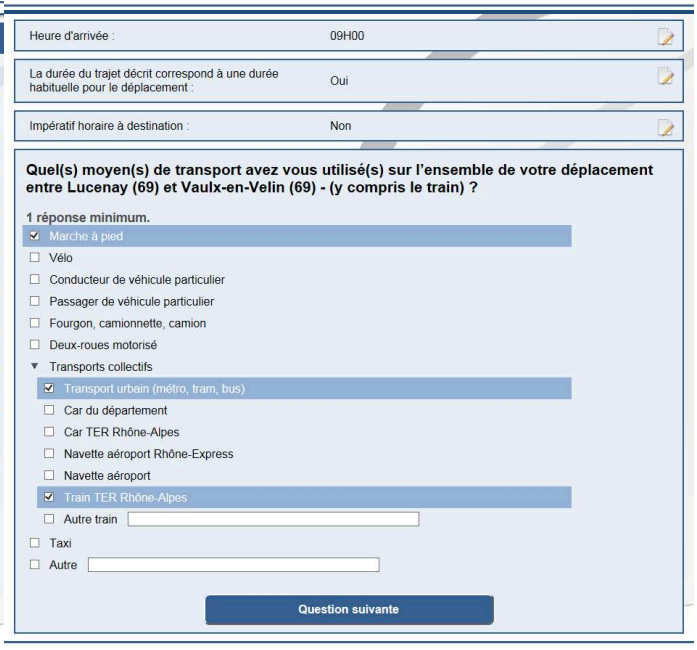
DESCRIPTION DU DÉPLACEMENT
Date du déplacement : Le 06/12/2014
Activité au départ : Domicile
Commune de départ : Lucenay (69)
Heure de départ : 08H00
Gare de départ : Anse
A quelle gare êtes-vous descendu ?
Gare d'arrivée
Lyon-Gorge De Loup
Lyon-Vaise
Lyon-Saint-Paul
Lyon-Gorge De Loup
Lyon-Perrache
Lyon-La Part Dieu
Lyon-Saint-Exupery-TGV
Lyon Jean Macé
Question suivante

DESCRIPTION DU DÉPLACEMENT
Date du déplacement : Le 06/12/2014
Activité au départ : Domicile
Commune de départ : Lucenay (69)
Heure de départ : 08H00
Gare de départ : Anse
Gare d'arrivée : Lyon-Vaise
Activité à l'arrivée : Travail
Commune d'arrivée : Vaulx-en-Velin (69)
A quelle heure êtes-vous arrivé à votre travail ?
Veuillez entrer l'heure au format "hhHmm", exemple : 13h22)
Heure d'arrivée :
Question suivante

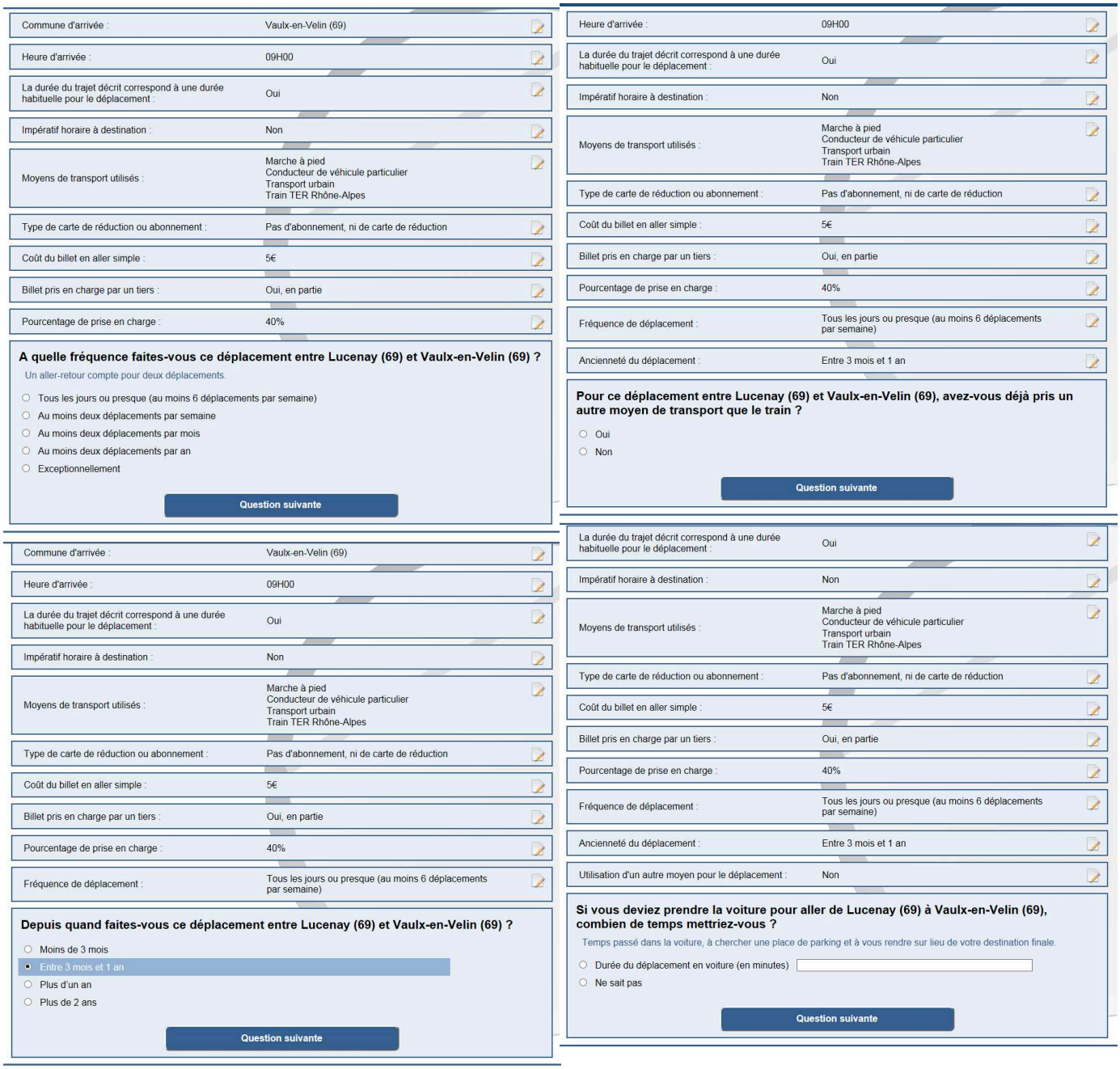

DESCRIPTION DU DÉPLACEMENT
Date du déplacement : Le 06/12/2014
Activité au départ : Domicile
Commune de départ : Lucenay (69)
Heure de départ : 08H00
Gare de départ : Anse
Gare d'arrivée : Lyon-Vaise
Activité à l'arrivée : Travail
Commune d'arrivée : Vaulx-en-Velin (69)
Heure d'arrivée : 09H00
le 06/12/2014, vous avez mis 1H pour faire le déplacement entre Lucenay (69) et Vaulx-en-Velin (69). Est-ce une durée habituelle pour ce déplacement ?
Oui
Non
Question suivante

DESCRIPTION DU DÉPLACEMENT
Date du déplacement : Le 06/12/2014
Activité au départ : Domicile
Commune de départ : Lucenay (69)
Heure de départ : 08H00
Gare de départ : Anse
Gare d'arrivée : Lyon-Vaise
Activité à l'arrivée : Travail
Commune d'arrivée : Vaulx-en-Velin (69)
Heure d'arrivée : 09H00
La durée du trajet décrit correspond à une durée habituelle pour le déplacement : Oui
Aviez-vous un impératif horaire à destination ?
Oui
Non
Question suivante

DESCRIPTION DU DÉPLACEMENT
Date du déplacement : Le 06/12/2014
Activité au départ : Domicile
Commune de départ : Lucenay (69)
Heure de départ : 08H00
Gare de départ : Anse
Gare d'arrivée : Lyon-Vaise
Activité à l'arrivée : Travail
Commune d'arrivée : Vaulx-en-Velin (69)
Heure d'arrivée : 09H00
La durée du trajet décrit correspond à une durée habituelle pour le déplacement : Non
Combien de temps mettez-vous habituellement pour faire ce déplacement entre Lucenay (69) et Vaulx-en-Velin (69) ?
Temps habituel pour le déplacement (en minutes) :
Question suivante


Heure d'arrivée : 09H00
La durée du trajet décrit correspond à une durée habituelle pour le déplacement : Oui
Impératif horaire à destination : Non
Quel(s) moyen(s) de transport avez vous utilisé(s) sur l'ensemble de votre déplacement entre Lucenay (69) et Vaulx-en-Velin (69) - (y compris le train) ?
1 réponse minimum.
Marche à pied
Vélo
Conducteur de véhicule particulier
Passager de véhicule particulier
Fourgon, camionnette, camion
Deux-roues motorisé
Transports collectifs
Transport urbain (métro, tram, bus)
Car du département
Car TER Rhône-Alpes
Navette aéroport Rhône-Express
Navette aéroport
Train TER Rhône-Alpes
Autre train
Taxi
Autre
Question suivante

Moyens de transport utilisés :
Marche à pied
Conducteur de véhicule particulier
Transport urbain
Train TER Rhône-Alpes
Quel type de réduction ou d'abonnement avez-vous utilisé pour ce déplacement ?
Pas d'abonnement, ni de carte de réduction
Abonnement hebdomadaire
Abonnement mensuel
Abonnement semestriel (6 mois)
Abonnement annuel
Carte Illico Liberté
Carte Illico Jeune
Carte Fréquence 1ère
Carte Fréquence 2de
Carte We
Carte Enfant+
Carte Famille
Carte Famille Nombreuse (30%)
Carte Famille Nombreuse (40%)
Carte Famille Nombreuse (50%)
Carte Famille Nombreuse (75%)
Militaire
Carte 12-27
Carte Sénior
Autre
Question suivante
Heure de départ : 08H00
Gare de départ : Anse
Gare d'arrivée : Lyon-Vaise
Activité à l'arrivée : Travail
Commune d'arrivée : Vaulx-en-Velin (69)
Heure d'arrivée : 09H00
La durée du trajet décrit correspond à une durée habituelle pour le déplacement : Oui
Impératif horaire à destination : Non
Moyens de transport utilisés : Marche à pied, Conducteur de véhicule particulier, Transport urbain, Train TER Rhône-Alpes
Type de carte de réduction ou abonnement : Abonnement hebdomadaire
Coût de l'abonnement par semaine : 65€
Le prix de cet abonnement de train est-il pris en charge par un tiers (employeur, etc.) ?
Oui, en totalité
Oui, en partie
Non
Question suivante
Commune de départ : Lucenay (69)
Heure de départ : 08H00
Gare de départ : Anse
Gare d'arrivée : Lyon-Vaise
Activité à l'arrivée : Travail
Commune d'arrivée : Vaulx-en-Velin (69)
Heure d'arrivée : 09H00
La durée du trajet décrit correspond à une durée habituelle pour le déplacement : Oui
Impératif horaire à destination : Non
Moyens de transport utilisés : Marche à pied, Conducteur de véhicule particulier, Transport urbain, Train TER Rhône-Alpes
Type de carte de réduction ou abonnement : Abonnement hebdomadaire
Combien coûte votre abonnement par semaine?
Indiquer l'ensemble du coût de l'abonnement même si celui-ci est pris, intégralement ou en partie, en charge par un tiers.
Si vous prenez plusieurs trains, plusieurs cars ou un train et un car, merci d'indiquer l'ensemble du coût de votre abonnement.
Coût de l'abonnement par semaine :
Question suivante
Heure de départ : 08H00
Gare de départ : Anse
Gare d'arrivée : Lyon-Vaise
Activité à l'arrivée : Travail
Commune d'arrivée : Vaulx-en-Velin (69)
Heure d'arrivée : 09H00
La durée du trajet décrit correspond à une durée habituelle pour le déplacement : Oui
Impératif horaire à destination : Non
Moyens de transport utilisés : Marche à pied, Conducteur de véhicule particulier, Transport urbain, Train TER Rhône-Alpes
Type de carte de réduction ou abonnement : Abonnement hebdomadaire
Coût de l'abonnement par semaine : 65€
Billet pris en charge par un tiers : Oui, en partie
Quel pourcentage est pris en charge ?
Pourcentage de prise en charge : 50 %
Question suivante



Commune d'arrivée :	Vaulx-en-Velin (69)
Heure d'arrivée :	09H00
La durée du trajet décrit correspond à une durée habituelle pour le déplacement :	Oui
Impératif horaire à destination :	Non
Moyens de transport utilisés :	Marche à pied Conducteur de véhicule particulier Transport urbain Train TER Rhône-Alpes
Type de carte de réduction ou abonnement :	Pas d'abonnement, ni de carte de réduction
Coût du billet en aller simple :	5€
Billet pris en charge par un tiers :	Oui, en partie
Pourcentage de prise en charge :	40%

A quelle fréquence faites-vous ce déplacement entre Lucenay (69) et Vaulx-en-Velin (69) ?

Un aller-retour compte pour deux déplacements.

- ○ Tous les jours ou presque (au moins 6 déplacements par semaine)
- ○ Au moins deux déplacements par semaine
- ○ Au moins deux déplacements par mois
- ○ Au moins deux déplacements par an
- ○ Exceptionnellement

Question suivante

Heure d'arrivée :	09H00
La durée du trajet décrit correspond à une durée habituelle pour le déplacement :	Oui
Impératif horaire à destination :	Non
Moyens de transport utilisés :	Marche à pied Conducteur de véhicule particulier Transport urbain Train TER Rhône-Alpes
Type de carte de réduction ou abonnement :	Pas d'abonnement, ni de carte de réduction
Coût du billet en aller simple :	5€
Billet pris en charge par un tiers :	Oui, en partie
Pourcentage de prise en charge :	40%
Fréquence de déplacement :	Tous les jours ou presque (au moins 6 déplacements par semaine)
Ancienneté du déplacement :	Entre 3 mois et 1 an

Pour ce déplacement entre Lucenay (69) et Vaulx-en-Velin (69), avez-vous déjà pris un autre moyen de transport que le train ?

- ○ Oui
- ○ Non

Question suivante

Commune d'arrivée :	Vaulx-en-Velin (69)
Heure d'arrivée :	09H00
La durée du trajet décrit correspond à une durée habituelle pour le déplacement :	Oui
Impératif horaire à destination :	Non
Moyens de transport utilisés :	Marche à pied Conducteur de véhicule particulier Transport urbain Train TER Rhône-Alpes
Type de carte de réduction ou abonnement :	Pas d'abonnement, ni de carte de réduction
Coût du billet en aller simple :	5€
Billet pris en charge par un tiers :	Oui, en partie
Pourcentage de prise en charge :	40%
Fréquence de déplacement :	Tous les jours ou presque (au moins 6 déplacements par semaine)

Depuis quand faites-vous ce déplacement entre Lucenay (69) et Vaulx-en-Velin (69) ?

- ○ Moins de 3 mois
- ◉ Entre 3 mois et 1 an
- ○ Plus d'un an
- ○ Plus de 2 ans

Question suivante

La durée du trajet décrit correspond à une durée habituelle pour le déplacement :	Oui
Impératif horaire à destination :	Non
Moyens de transport utilisés :	Marche à pied Conducteur de véhicule particulier Transport urbain Train TER Rhône-Alpes
Type de carte de réduction ou abonnement :	Pas d'abonnement, ni de carte de réduction
Coût du billet en aller simple :	5€
Billet pris en charge par un tiers :	Oui, en partie
Pourcentage de prise en charge :	40%
Fréquence de déplacement :	Tous les jours ou presque (au moins 6 déplacements par semaine)
Ancienneté du déplacement :	Entre 3 mois et 1 an
Utilisation d'un autre moyen pour le déplacement :	Non

Si vous deviez prendre la voiture pour aller de Lucenay (69) à Vaulx-en-Velin (69), combien de temps mettriez-vous ?

Temps passé dans la voiture, à chercher une place de parking et à vous rendre sur lieu de votre destination finale.

- ○ Durée du déplacement en voiture (en minutes)
- ○ Ne sait pas

Question suivante

VOS PRÉFÉRENCES

Choix entre 3 options - première série

Imaginez que vous ayez 3 options : deux en train et une en voiture pour effectuer le déplacement de Lucenay (69) à Vaulx-en-Velin (69). Merci de nous indiquer laquelle vous choisiriez si ces trois options étaient à votre disposition.

Votre réponse doit être faite en vous plaçant dans la situation du déplacement du 20/02/2015 (même motif de déplacement, même contraintes horaires...). L'ensemble du déplacement est pris en compte, de Lucenay (69) à Vaulx-en-Velin (69) (y compris les trajets pour accéder à la gare de départ et depuis la gare d'arrivée jusqu'à la destination finale).

Les caractéristiques proposées pour ce déplacement varient entre les questions et peuvent différer de la réalité. Nous vous demandons de vous baser sur les valeurs indiquées dans les questions et non sur la situation vécue le 20/02/2015 pour faire vos choix.

Selon quelles caractéristiques faire vos choix ?

Les 5 premiers choix que vous allez faire vont dépendre du retard éventuel, de la fréquence de passage des trains, du temps de déplacement et du coût du déplacement de Lucenay (69) à Vaulx-en-Velin (69) :

TRAIN	VOITURE
Retard : Un retard de 15 minutes 1 fois sur 10 signifie que tous les 10 déplacements, vous aurez 15 minutes de retard à cause d'accidents ou de bouchons en voiture, de problèmes techniques en train ou pour toutes autres raisons. Les autres fois, vous serez à l'heure Le retard est donc défini par **le nombre de retards** : moins d'1 train sur 10 (5%), 1 train sur 10 (10%) ou 2 trains sur 10 (20%) et la **durée de retard** : 10, 15 ou 30 minutes.	
Fréquence : Nombre de trains qui passent chaque heure. Il peut passer un train toutes les deux heures ; 1, 2 ou 4 trains par heure. Exemple avec 4 trains par heure : 4 trains par heure	
Temps de déplacement : temps d'accès à la gare (à pied, en bus, à vélo...), temps passé dans le train et temps pour vous rendre à votre destination finale.	**Temps de déplacement** : temps pour récupérer votre voiture, temps passé dans la voiture, à chercher une place de parking et à vous rendre à votre destination finale.
Coût du déplacement : Coût du ou des titres de transport, y compris le prix du ticket de bus ou de tram ou le coût éventuel de l'essence pour aller ou partir de la gare. Le coût est exprimé par déplacement (aller simple).	**Coût du déplacement** : Coût de l'essence, et éventuellement du parking et du péage. Le coût est exprimé par déplacement (aller simple).

Exemple de choix à faire

	Option A	Option B	Option C
Moyen de transport			
Retard	Retard de 15 minutes 2 fois sur 10 (20%)	Retard de 30 minutes 1 fois sur 10 (10%)	Retard de 15 minutes 1 fois sur 10 (10%)
Fréquence	1 train par heure	2 trains par heure	Départ quand vous le voulez
Temps de déplacement	54 minutes	1H21	1H30
Coût du déplacement	23€ / déplacement	18€ / déplacement	21€ / déplacement
Cochez l'option choisie :	○	◉	○

Exemple de réponse

	Option A	Option B	Option C
Moyen de transport			
Retard		Retard de 30 minutes 1 fois sur 10 (10%)	Retard de 15 minutes 1 fois sur 10 (10%)
Fréquence		2 trains par heure	Départ quand vous le voulez
Temps de déplacement		1H21	1H30
Coût du déplacement		18€ / déplacement	21€ / déplacement
Cochez l'option choisie :	○	◉	○

Si vous choisissez l'option B, vous choisissez de vous déplacer en train.

9 fois sur 10 vous êtes à l'heure. Mais 1 fois sur 10, vous êtes en retard de 30 minutes.

Il passe 2 trains par heure.

Le déplacement dure 1H21 et vous coûte 18 euros par déplacement.

Question suivante

VOS PRÉFÉRENCES

Descriptif de l'exercice 1.

Cas 1.1 : Que choisissez-vous ?

	Option A	Option B	Option C
Moyen de transport			
Retard	Retard de 15 minutes 1 fois sur 10 (10%)	Retard de 30 minutes 1 fois sur 10 (10%)	Retard de 30 minutes Moins de 1 fois sur 10 (5%)
Fréquence	2 trains par heure	1 train par heure	Départ quand vous le voulez
Temps de déplacement	1H30	1H30	34 minutes
Coût du déplacement	7€ / déplacement	5€ / déplacement	4€ / déplacement
Cochez l'option choisie :	○	○	○

Question suivante

VOS PRÉFÉRENCES

Descriptif de l'exercice 1.

Cas 1.1 : Je choisis l'option A

Cas 1.2 : Que choisissez-vous ?

	Option A	Option B	Option C
Moyen de transport			
Retard	Retard de 15 minutes 1 fois sur 10 (10%)	Retard de 15 minutes 1 fois sur 10 (10%)	Retard de 10 minutes Moins de 1 fois sur 10 (5%)
Fréquence	2 trains par heure	4 trains par heure	Départ quand vous le voulez
Temps de déplacement	1H30	1H07	56 minutes
Coût du déplacement	7€ / déplacement	9€ / déplacement	4€ / déplacement
Cochez l'option choisie :	○	○	○

Question suivante

VOS PRÉFÉRENCES

Descriptif de l'exercice 1.

Cas 1.1 : Je choisis l'option A

Cas 1.2 : Je choisis l'option B

Cas 1.3 : Que choisissez-vous ?

	Option A	Option B	Option C
Moyen de transport			
Retard	Retard de 15 minutes 1 fois sur 10 (10%)	Retard de 10 minutes 2 fois sur 10 (20%)	Retard de 15 minutes 1 fois sur 10 (10%)
Fréquence	2 trains par heure	1 train toutes les 2 heures	Départ quand vous le voulez
Temps de déplacement	1H30	1H52	56 minutes
Coût du déplacement	7€ / déplacement	7€ / déplacement	6€ / déplacement
Cochez l'option choisie :	○	○	○

Question suivante

Descriptif de l'exercice 1.

Cas 1.1 : Je choisis l'option A

Cas 1.2 : Je choisis l'option B

Cas 1.3 : Je choisis l'option C

Cas 1.4 : Je choisis l'option A

Cas 1.5 : Que choisissez-vous ?

	Option A	Option B	Option C
Moyen de transport			
Retard	Retard de 15 minutes 1 fois sur 10 (10%)	Retard de 30 minutes 1 fois sur 10 (10%)	Retard de 30 minutes 1 fois sur 10 (10%)
Fréquence	2 trains par heure	2 trains par heure	Départ quand vous le voulez
Temps de déplacement	1H30	1H30	56 minutes
Coût du déplacement	7€ / déplacement	7€ / déplacement	6€ / déplacement
Cochez l'option choisie :	○	○	○

Question suivante

VOS PRÉFÉRENCES

Descriptif de l'exercice 1.

Cas 1.1 : Je choisis l'option A

Cas 1.2 : Je choisis l'option B

Cas 1.3 : Je choisis l'option C

Cas 1.4 : Que choisissez-vous ?

	Option A	Option B	Option C
Moyen de transport			
Retard	Retard de 15 minutes 1 fois sur 10 (10%)	Retard de 30 minutes Moins de 1 fois sur 10 (5%)	Retard de 10 minutes 2 fois sur 10 (20%)
Fréquence	2 trains par heure	4 trains par heure	Départ quand vous le voulez
Temps de déplacement	1H30	1H07	45 minutes
Coût du déplacement	7€ / déplacement	5€ / déplacement	6€ / déplacement
Cochez l'option choisie :	○	○	○

Question suivante

Cas 1.5 : Je choisis l'option A

Choix entre 3 options - deuxième série

Vous allez maintenant faire une deuxième série de choix, incluant entre autres le cadencement.

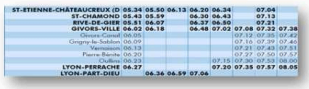
Qu'est-ce que le cadencement ?

Le cadencement signifie que les trains circulent à intervalles de temps réguliers.
Actuellement, la plupart des trains partent à heure fixe d'une gare donnée, par exemple : 5h20, 5h50, 6h20, 6h50...
Le cadencement permet de savoir quand partira le train, quelle que soit l'heure où on rejoint la gare, puisque les minutes de départ sont les mêmes à toutes les heures.

Horaires cadencés

Dans cet exemple de grille horaire, les horaires de train sont cadencés.

Horaires non cadencés

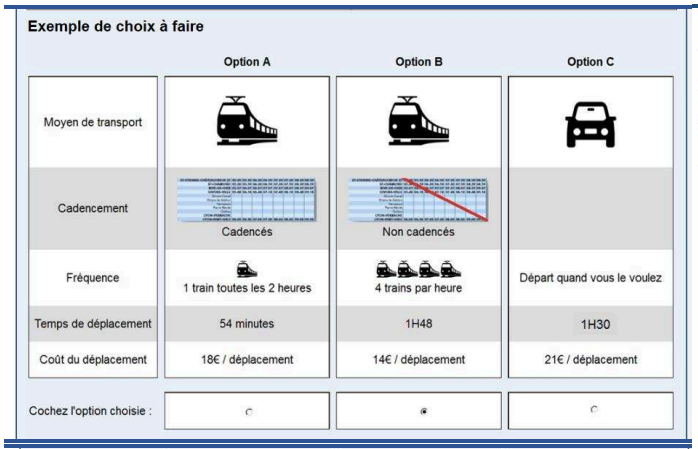


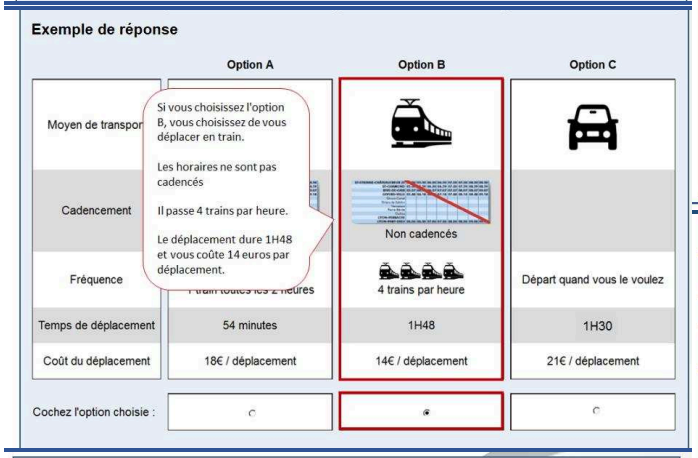
Dans cet exemple de grille horaire, les trains passent à intervalles de temps irréguliers, les horaires de train ne sont pas cadencés.

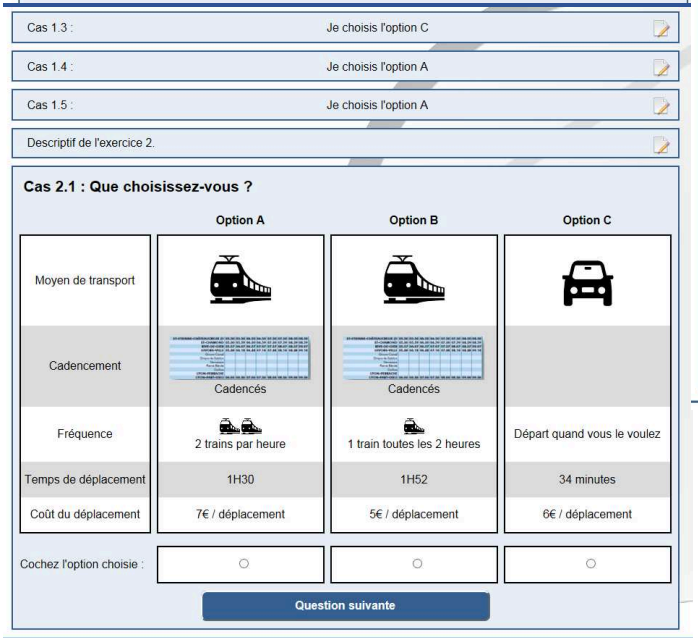
Selon quelles caractéristiques faire vos choix ?

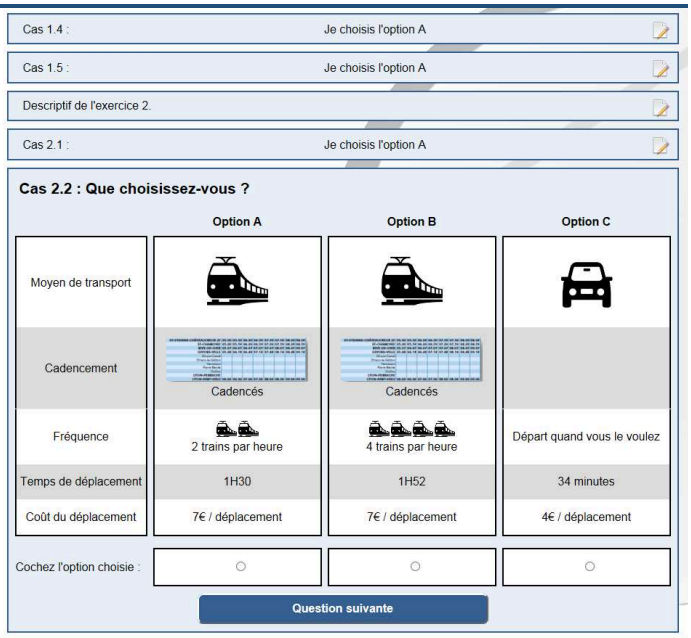
Les 5 choix que vous allez faire maintenant vont dépendre du cadencement des horaires de train, de la fréquence, du temps de déplacement et du coût du déplacement de Lucenay (69) à Vaulx-en-Velin (69) :

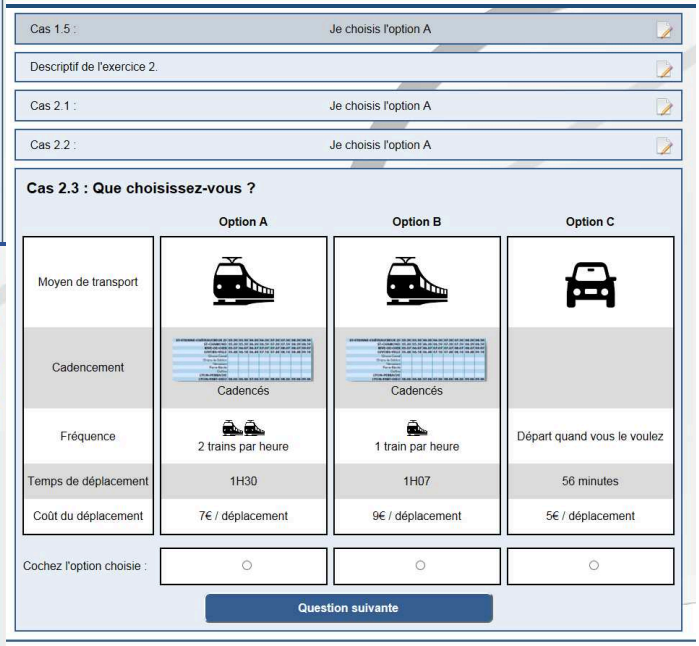
TRAIN	VOITURE
Cadencement : Un horaire cadencé signifie que le train part toutes les heures aux mêmes minutes ; ex : 8h17, 8h37, 8h57, 9h17, 9h37, 9h57... Si les horaires ne sont pas cadencés, les trains passent à intervalle de temps irrégulier. Horaires cadencés Horaires non cadencés	
Fréquence : Nombre de trains qui passent chaque heure. Il peut passer un train toutes les deux heures ; 1, 2 ou 4 trains par heure. Exemple avec 4 trains par heure : 4 trains par heure	
Temps de déplacement : temps d'accès à la gare (à pied, en bus, à vélo...), temps passé dans le train et temps pour vous rendre à votre destination finale.	**Temps de déplacement** : temps pour récupérer votre voiture, temps passé dans la voiture, à chercher une place de parking et à vous rendre à votre destination finale.
Coût du déplacement : Coût du ou des titres de transport, y compris le prix du ticket de bus ou de tram ou le coût éventuel de l'essence pour aller ou partir de la gare. Le coût est exprimé par déplacement (aller simple).	**Coût du déplacement** : Coût de l'essence, et éventuellement du parking et du péage. Le coût est exprimé par déplacement (aller simple).


Exemple de choix à faire
Option A
Option B
Option C
Moyen de transport
Cadencement
Cadencés
Non cadencés
Fréquence
1 train toutes les 2 heures
4 trains par heure
Départ quand vous le voulez
Temps de déplacement
54 minutes
1H48
1H30
Coût du déplacement
18€ / déplacement
14€ / déplacement
21€ / déplacement
Cochez l'option choisie :


Exemple de réponse
Option A
Option B
Option C
Si vous choisissez l'option B, vous choisissez de vous déplacer en train.
Les horaires ne sont pas cadencés
Il passe 4 trains par heure.
Le déplacement dure 1H48 et vous coûte 14 euros par déplacement.
Moyen de transport
Cadencement
Non cadencés
Fréquence
4 trains par heure
Départ quand vous le voulez
Temps de déplacement
54 minutes
1H48
1H30
Coût du déplacement
18€ / déplacement
14€ / déplacement
21€ / déplacement
Cochez l'option choisie :


Cas 1.3 : Je choisis l'option C
Cas 1.4 : Je choisis l'option A
Cas 1.5 : Je choisis l'option A
Descriptif de l'exercice 2.
Cas 2.1 : Que choisissez-vous ?
Option A
Option B
Option C
Moyen de transport
Cadencement
Cadencés
Cadencés
Fréquence
2 trains par heure
1 train toutes les 2 heures
Départ quand vous le voulez
Temps de déplacement
1H30
1H52
34 minutes
Coût du déplacement
7€ / déplacement
5€ / déplacement
6€ / déplacement
Cochez l'option choisie :
Question suivante


Cas 1.4 : Je choisis l'option A
Cas 1.5 : Je choisis l'option A
Descriptif de l'exercice 2.
Cas 2.1 : Je choisis l'option A
Cas 2.2 : Que choisissez-vous ?
Option A
Option B
Option C
Moyen de transport
Cadencement
Cadencés
Cadencés
Fréquence
2 trains par heure
4 trains par heure
Départ quand vous le voulez
Temps de déplacement
1H30
1H52
34 minutes
Coût du déplacement
7€ / déplacement
7€ / déplacement
4€ / déplacement
Cochez l'option choisie :
Question suivante


Cas 1.5 : Je choisis l'option A
Descriptif de l'exercice 2.
Cas 2.1 : Je choisis l'option A
Cas 2.2 : Je choisis l'option A
Cas 2.3 : Que choisissez-vous ?
Option A
Option B
Option C
Moyen de transport
Cadencement
Cadencés
Cadencés
Fréquence
2 trains par heure
1 train par heure
Départ quand vous le voulez
Temps de déplacement
1H30
1H07
56 minutes
Coût du déplacement
7€ / déplacement
9€ / déplacement
5€ / déplacement
Cochez l'option choisie :
Question suivante

Descriptif de l'exercice 2.

Cas 2.1 : Je choisis l'option A

Cas 2.2 : Je choisis l'option A

Cas 2.3 : Je choisis l'option A

Cas 2.4 : Que choisissez-vous ?

	Option A	Option B	Option C
Moyen de transport			
Cadencement	Cadencés	Non cadencés	
Fréquence	2 trains par heure	2 trains par heure	Départ quand vous le voulez
Temps de déplacement	1H30	1H52	34 minutes
Coût du déplacement	7€ / déplacement	5€ / déplacement	6€ / déplacement
Cochez l'option choisie :	○	○	○

Question suivante

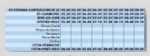
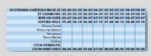
Cas 2.1 : Je choisis l'option A

Cas 2.2 : Je choisis l'option A

Cas 2.3 : Je choisis l'option A

Cas 2.4 : Je choisis l'option A

Cas 2.5 : Que choisissez-vous ?

	Option A	Option B	Option C
Moyen de transport			
Cadencement	 Cadencés	 Cadencés	
Fréquence	2 trains par heure	1 train par heure	Départ quand vous le voulez
Temps de trajet	1H30	1H07	56 minutes
Coût de trajet	7€ / déplacement	7€ / déplacement	4€ / déplacement
Option choisie =>	○	○	○

Question suivante

Choix entre 3 options - dernière série

Vous allez maintenant faire une dernière série de choix avec une option en train, une option en car et une option en voiture.

Selon quelles caractéristiques faire vos choix ?

Les 4 choix que vous allez faire maintenant vont dépendre de la garantie ou non d'avoir une place assise, du temps de déplacement et du coût du déplacement de Lucenay (69) à Vaulx-en-Velin (69) :

TRAIN	CAR	VOITURE
Place assise : Garantie ou non d'avoir une place assise dans le train. Si la place assise n'est pas garantie, il peut vous arriver de voyager debout sur tout ou une partie du déplacement. La place assise est toujours garantie dans le car. Place assise non garantie : Place assise garantie :		
Temps de déplacement : temps d'accès à la gare ou à l'arrêt de car (à pied, en bus, à vélo...), temps passé dans le train ou le car et temps pour vous rendre à votre destination finale.		**Temps de déplacement** : temps pour récupérer votre voiture, temps passé dans la voiture, à chercher une place de parking et à vous rendre à votre destination finale.
Coût du déplacement : Coût du ou des titres de transport, y compris le prix du ticket de bus ou de tram ou le coût éventuel de l'essence pour aller ou partir de la gare ou de l'arrêt de car. Le coût est exprimé par déplacement (aller simple).		**Coût du déplacement** : Coût de l'essence, et éventuellement du parking et du péage. Le coût est exprimé par déplacement (aller simple).

Exemple de réponse

	Option A	Option B	Option C
Moyen de transport			
Place Assise	Si vous choisissez l'option B, vous choisissez de vous déplacer en car. Vous êtes assuré(e) de disposer d'une place assise	Place assise garantie	
Temps de déplacement	Le déplacement dure 1H21 et vous coûte 23 euros par déplacement.	1H21	1H30
Coût du déplacement	18€ / déplacement	23€ / déplacement	21€ / déplacement
Cochez l'option choisie :	○	◉	○

Cas 2.1 : Je choisis l'option A

Cas 2.2 : Je choisis l'option A

Cas 2.3 : Je choisis l'option A

Cas 2.4 : Je choisis l'option A

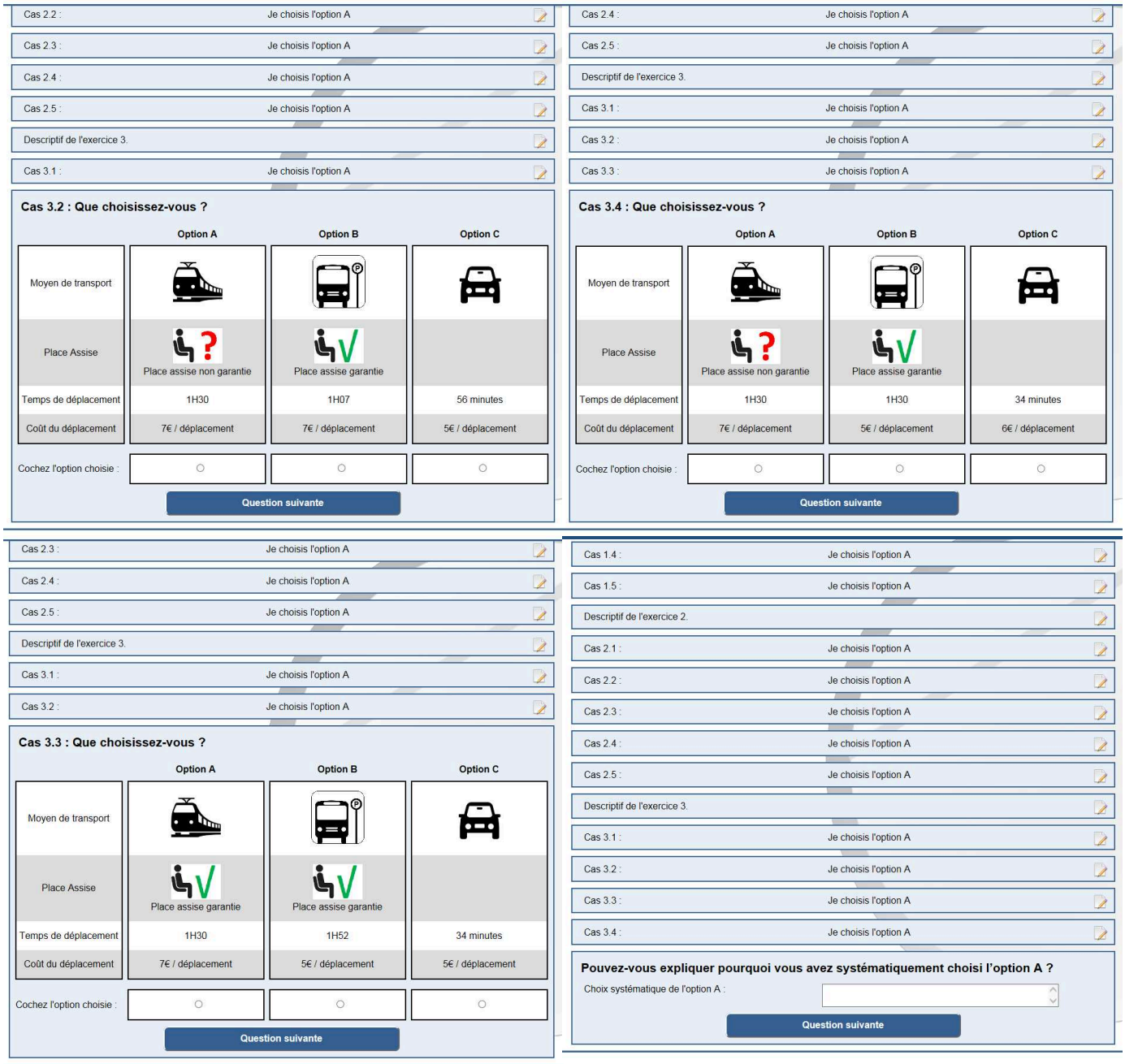
Cas 2.5 : Je choisis l'option A

Descriptif de l'exercice 3.

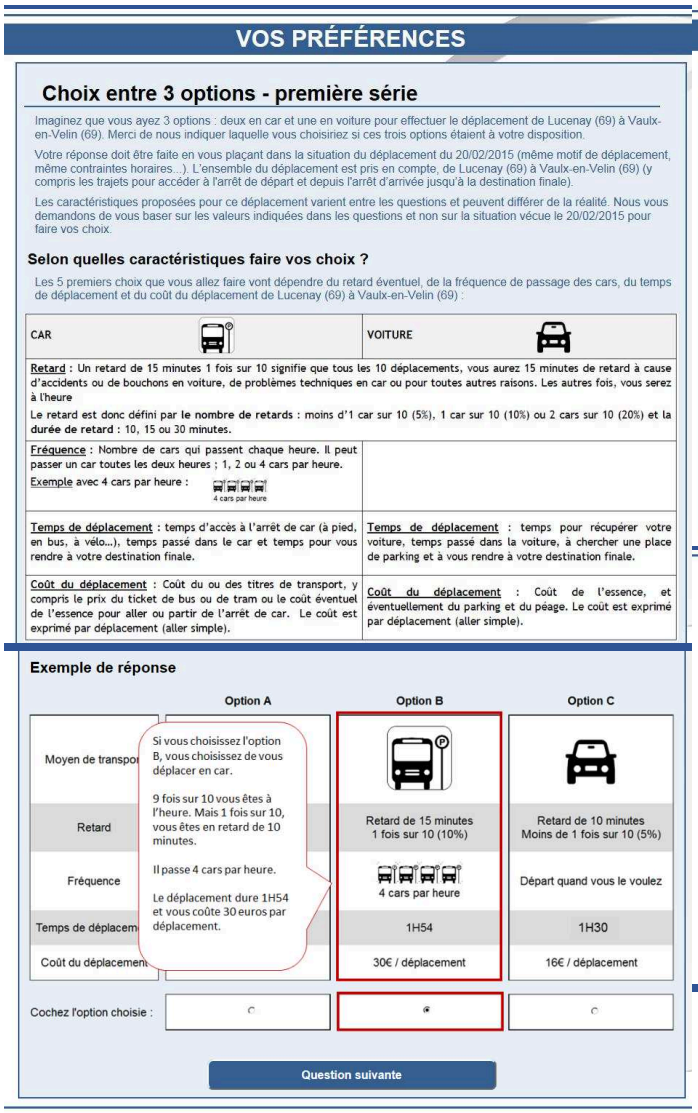
Cas 3.1 : Que choisissez-vous ?

	Option A	Option B	Option C
Moyen de transport			
Place Assise	Place assise non garantie	Place assise garantie	
Temps de déplacement	1H30	1H52	34 minutes
Coût du déplacement	7€ / déplacement	5€ / déplacement	6€ / déplacement
Cochez l'option choisie :	○	○	○

Question suivante


Cas 2.2 : Je choisis l'option A
Cas 2.3 : Je choisis l'option A
Cas 2.4 : Je choisis l'option A
Cas 2.5 : Je choisis l'option A
Descriptif de l'exercice 3.
Cas 3.1 : Je choisis l'option A
Cas 3.2 : Que choisissez-vous ?
Option A
Option B
Option C
Moyen de transport
Place Assise
Place assise non garantie
Place assise garantie
Temps de déplacement
1H30
1H07
56 minutes
Coût du déplacement
7€ / déplacement
7€ / déplacement
5€ / déplacement
Cochez l'option choisie :
Question suivante
Cas 2.4 : Je choisis l'option A
Cas 2.5 : Je choisis l'option A
Descriptif de l'exercice 3.
Cas 3.1 : Je choisis l'option A
Cas 3.2 : Je choisis l'option A
Cas 3.3 : Je choisis l'option A
Cas 3.4 : Que choisissez-vous ?
Option A
Option B
Option C
Moyen de transport
Place Assise
Place assise non garantie
Place assise garantie
Temps de déplacement
1H30
1H30
34 minutes
Coût du déplacement
7€ / déplacement
5€ / déplacement
6€ / déplacement
Cochez l'option choisie :
Question suivante
Cas 2.3 : Je choisis l'option A
Cas 2.4 : Je choisis l'option A
Cas 2.5 : Je choisis l'option A
Descriptif de l'exercice 3.
Cas 3.1 : Je choisis l'option A
Cas 3.2 : Je choisis l'option A
Cas 3.3 : Que choisissez-vous ?
Option A
Option B
Option C
Moyen de transport
Place Assise
Place assise garantie
Place assise garantie
Temps de déplacement
1H30
1H52
34 minutes
Coût du déplacement
7€ / déplacement
5€ / déplacement
5€ / déplacement
Cochez l'option choisie :
Question suivante
Cas 1.4 : Je choisis l'option A
Cas 1.5 : Je choisis l'option A
Descriptif de l'exercice 2.
Cas 2.1 : Je choisis l'option A
Cas 2.2 : Je choisis l'option A
Cas 2.3 : Je choisis l'option A
Cas 2.4 : Je choisis l'option A
Cas 2.5 : Je choisis l'option A
Descriptif de l'exercice 3.
Cas 3.1 : Je choisis l'option A
Cas 3.2 : Je choisis l'option A
Cas 3.3 : Je choisis l'option A
Cas 3.4 : Je choisis l'option A
Pouvez-vous expliquer pourquoi vous avez systématiquement choisi l'option A ?
Choix systématique de l'option A :
Question suivante

Cas d'un déplacement en car



VOS PRÉFÉRENCES

Choix entre 3 options - première série

Imaginez que vous ayez 3 options : deux en car et une en voiture pour effectuer le déplacement de Lucenay (69) à Vaulx-en-Velin (69). Merci de nous indiquer laquelle vous choisiriez si ces trois options étaient à votre disposition.

Votre réponse doit être faite en vous plaçant dans la situation du déplacement du 20/02/2015 (même motif de déplacement, même contraintes horaires...). L'ensemble du déplacement est pris en compte, de Lucenay (69) à Vaulx-en-Velin (69) (y compris les trajets pour accéder à l'arrêt de départ et depuis l'arrêt d'arrivée jusqu'à la destination finale).

Les caractéristiques proposées pour ce déplacement varient entre les questions et peuvent différer de la réalité. Nous vous demandons de vous baser sur les valeurs indiquées dans les questions et non sur la situation vécue le 20/02/2015 pour faire vos choix.

Selon quelles caractéristiques faire vos choix ?

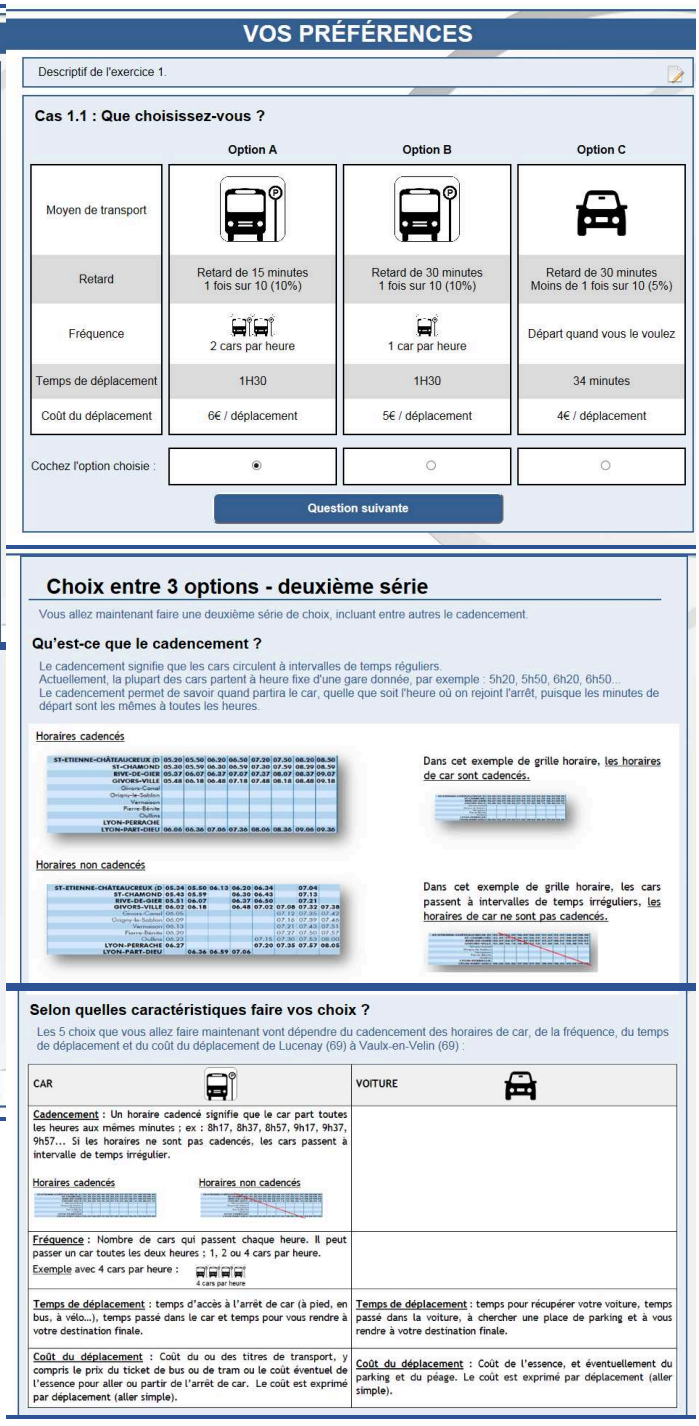
Les 5 premiers choix que vous allez faire vont dépendre du retard éventuel, de la fréquence de passage des cars, du temps de déplacement et du coût du déplacement de Lucenay (69) à Vaulx-en-Velin (69) :

CAR	VOITURE
Retard : Un retard de 15 minutes 1 fois sur 10 signifie que tous les 10 déplacements, vous aurez 15 minutes de retard à cause d'accidents ou de bouchons en voiture, de problèmes techniques en car ou pour toutes autres raisons. Les autres fois, vous serez à l'heure Le retard est donc défini par **le nombre de retards** : moins d'1 car sur 10 (5%), 1 car sur 10 (10%) ou 2 cars sur 10 (20%) et la **durée de retard** : 10, 15 ou 30 minutes.	
Fréquence : Nombre de cars qui passent chaque heure. Il peut passer un car toutes les deux heures ; 1, 2 ou 4 cars par heure. Exemple avec 4 cars par heure : 4 cars par heure	
Temps de déplacement : temps d'accès à l'arrêt de car (à pied, en bus, à vélo...), temps passé dans le car et temps pour vous rendre à votre destination finale.	**Temps de déplacement** : temps pour récupérer votre voiture, temps passé dans la voiture, à chercher une place de parking et à vous rendre à votre destination finale.
Coût du déplacement : Coût du ou des titres de transport, y compris le prix du ticket de bus ou de tram ou le coût éventuel de l'essence pour aller ou partir de l'arrêt de car. Le coût est exprimé par déplacement (aller simple).	**Coût du déplacement** : Coût de l'essence, et éventuellement du parking et du péage. Le coût est exprimé par déplacement (aller simple).

Exemple de réponse

	Option A	Option B	Option C
Moyen de transport	Si vous choisissez l'option B, vous choisissez de vous déplacer en car.		
Retard	9 fois sur 10 vous êtes à l'heure. Mais 1 fois sur 10, vous êtes en retard de 10 minutes.	Retard de 15 minutes 1 fois sur 10 (10%)	Retard de 10 minutes Moins de 1 fois sur 10 (5%)
Fréquence	Il passe 4 cars par heure.	4 cars par heure	Départ quand vous le voulez
Temps de déplacem	Le déplacement dure 1H54 et vous coûte 30 euros par déplacement.	1H54	1H30
Coût du déplacement		30€ / déplacement	16€ / déplacement
Cochez l'option choisie :	○	◉	○

Question suivante



VOS PRÉFÉRENCES

Descriptif de l'exercice 1.

Cas 1.1 : Que choisissez-vous ?

	Option A	Option B	Option C
Moyen de transport			
Retard	Retard de 15 minutes 1 fois sur 10 (10%)	Retard de 30 minutes 1 fois sur 10 (10%)	Retard de 30 minutes Moins de 1 fois sur 10 (5%)
Fréquence	2 cars par heure	1 car par heure	Départ quand vous le voulez
Temps de déplacement	1H30	1H30	34 minutes
Coût du déplacement	6€ / déplacement	5€ / déplacement	4€ / déplacement
Cochez l'option choisie :	◉	○	○

Question suivante

Choix entre 3 options - deuxième série

Vous allez maintenant faire une deuxième série de choix, incluant entre autres le cadencement.

Qu'est-ce que le cadencement ?

Le cadencement signifie que les cars circulent à intervalles de temps réguliers.
Actuellement, la plupart des cars partent à heure fixe d'une gare donnée, par exemple : 5h20, 5h50, 6h20, 6h50...
Le cadencement permet de savoir quand partira le car, quelle que soit l'heure où on rejoint l'arrêt, puisque les minutes de départ sont les mêmes à toutes les heures.

Horaires cadencés

Dans cet exemple de grille horaire, les horaires de car sont cadencés.

Horaires non cadencés

Dans cet exemple de grille horaire, les cars passent à intervalles de temps irréguliers, les horaires de car ne sont pas cadencés.

Selon quelles caractéristiques faire vos choix ?

Les 5 choix que vous allez faire maintenant vont dépendre du cadencement des horaires de car, de la fréquence, du temps de déplacement et du coût du déplacement de Lucenay (69) à Vaulx-en-Velin (69) :

CAR	VOITURE
Cadencement : Un horaire cadencé signifie que le car part toutes les heures aux mêmes minutes ; ex : 8h17, 8h37, 8h57, 9h17, 9h37, 9h57... Si les horaires ne sont pas cadencés, les cars passent à intervalle de temps irrégulier. Horaires cadencés Horaires non cadencés	
Fréquence : Nombre de cars qui passent chaque heure. Il peut passer un car toutes les deux heures ; 1, 2 ou 4 cars par heure. Exemple avec 4 cars par heure : 4 cars par heure	
Temps de déplacement : temps d'accès à l'arrêt de car (à pied, en bus, à vélo...), temps passé dans le car et temps pour vous rendre à votre destination finale.	**Temps de déplacement** : temps pour récupérer votre voiture, temps passé dans la voiture, à chercher une place de parking et à vous rendre à votre destination finale.
Coût du déplacement : Coût du ou des titres de transport, y compris le prix du ticket de bus ou de tram ou le coût éventuel de l'essence pour aller ou partir de l'arrêt de car. Le coût est exprimé par déplacement (aller simple).	**Coût du déplacement** : Coût de l'essence, et éventuellement du parking et du péage. Le coût est exprimé par déplacement (aller simple).

Exemple de choix à faire

	Option A	Option B	Option C
Moyen de transport			
Cadencement	Cadencés	Cadencés	
Fréquence	2 cars par heure	1 car par heure	Départ quand vous le voulez
Temps de déplacement	2H33	1H54	1H30
Coût du déplacement	24€ / déplacement	30€ / déplacement	21€ / déplacement
Cochez l'option choisie :	○	◉	○

Exemple de réponse

	Option A	Option B	Option C
Moyen de transpo			
Cadencement		Cadencés	
Fréquence		1 car par heure	Départ quand vous le voulez
Temps de déplacement	2H33	1H54	1H30
Coût du déplacement	24€ / déplacement	30€ / déplacement	21€ / déplacement
Cochez l'option choisie :	○	◉	○

Si vous choisissez l'option B, vous choisissez de vous déplacer en car.

Les horaires sont cadencés

Il passe 1 car par heure.

Le déplacement dure 1H54 et vous coûte 30 euros par déplacement.

Cas 1.3 :	Je choisis l'option A
Cas 1.4 :	Je choisis l'option A
Cas 1.5 :	Je choisis l'option A
Descriptif de l'exercice 2.	

Cas 2.1 : Que choisissez-vous ?

	Option A	Option B	Option C
Moyen de transport			
Cadencement	Cadencés	Cadencés	
Fréquence	2 cars par heure	1 car toutes les 2 heures	Départ quand vous le voulez
Temps de déplacement	1H30	1H52	34 minutes
Coût du déplacement	6€ / déplacement	5€ / déplacement	6€ / déplacement
Cochez l'option choisie :	◉	○	○

Choix entre 3 options - dernière série

Vous allez maintenant faire une dernière série de choix avec une option en train, une option en car et une option en voiture.

Selon quelles caractéristiques faire vos choix ?

Les 4 choix que vous allez faire maintenant vont dépendre de la garantie ou non d'avoir une place assise, du temps de déplacement et du coût du déplacement de Lucenay (69) à Vaulx-en-Velin (69) :

TRAIN	CAR	VOITURE
Place assise : Garantie ou non d'avoir une place assise dans le train. Si la place assise n'est pas garantie, il peut vous arriver de voyager debout sur tout ou une partie du déplacement. La place assise est toujours garantie dans le car. Place assise non garantie : Place assise garantie :		
Temps de déplacement : temps d'accès à la gare ou à l'arrêt de car (à pied, en bus, à vélo...), temps passé dans le train ou le car et temps pour vous rendre à votre destination finale.		Temps de déplacement : temps pour récupérer votre voiture, temps passé dans la voiture, à chercher une place de parking et à vous rendre à votre destination finale.
Coût du déplacement : Coût du ou des titres de transport, y compris le prix du ticket de bus ou de tram ou le coût éventuel de l'essence pour aller ou partir de la gare ou de l'arrêt de car. Le coût est exprimé par déplacement (aller simple).		Coût du déplacement : Coût de l'essence, et éventuellement du parking et du péage. Le coût est exprimé par déplacement (aller simple).

Exemple de choix à faire

	Option A	Option B	Option C
Moyen de transport			
Place Assise	Place assise non garantie	Place assise garantie	
Temps de déplacement	1H48	1H21	1H30
Coût du déplacement	24€ / déplacement	24€ / déplacement	21€ / déplacement
Cochez l'option choisie :	○	◉	○

Exemple de réponse

	Option A	Option B	Option C
Moyen de transport			
Place Assise		Place assise garantie	
Temps de déplaceme		1H21	1H30
Coût du déplacement	24€ / déplacement	24€ / déplacement	21€ / déplacement
Cochez l'option choisie :	○	◉	○

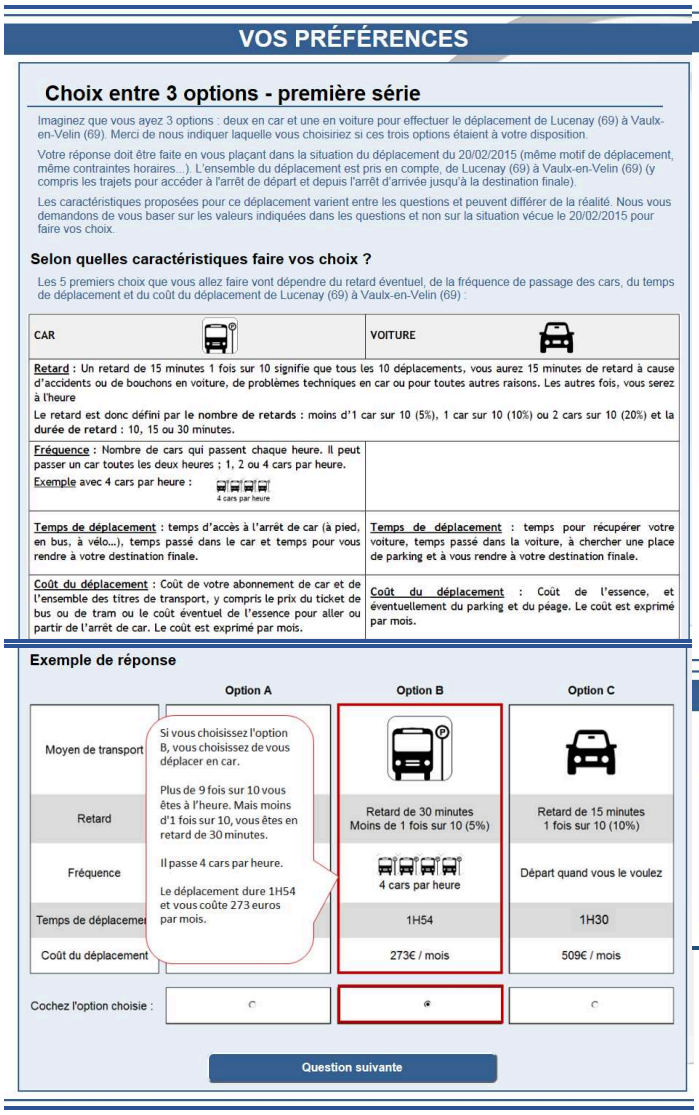
Si vous choisissez l'option B, vous choisissez de vous déplacer en car.

Vous êtes assuré(e) de disposer d'une place assise

Le déplacement dure 1H21 et vous coûte 24 euros par déplacement.

Question suivante

Cas abonnement car



VOS PRÉFÉRENCES

Choix entre 3 options - première série

Imaginez que vous ayez 3 options : deux en car et une en voiture pour effectuer le déplacement de Lucenay (69) à Vaulx-en-Velin (69). Merci de nous indiquer laquelle vous choisiriez si ces trois options étaient à votre disposition.

Votre réponse doit être faite en vous plaçant dans la situation du déplacement du 20/02/2015 (même motif de déplacement, même contraintes horaires...). L'ensemble du déplacement est pris en compte, de Lucenay (69) à Vaulx-en-Velin (69) (y compris les trajets pour accéder à l'arrêt de départ et depuis l'arrêt d'arrivée jusqu'à la destination finale).

Les caractéristiques proposées pour ce déplacement varient entre les questions et peuvent différer de la réalité. Nous vous demandons de vous baser sur les valeurs indiquées dans les questions et non sur la situation vécue le 20/02/2015 pour faire vos choix.

Selon quelles caractéristiques faire vos choix ?

Les 5 premiers choix que vous allez faire vont dépendre du retard éventuel, de la fréquence de passage des cars, du temps de déplacement et du coût du déplacement de Lucenay (69) à Vaulx-en-Velin (69) :

CAR	VOITURE
Retard : Un retard de 15 minutes 1 fois sur 10 signifie que tous les 10 déplacements, vous aurez 15 minutes de retard à cause d'accidents ou de bouchons en voiture, de problèmes techniques en car ou pour toutes autres raisons. Les autres fois, vous serez à l'heure Le retard est donc défini par le nombre de retards : moins d'1 car sur 10 (5%), 1 car sur 10 (10%) ou 2 cars sur 10 (20%) et la durée de retard : 10, 15 ou 30 minutes.	
Fréquence : Nombre de cars qui passent chaque heure. Il peut passer un car toutes les deux heures ; 1, 2 ou 4 cars par heure. Exemple avec 4 cars par heure : 4 cars par heure	
Temps de déplacement : temps d'accès à l'arrêt de car (à pied, en bus, à vélo...), temps passé dans le car et temps pour vous rendre à votre destination finale.	Temps de déplacement : temps pour récupérer votre voiture, temps passé dans la voiture, à chercher une place de parking et à vous rendre à votre destination finale.
Coût du déplacement : Coût de votre abonnement de car et de l'ensemble des titres de transport, y compris le prix du ticket de bus ou de tram ou le coût éventuel de l'essence pour aller ou partir de l'arrêt de car. Le coût est exprimé par mois.	Coût du déplacement : Coût de l'essence, et éventuellement du parking et du péage. Le coût est exprimé par mois.

Exemple de réponse

	Option A	Option B	Option C
Moyen de transport	Si vous choisissez l'option B, vous choisissez de vous déplacer en car.		
Retard	Plus de 9 fois sur 10 vous êtes à l'heure. Mais moins d'1 fois sur 10, vous êtes en retard de 30 minutes.	Retard de 30 minutes Moins de 1 fois sur 10 (5%)	Retard de 15 minutes 1 fois sur 10 (10%)
Fréquence	Il passe 4 cars par heure.	4 cars par heure	Départ quand vous le voulez
Temps de déplacemer	Le déplacement dure 1H54 et vous coûte 273 euros par mois.	1H54	1H30
Coût du déplacement		273€ / mois	509€ / mois
Cochez l'option choisie :	○	◉	○

Question suivante

PERCEPTION DU DÉPLACEMENT

Pendant ce déplacement, quelles activités particulières avez-vous eues ?

Par exemple : lire, téléphoner, discuter

1 réponse minimum - 3 réponses maximum.

- ☐ Rien d'autre que des activités nécessaires au transport
- ☐ Travailler, étudier
- ☐ Lire, regarder un film, une vidéo
- ☐ Téléphoner, envoyer des messages
- ☐ Discuter avec d'autres personnes
- ☐ Jouer seul ou avec d'autres personnes ou avoir une activité manuelle
- ☐ Ecouter de la musique ou la radio
- ☐ Penser, rester seul
- ☐ Regarder le paysage, les vitrines, les gens
- ☐ Manger, boire, fumer
- ☐ Dormir, somnoler
- ☐ Autre
- ☐ Non réponse

Question suivante

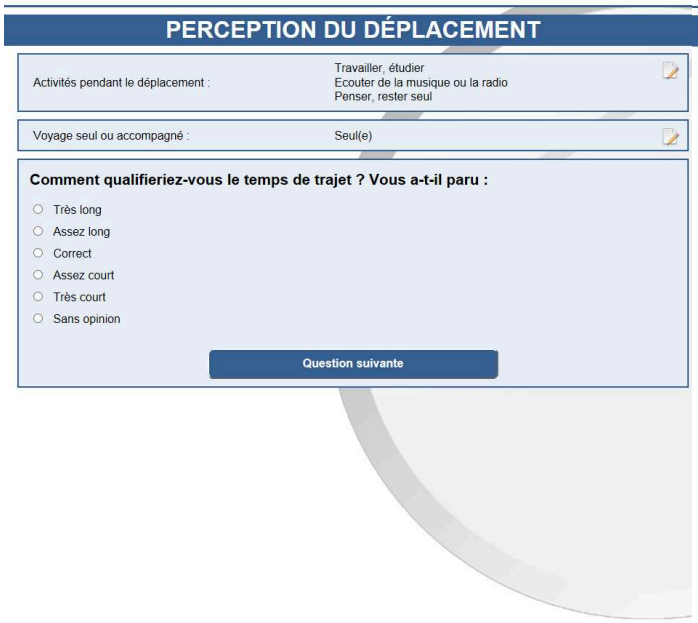
PERCEPTION DU DÉPLACEMENT

Activités pendant le déplacement : Travailler, étudier
Ecouter de la musique ou la radio
Penser, rester seul

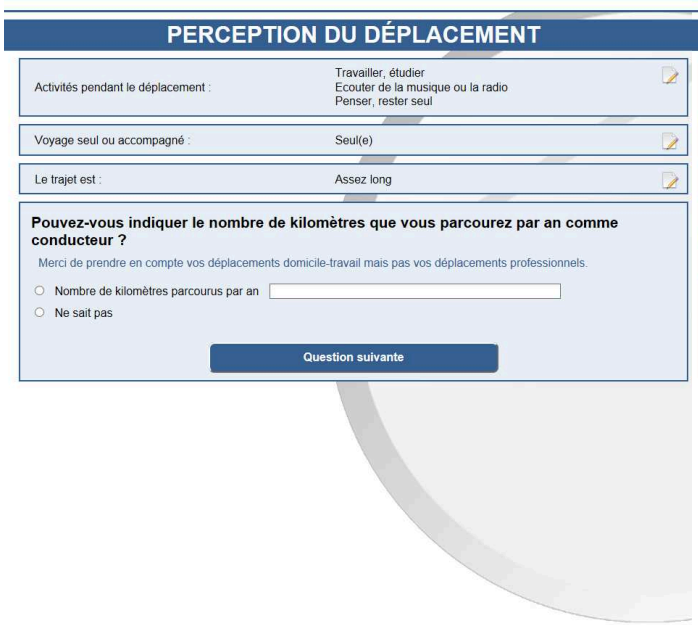
Avez-vous voyagé seul(e) ou accompagné(e) ?

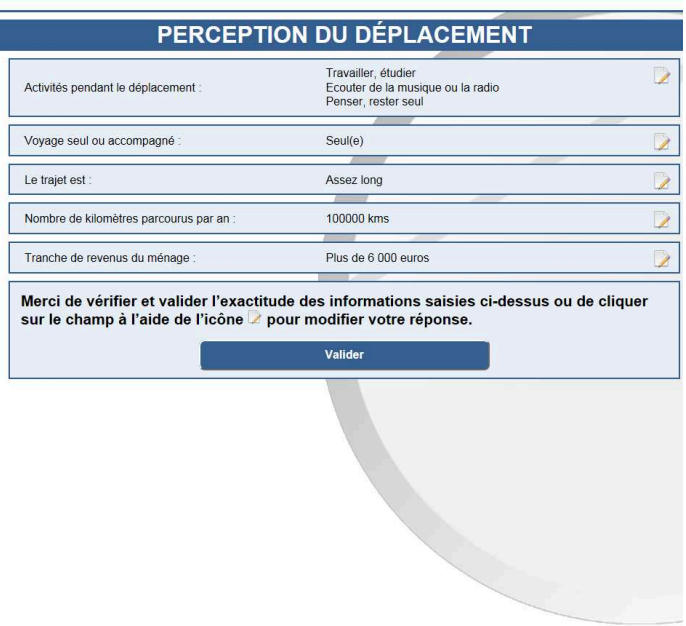
- ○ Seul(e)
- ○ Accompagné(e)

Question suivante


PERCEPTION DU DÉPLACEMENT
Activités pendant le déplacement :
Travailler, étudier
Ecouter de la musique ou la radio
Penser, rester seul
Voyage seul ou accompagné :
Seul(e)
Comment qualifieriez-vous le temps de trajet ? Vous a-t-il paru :
Très long
Assez long
Correct
Assez court
Très court
Sans opinion
Question suivante

PERCEPTION DU DÉPLACEMENT
Activités pendant le déplacement :
Travailler, étudier
Ecouter de la musique ou la radio
Penser, rester seul
Voyage seul ou accompagné :
Seul(e)
Le trajet est :
Assez long
Nombre de kilomètres parcourus par an :
100000 kms
Pouvez-vous nous indiquer dans quelle tranche se situe le montant des revenus mensuels nets de votre ménage ?
Merci de prendre en compte les primes et le 13ème mois éventuels, les revenus annexes, l'ensemble des prestations sociales, etc.
Le traitement des données se fait de façon anonyme et confidentielle. Votre réponse est utile pour analyser l'ensemble des réponses.
Moins de 1 000 euros
Entre 1 000 et 1 500 euros
Entre 1 500 et 2 000 euros
Entre 2 000 et 3 000 euros
Entre 3 000 et 4 000 euros
Entre 4 000 et 5 000 euros
Entre 5 000 et 6 000 euros
Plus de 6 000 euros
Ne souhaite pas répondre
Question suivante


PERCEPTION DU DÉPLACEMENT
Activités pendant le déplacement :
Travailler, étudier
Ecouter de la musique ou la radio
Penser, rester seul
Voyage seul ou accompagné :
Seul(e)
Le trajet est :
Assez long
Pouvez-vous indiquer le nombre de kilomètres que vous parcourez par an comme conducteur ?
Merci de prendre en compte vos déplacements domicile-travail mais pas vos déplacements professionnels.
Nombre de kilomètres parcourus par an
Ne sait pas
Question suivante


PERCEPTION DU DÉPLACEMENT
Activités pendant le déplacement :
Travailler, étudier
Ecouter de la musique ou la radio
Penser, rester seul
Voyage seul ou accompagné :
Seul(e)
Le trajet est :
Assez long
Nombre de kilomètres parcourus par an :
100000 kms
Tranche de revenus du ménage :
Plus de 6 000 euros
Merci de vérifier et valider l'exactitude des informations saisies ci-dessus ou de cliquer sur le champ à l'aide de l'icône pour modifier votre réponse.
Valider

QUESTIONS OPINION

Pour chacune des affirmations suivantes, merci d'indiquer dans quelle mesure vous êtes d'accord ou non.

	Pas du tout d'accord	Pas d'accord	Neutre	Plutôt d'accord	Complètement d'accord	Non réponse
L'équilibre de la nature est fragile	○	○	◉	○	○	○
Chacun doit être libre d'utiliser sa voiture quand il veut et où il veut	○	○	◉	○	○	○
Nous avons besoin de construire plus de parkings en centre-ville	○	○	◉	○	○	○
Le prix de l'essence devrait être augmenté pour réduire les bouchons et la pollution	○	○	◉	○	○	○
Les pouvoirs publics ont le droit de réduire l'utilisation de la voiture pour préserver l'environnement et la qualité de vie	○	○	◉	○	○	○
Nous allons au-devant d'une catastrophe écologique majeure	○	○	◉	○	○	○
La soi-disant « crise écologique » et ses conséquences pour l'Homme sont très exagérées	○	○	◉	○	○	○
Nous avons besoin de plus de transports en commun même si cela implique d'augmenter les impôts	○	○	◉	○	○	○

Question suivante

Nous allons au-devant d'une catastrophe écologique majeure : Neutre
L'équilibre de la nature est fragile : Neutre
La soi-disant « crise écologique » et ses conséquences pour l'Homme sont très exagérées : Neutre
Nous avons besoin de plus de transports en commun même si cela implique d'augmenter les impôts : Neutre
Les pouvoirs publics ont le droit de réduire l'utilisation de la voiture pour préserver l'environnement et la qualité de vie : Neutre
Nous avons besoin de construire plus de parkings en centre-ville : Neutre
Le prix de l'essence devrait être augmenté pour réduire les bouchons et la pollution : Neutre
Chacun doit être libre d'utiliser sa voiture quand il veut et où il veut : Neutre

L'insécurité routière : Neutre
La qualité de vie amoindrie dans les villes du fait du bruit du trafic et des odeurs : Neutre
La pollution de l'air : Neutre
L'utilisation de l'espace par la voiture au détriment des cyclistes, piétons et enfants : Neutre

Pour chacune des affirmations suivantes, merci d'indiquer dans quelle mesure vous êtes d'accord ou non.

	Pas du tout d'accord	Pas d'accord	Neutre	Plutôt d'accord	Complètement d'accord	Non réponse
J'aime conduire vite	○	○	○	○	○	○
Quand je conduis, j'ai peur du risque d'accident	○	○	○	○	○	○
Je trouve difficile de prendre les transports en commun quand j'ai des bagages ou des courses à transporter	○	○	○	○	○	○
Je n'accorde pas d'importance au type de voiture que je conduis	○	○	○	○	○	○
Que ce soit en voiture ou dans les transports en commun, je n'aime pas voyager avec des personnes que je ne connais pas	○	○	○	○	○	○
Je me sens en sécurité dans ma voiture	○	○	○	○	○	○
Je trouve difficile de prendre les transports en commun quand je voyage avec des enfants	○	○	○	○	○	○
J'aime ma voiture	○	○	○	○	○	○
Votre voiture reflète votre succès	○	○	○	○	○	○
La voiture est un objet qui permet de montrer aux autres ce que vous êtes, ainsi que vos goûts	○	○	○	○	○	○

Question suivante

QUESTIONS OPINION

Nous allons au-devant d'une catastrophe écologique majeure : Neutre
L'équilibre de la nature est fragile : Neutre
La soi-disant « crise écologique » et ses conséquences pour l'Homme sont très exagérées : Neutre
Nous avons besoin de plus de transports en commun même si cela implique d'augmenter les impôts : Neutre
Les pouvoirs publics ont le droit de réduire l'utilisation de la voiture pour préserver l'environnement et la qualité de vie : Neutre
Nous avons besoin de construire plus de parkings en centre-ville : Neutre
Le prix de l'essence devrait être augmenté pour réduire les bouchons et la pollution : Neutre
Chacun doit être libre d'utiliser sa voiture quand il veut et où il veut : Neutre

Merci d'indiquer dans quelle mesure vous vous sentez concerné(e) par chacun des problèmes suivants causés par l'utilisation de la voiture :

	Pas du tout concerné (e)	Pas vraiment concerné (e)	Neutre	Plutôt concerné (e)	Complètement concerné(e)	Non réponse
La qualité de vie amoindrie dans les villes du fait du bruit du trafic et des odeurs	○	○	○	○	○	○
La pollution de l'air	○	○	○	○	○	○
L'insécurité routière	○	○	○	○	○	○
L'utilisation de l'espace par la voiture au détriment des cyclistes, piétons et enfants	○	○	○	○	○	○

Question suivante

Nous avons besoin de construire plus de parkings en centre-ville : Neutre
Le prix de l'essence devrait être augmenté pour réduire les bouchons et la pollution : Neutre
Chacun doit être libre d'utiliser sa voiture quand il veut et où il veut : Neutre

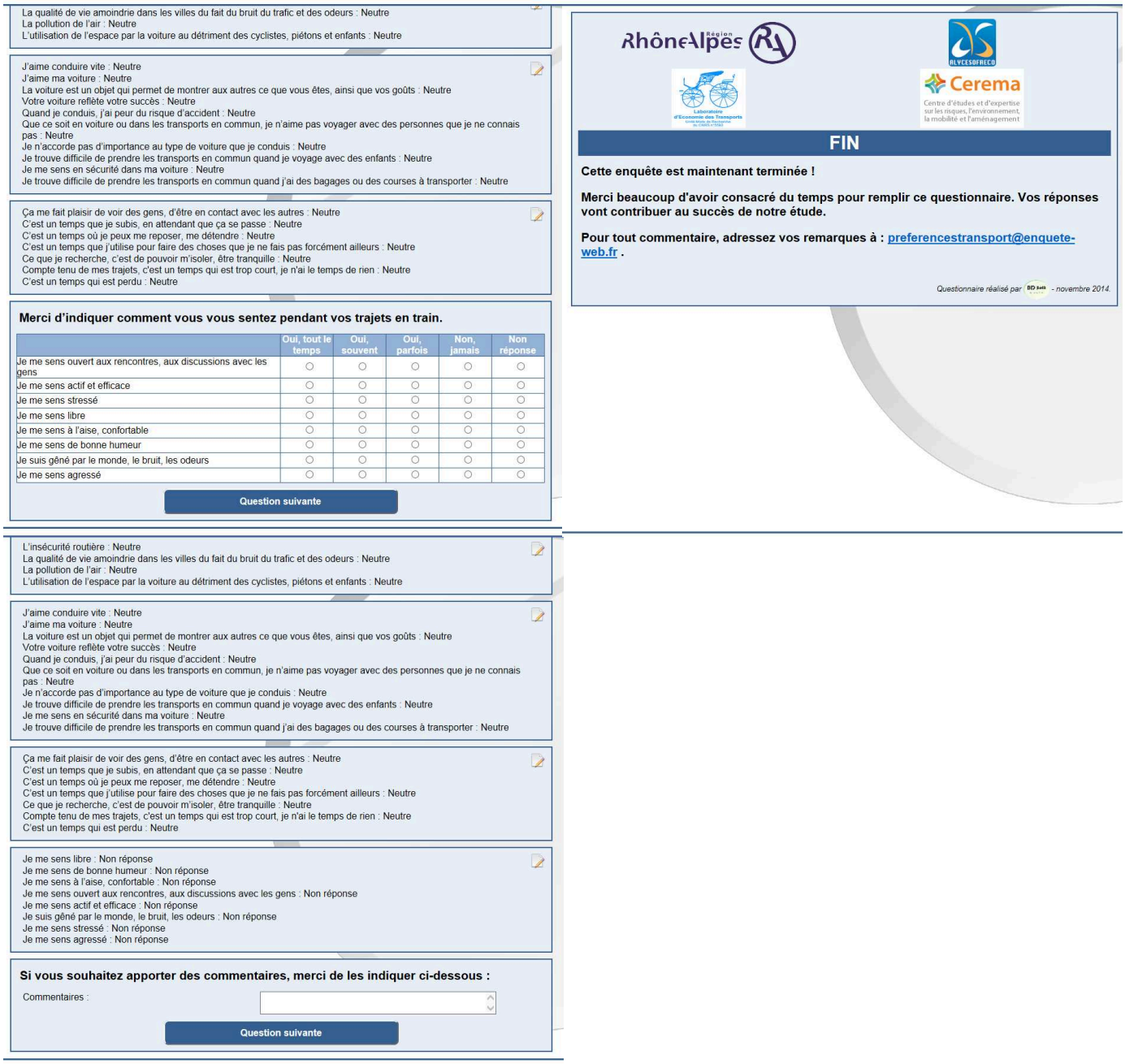
L'insécurité routière : Neutre
La qualité de vie amoindrie dans les villes du fait du bruit du trafic et des odeurs : Neutre
La pollution de l'air : Neutre
L'utilisation de l'espace par la voiture au détriment des cyclistes, piétons et enfants : Neutre

J'aime conduire vite : Neutre
J'aime ma voiture : Neutre
La voiture est un objet qui permet de montrer aux autres ce que vous êtes, ainsi que vos goûts : Neutre
Votre voiture reflète votre succès : Neutre
Quand je conduis, j'ai peur du risque d'accident : Neutre
Que ce soit en voiture ou dans les transports en commun, je n'aime pas voyager avec des personnes que je ne connais pas : Neutre
Je n'accorde pas d'importance au type de voiture que je conduis : Neutre
Je trouve difficile de prendre les transports en commun quand je voyage avec des enfants : Neutre
Je me sens en sécurité dans ma voiture : Neutre
Je trouve difficile de prendre les transports en commun quand j'ai des bagages ou des courses à transporter : Neutre

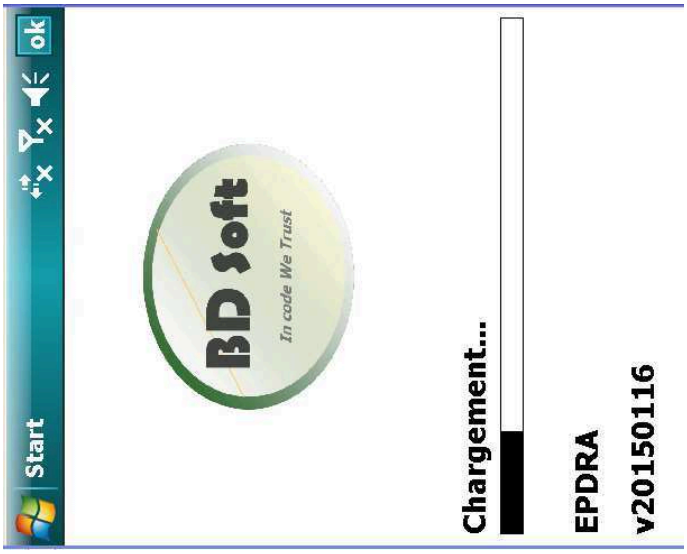
Voici des phrases qu'on pourrait dire sur les trajets en train. Merci d'indiquer dans quelle mesure vous êtes d'accord ou non.

	Pas du tout d'accord	Pas d'accord	Neutre	Plutôt d'accord	Complètement d'accord	Non réponse
C'est un temps qui est perdu	○	○	○	○	○	○
C'est un temps que j'utilise pour faire des choses que je ne fais pas forcément ailleurs	○	○	○	○	○	○
Ce que je recherche, c'est de pouvoir m'isoler, être tranquille	○	○	○	○	○	○
Compte tenu de mes trajets, c'est un temps qui est trop court, je n'ai le temps de rien	○	○	○	○	○	○
C'est un temps où je peux me reposer, me détendre	○	○	○	○	○	○
C'est un temps que je subis, en attendant que ça se passe	○	○	○	○	○	○
Ça me fait plaisir de voir des gens, d'être en contact avec les autres	○	○	○	○	○	○

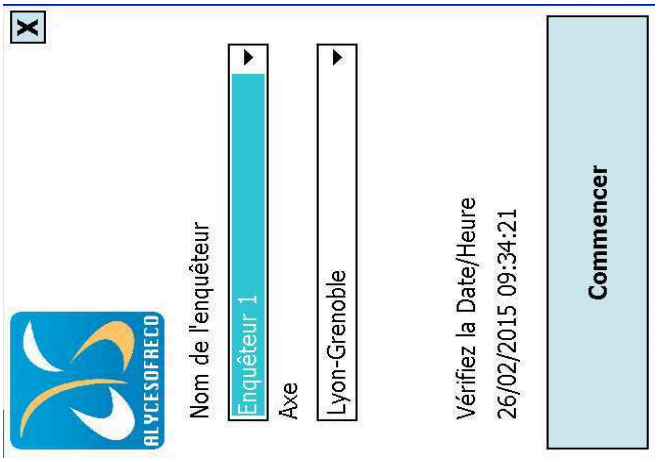
Question suivante

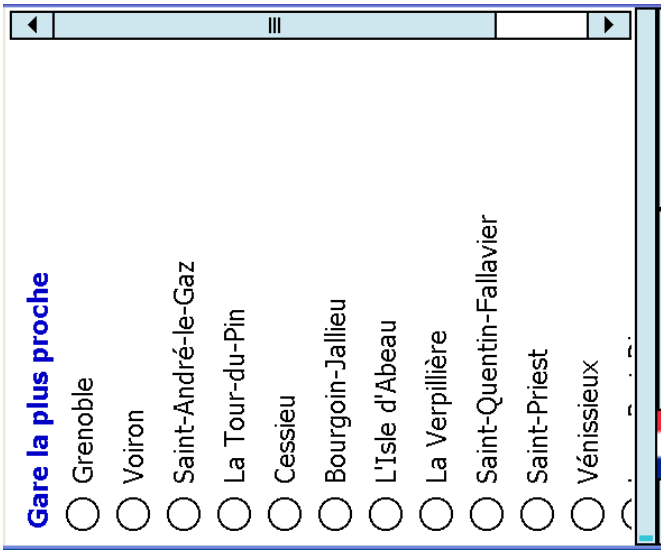

La qualité de vie amoindrie dans les villes du fait du bruit du trafic et des odeurs : Neutre
La pollution de l'air : Neutre
L'utilisation de l'espace par la voiture au détriment des cyclistes, piétons et enfants : Neutre
J'aime conduire vite : Neutre
J'aime ma voiture : Neutre
La voiture est un objet qui permet de montrer aux autres ce que vous êtes, ainsi que vos goûts : Neutre
Votre voiture reflète votre succès : Neutre
Quand je conduis, j'ai peur du risque d'accident : Neutre
Que ce soit en voiture ou dans les transports en commun, je n'aime pas voyager avec des personnes que je ne connais pas : Neutre
Je n'accorde pas d'importance au type de voiture que je conduis : Neutre
Je trouve difficile de prendre les transports en commun quand je voyage avec des enfants : Neutre
Je me sens en sécurité dans ma voiture : Neutre
Je trouve difficile de prendre les transports en commun quand j'ai des bagages ou des courses à transporter : Neutre
Ça me fait plaisir de voir des gens, d'être en contact avec les autres : Neutre
C'est un temps que je subis, en attendant que ça se passe : Neutre
C'est un temps où je peux me reposer, me détendre : Neutre
C'est un temps que j'utilise pour faire des choses que je ne fais pas forcément ailleurs : Neutre
Ce que je recherche, c'est de pouvoir m'isoler, être tranquille : Neutre
Compte tenu de mes trajets, c'est un temps qui est trop court, je n'ai le temps de rien : Neutre
C'est un temps qui est perdu : Neutre
Merci d'indiquer comment vous vous sentez pendant vos trajets en train.
Oui, tout le temps
Oui, souvent
Oui, parfois
Non, jamais
Non réponse
Je me sens ouvert aux rencontres, aux discussions avec les gens
Je me sens actif et efficace
Je me sens stressé
Je me sens libre
Je me sens à l'aise, confortable
Je me sens de bonne humeur
Je suis gêné par le monde, le bruit, les odeurs
Je me sens agressé
Question suivante
RhôneAlpes Région
ALYCESOFRECO
Cerema
Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement
FIN
Cette enquête est maintenant terminée !
Merci beaucoup d'avoir consacré du temps pour remplir ce questionnaire. Vos réponses vont contribuer au succès de notre étude.
Pour tout commentaire, adressez vos remarques à : preferencestransport@enquete-web.fr .
Questionnaire réalisé par - novembre 2014.
L'insécurité routière : Neutre
La qualité de vie amoindrie dans les villes du fait du bruit du trafic et des odeurs : Neutre
La pollution de l'air : Neutre
L'utilisation de l'espace par la voiture au détriment des cyclistes, piétons et enfants : Neutre
J'aime conduire vite : Neutre
J'aime ma voiture : Neutre
La voiture est un objet qui permet de montrer aux autres ce que vous êtes, ainsi que vos goûts : Neutre
Votre voiture reflète votre succès : Neutre
Quand je conduis, j'ai peur du risque d'accident : Neutre
Que ce soit en voiture ou dans les transports en commun, je n'aime pas voyager avec des personnes que je ne connais pas : Neutre
Je n'accorde pas d'importance au type de voiture que je conduis : Neutre
Je trouve difficile de prendre les transports en commun quand je voyage avec des enfants : Neutre
Je me sens en sécurité dans ma voiture : Neutre
Je trouve difficile de prendre les transports en commun quand j'ai des bagages ou des courses à transporter : Neutre
Ça me fait plaisir de voir des gens, d'être en contact avec les autres : Neutre
C'est un temps que je subis, en attendant que ça se passe : Neutre
C'est un temps où je peux me reposer, me détendre : Neutre
C'est un temps que j'utilise pour faire des choses que je ne fais pas forcément ailleurs : Neutre
Ce que je recherche, c'est de pouvoir m'isoler, être tranquille : Neutre
Compte tenu de mes trajets, c'est un temps qui est trop court, je n'ai le temps de rien : Neutre
C'est un temps qui est perdu : Neutre
Je me sens libre : Non réponse
Je me sens de bonne humeur : Non réponse
Je me sens à l'aise, confortable : Non réponse
Je me sens ouvert aux rencontres, aux discussions avec les gens : Non réponse
Je me sens actif et efficace : Non réponse
Je suis gêné par le monde, le bruit, les odeurs : Non réponse
Je me sens stressé : Non réponse
Je me sens agressé : Non réponse
Si vous souhaitez apporter des commentaires, merci de les indiquer ci-dessous :
Commentaires :
Question suivante

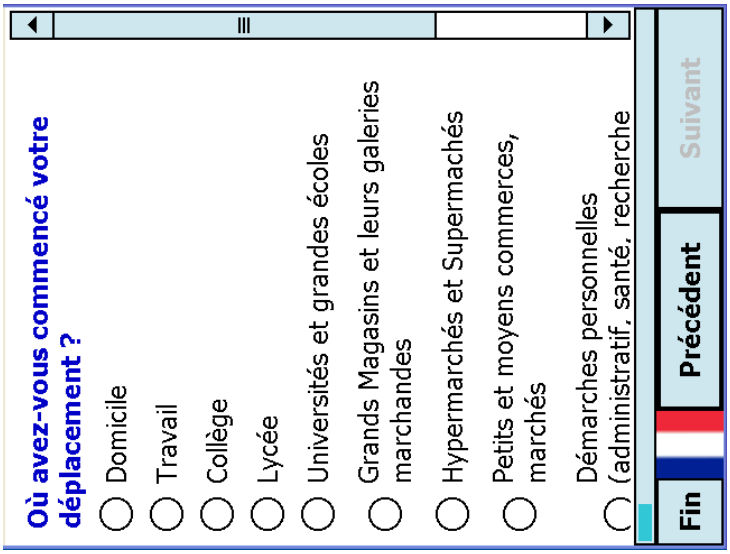
C.2 CAPI



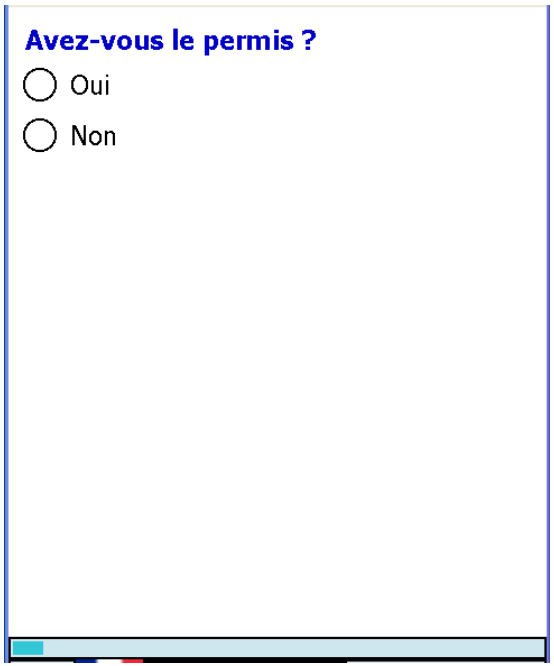
Partie « configuration »

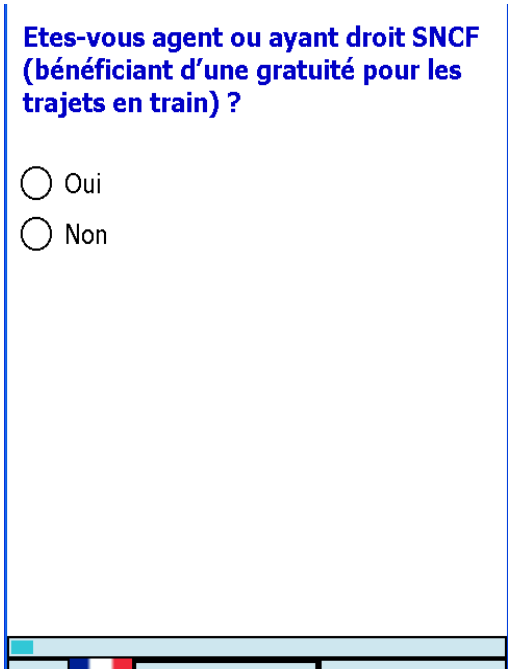


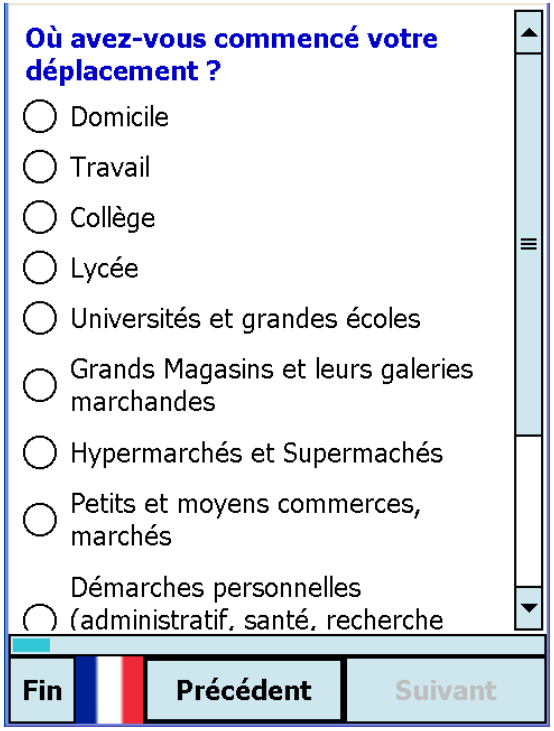


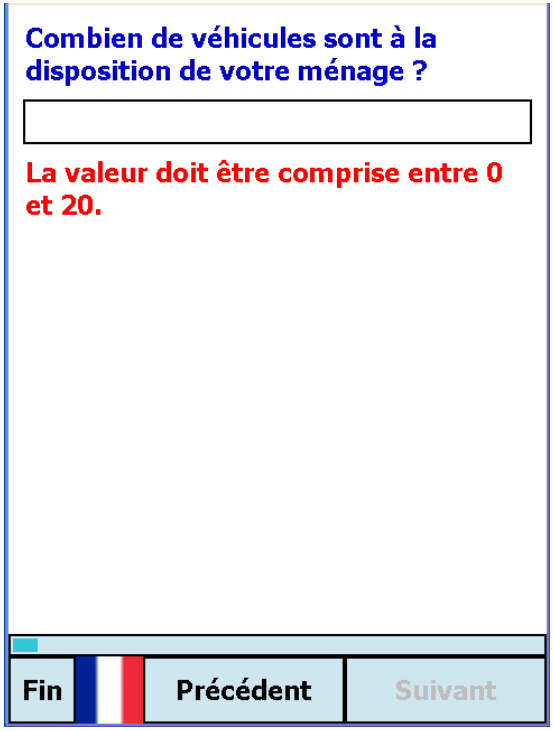


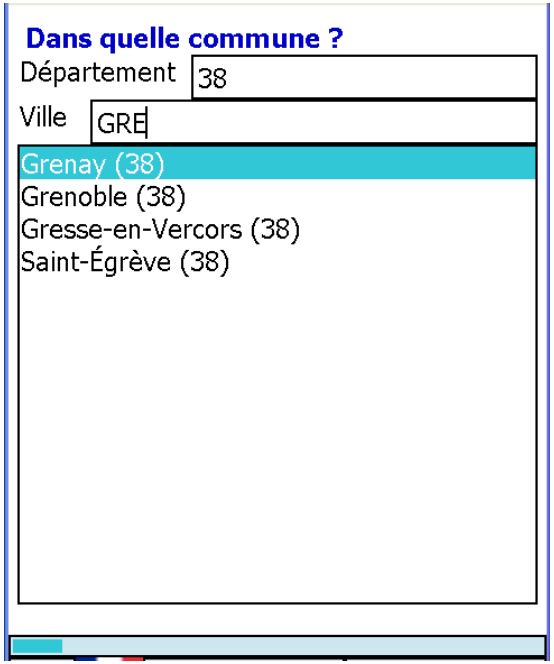
Partie « Initialisation »

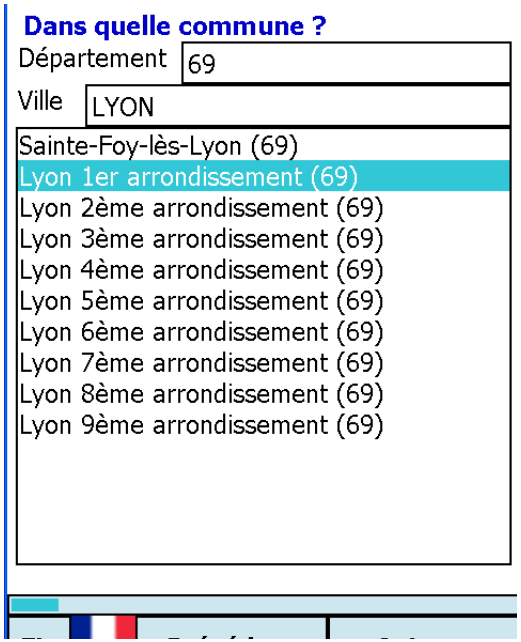

Avez-vous le permis ?
Oui
Non

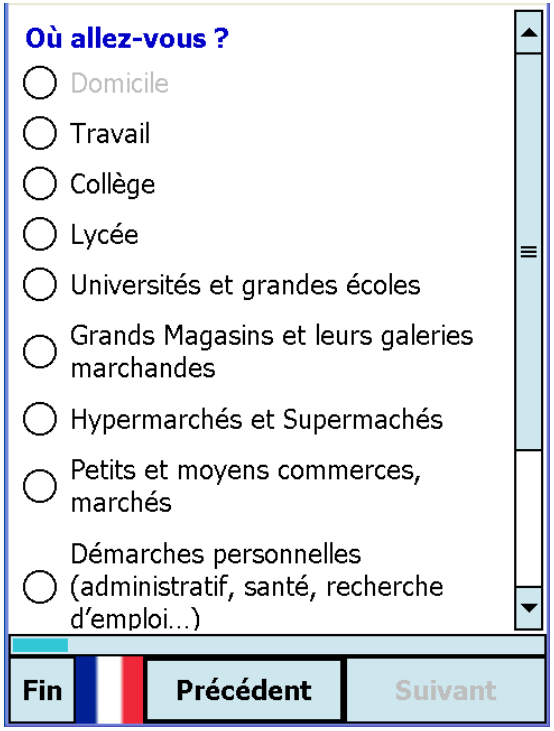

Etes-vous agent ou ayant droit SNCF (bénéficiant d'une gratuité pour les trajets en train) ?
Oui
Non

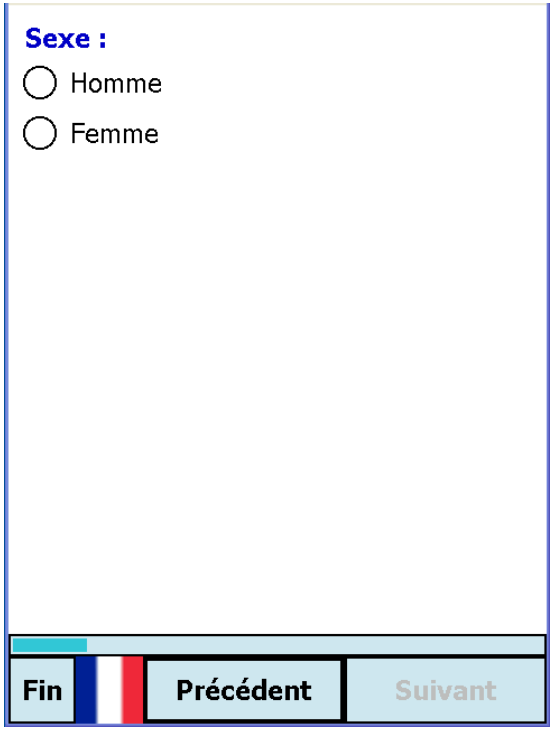

Où avez-vous commencé votre déplacement ?
Domicile
Travail
Collège
Lycée
Universités et grandes écoles
Grands Magasins et leurs galeries marchandes
Hypermarchés et Supermachés
Petits et moyens commerces, marchés
Démarches personnelles (administratif, santé, recherche
Fin
Précédent
Suivant

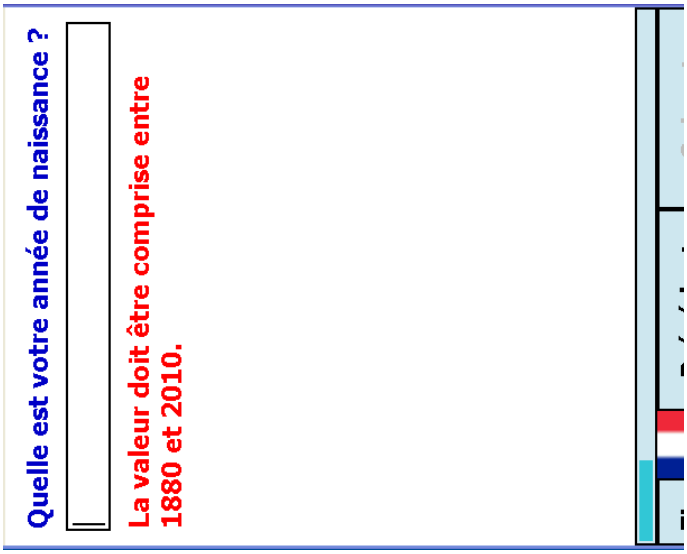

Combien de véhicules sont à la disposition de votre ménage ?
La valeur doit être comprise entre 0 et 20.
Fin
Précédent
Suivant


Dans quelle commune ?
Département 38
Ville GRE
Grenay (38)
Grenoble (38)
Gresse-en-Vercors (38)
Saint-Égrève (38)

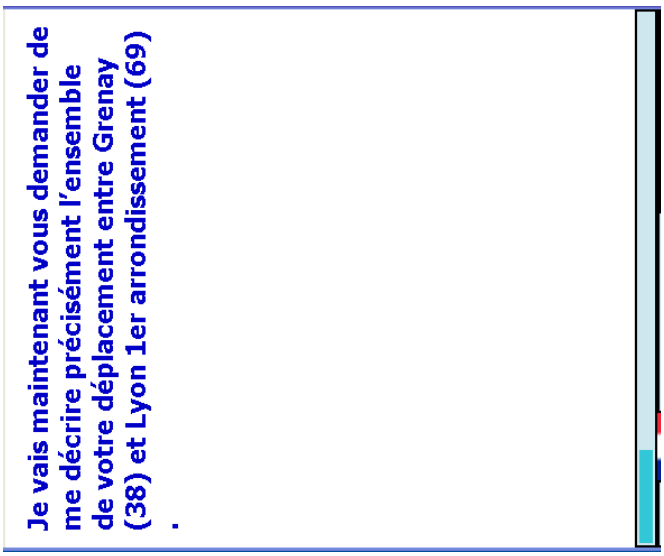

Dans quelle commune ?
Département 69
Ville LYON
Sainte-Foy-lès-Lyon (69)
Lyon 1er arrondissement (69)
Lyon 2ème arrondissement (69)
Lyon 3ème arrondissement (69)
Lyon 4ème arrondissement (69)
Lyon 5ème arrondissement (69)
Lyon 6ème arrondissement (69)
Lyon 7ème arrondissement (69)
Lyon 8ème arrondissement (69)
Lyon 9ème arrondissement (69)

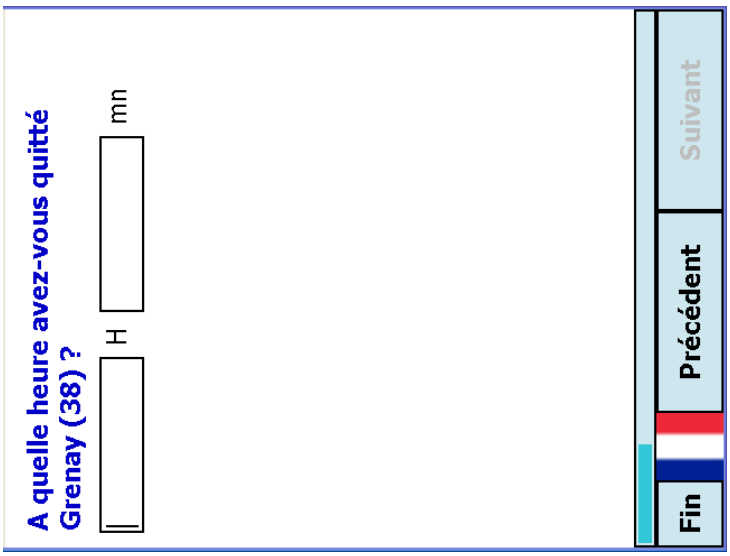

Où allez-vous ?
Domicile
Travail
Collège
Lycée
Universités et grandes écoles
Grands Magasins et leurs galeries marchandes
Hypermarchés et Supermachés
Petits et moyens commerces, marchés
Démarches personnelles (administratif, santé, recherche d'emploi…)
Fin
Précédent
Suivant

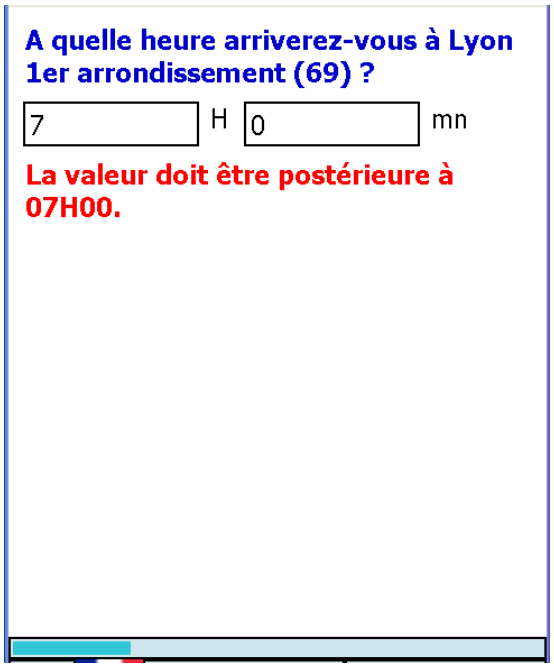

Sexe :
Homme
Femme
Fin
Précédent
Suivant

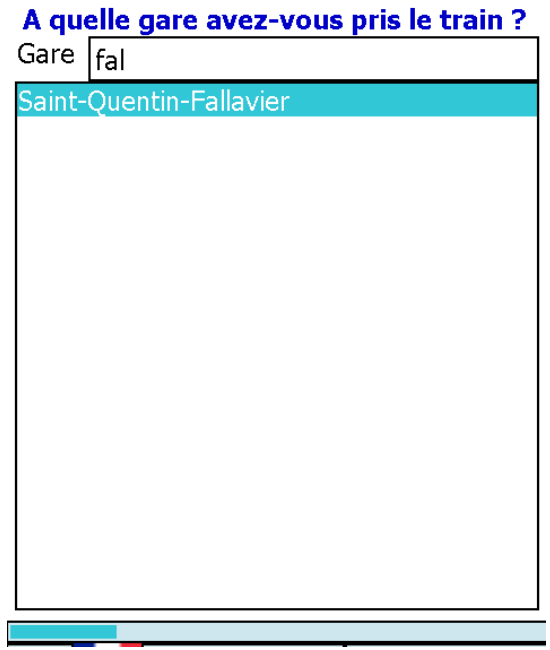


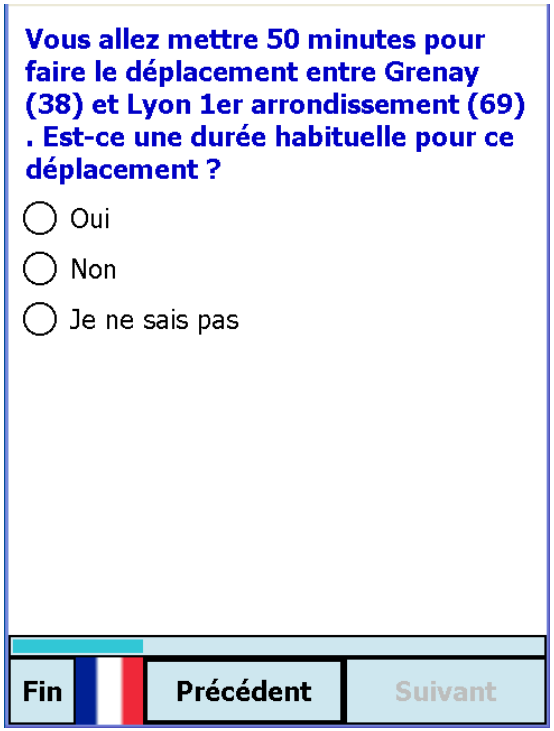
Partie « Description du déplacement »

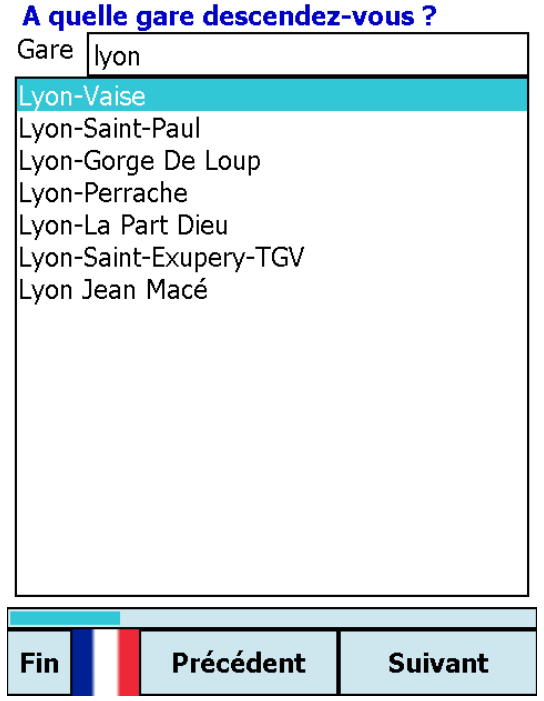





A quelle heure arriverez-vous à Lyon 1er arrondissement (69) ?
7
H
0
mn
La valeur doit être postérieure à 07H00.


A quelle gare avez-vous pris le train ?
Gare
fal
Saint-Quentin-Fallavier


Vous allez mettre 50 minutes pour faire le déplacement entre Grenay (38) et Lyon 1er arrondissement (69) . Est-ce une durée habituelle pour ce déplacement ?
Oui
Non
Je ne sais pas
Fin
Précédent
Suivant


A quelle gare descendez-vous ?
Gare
lyon
Lyon-Vaise
Lyon-Saint-Paul
Lyon-Gorge De Loup
Lyon-Perrache
Lyon-La Part Dieu
Lyon-Saint-Exupery-TGV
Lyon Jean Macé
Fin
Précédent
Suivant

Quel(s) moyen(s) de transport utilisez-vous sur l'ensemble de votre déplacement entre Grenay (38) et Lyon 1er arrondissement (69) ?

- ☐ Marche à pied
- ☐ Vélo
- ☐ Conducteur de véhicule particulier
- ☐ Passager de véhicule particulier
- ☐ Fourgon, camionnette, camion
- ☐ Deux-roues motorisé
- ☐ Transport urbain (métro, tram, bus)
- ☐ Car du département

Combien de temps mettez-vous habituellement pour faire ce déplacement entre Grenay (38) et Lyon 1er arrondissement (69) ?

[ ] H [ ] mn

Pour ce déplacement, utilisez-vous une carte de réduction ou un abonnement ?

- ◯ Pas d'abonnement, ni de carte de réduction
- ◯ Abonnement hebdomadaire
- ◯ Abonnement mensuel
- ◯ Abonnement semestriel (6 mois)
- ◯ Abonnement annuel
- ◯ Carte Illico Liberté
- ◯ Carte Illico Jeune
- ◯ Carte Fréquence 1ère
- ◯ Carte Fréquence 2de

Fin | Précédent | Suivant

Avez-vous un impératif horaire à destination ?

- ◯ Oui
- ◯ Non

Fin | Précédent | Suivant

Combien coûte votre abonnement par mois ?

La valeur doit être comprise entre 1 et 2000.

Combien vous a coûté votre billet de train par personne (pour un seul trajet en aller-simple) ?

La valeur doit être comprise entre 1 et 1000.

Fin | Précédent | Suivant

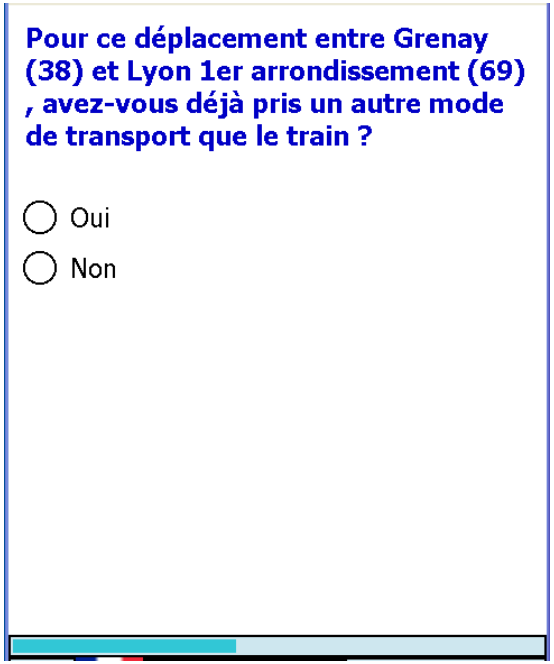
Le prix de cet abonnement de train est-il pris en charge par un tiers (employeur, etc.) ?

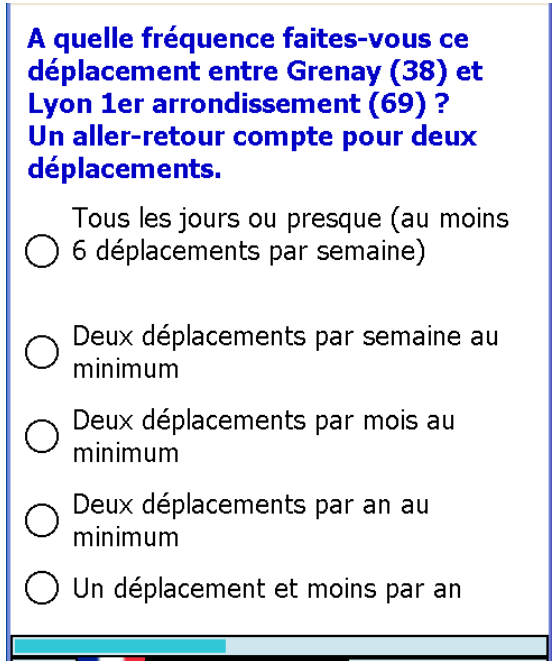
- ◯ Oui, en totalité
- ◯ Oui, en partie
- ◯ Non

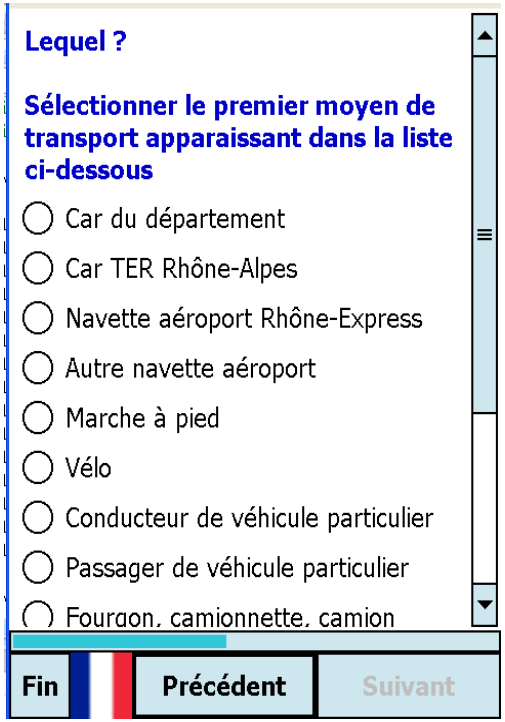
Savez-vous que votre employeur est tenu de vous rembourser la moitié du prix de votre abonnement, si vous utilisez les transports collectifs pour aller au travail ?

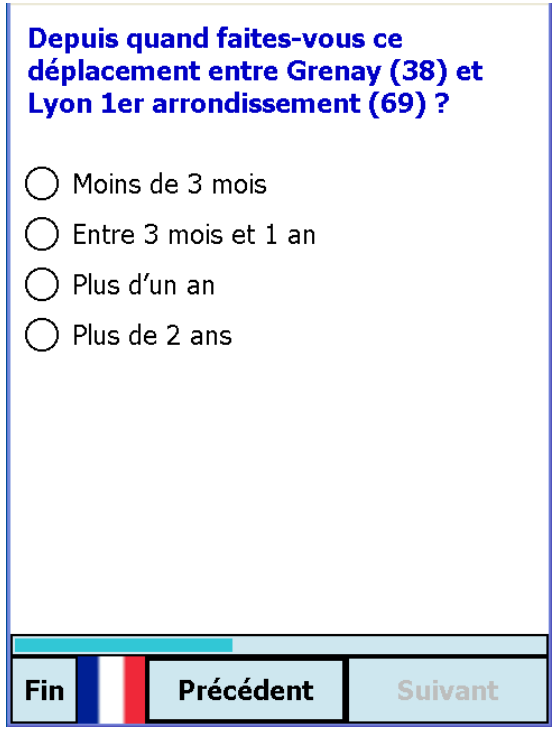
- ◯ Oui, je le sais
- ◯ Non, je ne le sais pas

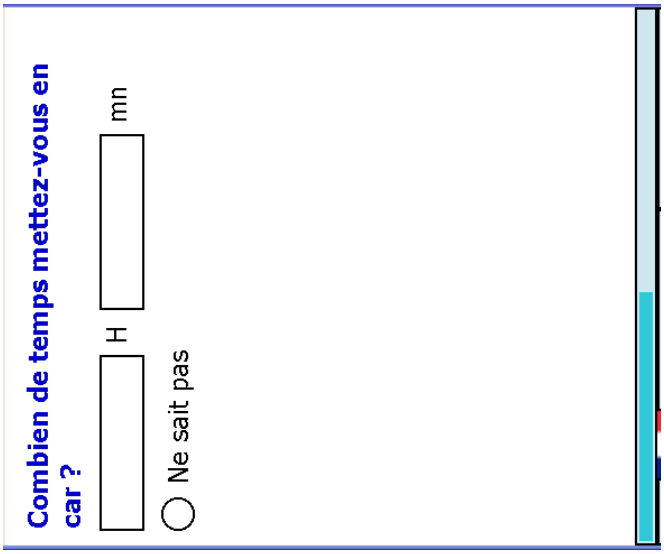
Fin | Précédent | Suivant

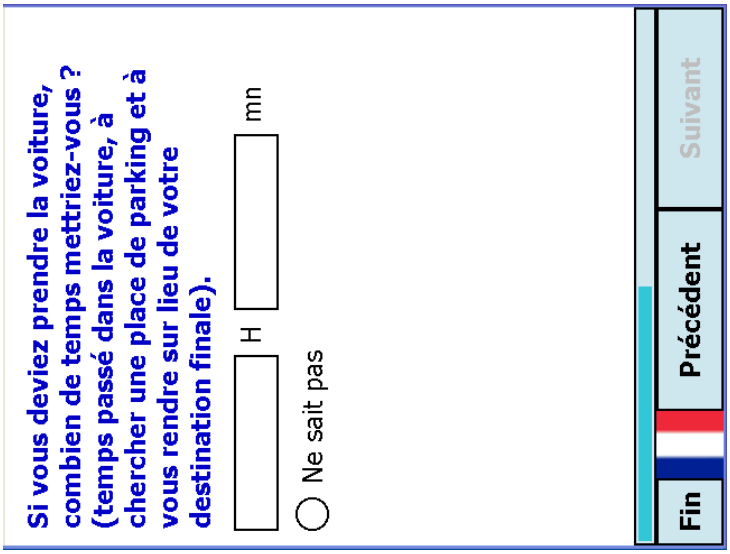

Pour ce déplacement entre Grenay (38) et Lyon 1er arrondissement (69) , avez-vous déjà pris un autre mode de transport que le train ?
Oui
Non


A quelle fréquence faites-vous ce déplacement entre Grenay (38) et Lyon 1er arrondissement (69) ?
Un aller-retour compte pour deux déplacements.
Tous les jours ou presque (au moins 6 déplacements par semaine)
Deux déplacements par semaine au minimum
Deux déplacements par mois au minimum
Deux déplacements par an au minimum
Un déplacement et moins par an

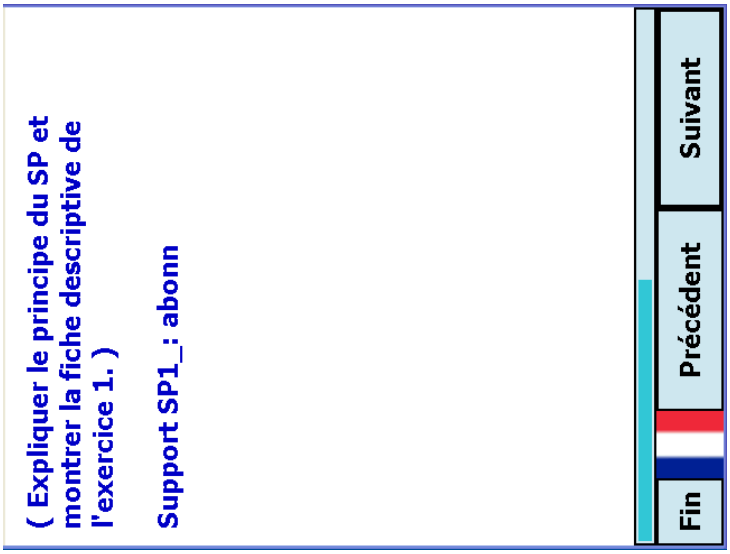

Lequel ?
Sélectionner le premier moyen de transport apparaissant dans la liste ci-dessous
Car du département
Car TER Rhône-Alpes
Navette aéroport Rhône-Express
Autre navette aéroport
Marche à pied
Vélo
Conducteur de véhicule particulier
Passager de véhicule particulier
Fourgon, camionnette, camion
Fin
Précédent
Suivant

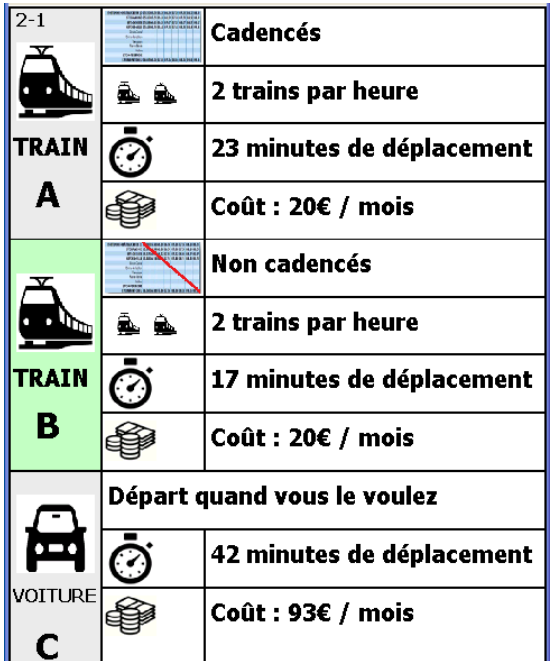

Depuis quand faites-vous ce déplacement entre Grenay (38) et Lyon 1er arrondissement (69) ?
Moins de 3 mois
Entre 3 mois et 1 an
Plus d'un an
Plus de 2 ans
Fin
Précédent
Suivant





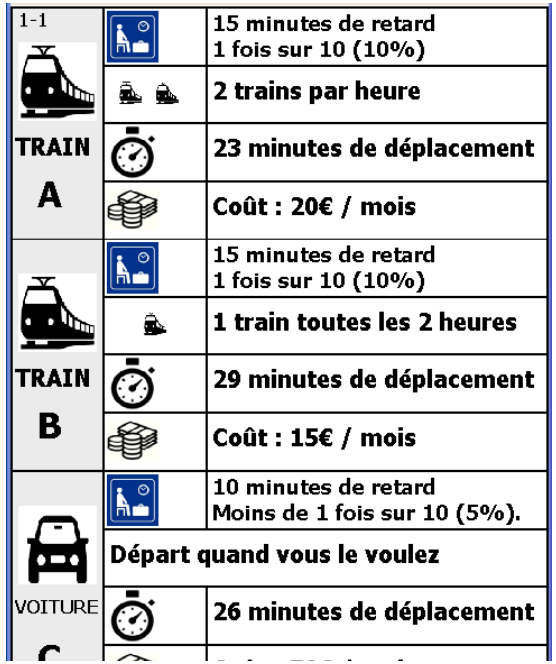
Partie « Préférence déclarée »





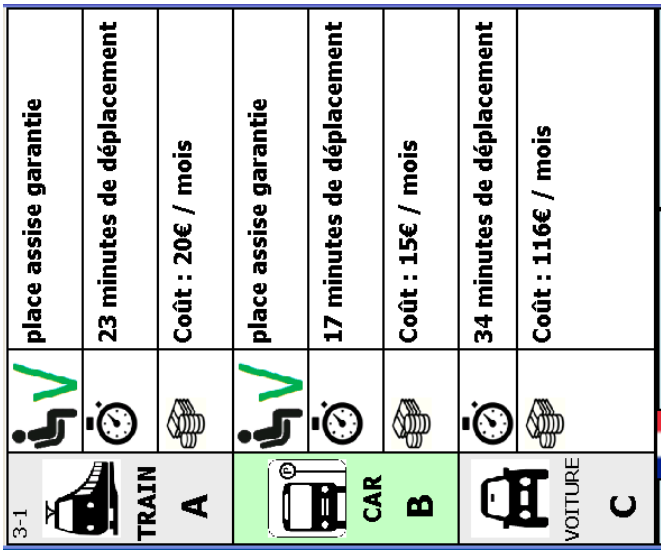
(Expliquer le principe du SP et montrer la fiche descriptive de l'exercice 3.)

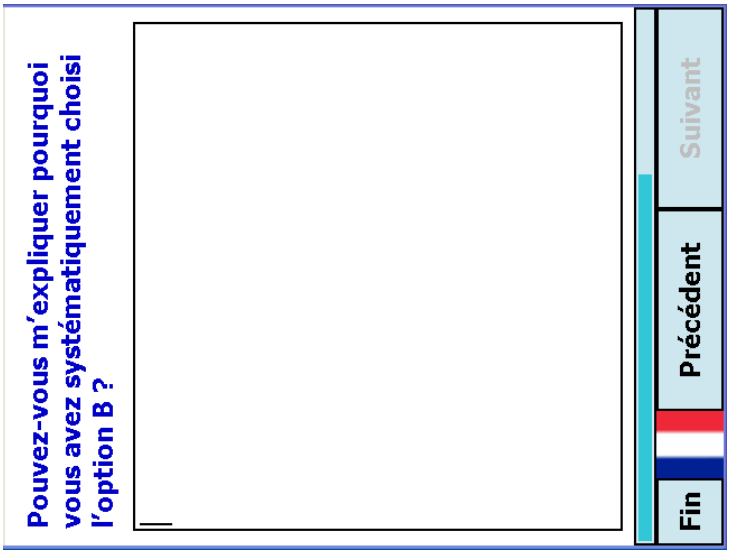
Support SP3_: abonn



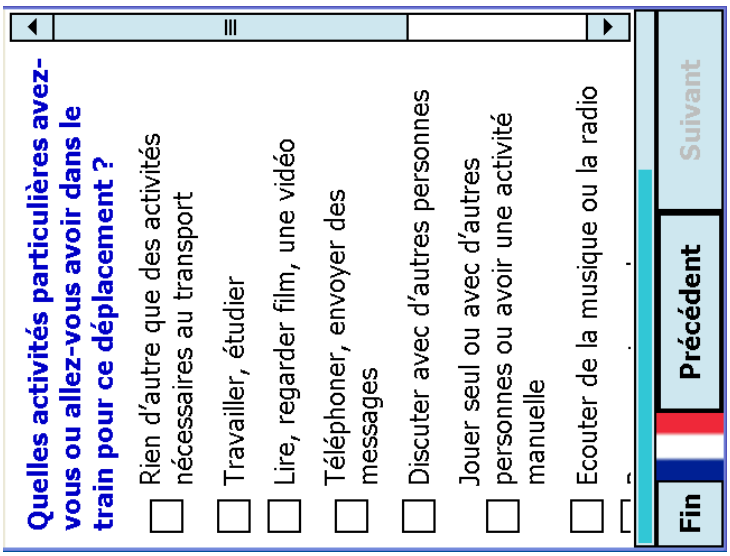
(Expliquer le principe du SP et montrer la fiche descriptive de l'exercice 2.)

Support SP2_: abonn





Partie « Socio Economique »



Quelle est ou a été votre profession ?

- ◯ Agriculteur
- ◯ Artisan, commerçant, chef d'entreprise
- ◯ Cadre et profession intellectuelle supérieure
- ◯ Profession intermédiaire
- ◯ Employé
- ◯ Ouvrier
- ◯ Elève, étudiant
- ◯ Chômeur n'ayant jamais travaillé
- ◯ Autre inactif n'ayant jamais travaillé

Voyagez-vous seul(e) ou accompagné(e) ?

- ◯ Seul(e)
- ◯ Accompagné(e)

Quelle est votre occupation principale ?

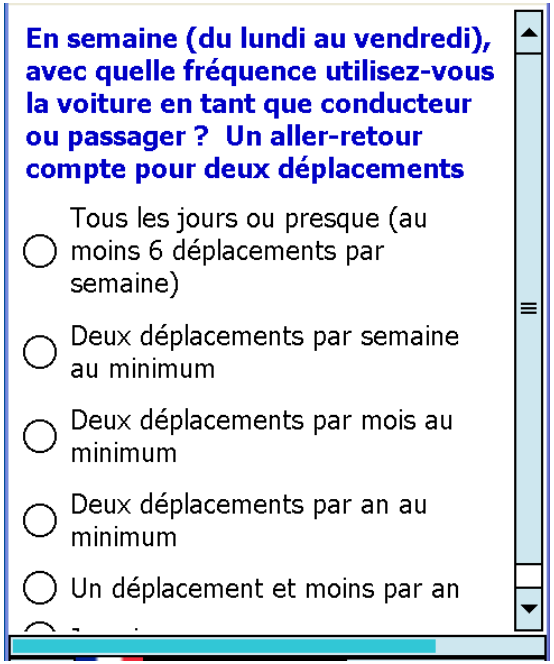
- ◯ Travail à plein temps
- ◯ Travail à temps partiel
- ◯ Formation en alternance (apprentissage, professionnalisation), stage
- ◉ Etudiant
- ◯ Scolaire jusqu'au BAC
- ◯ Chômeur et/ou en recherche d'emploi
- ◯ Retraité
- ◯ Reste au foyer
- ◯ Autre

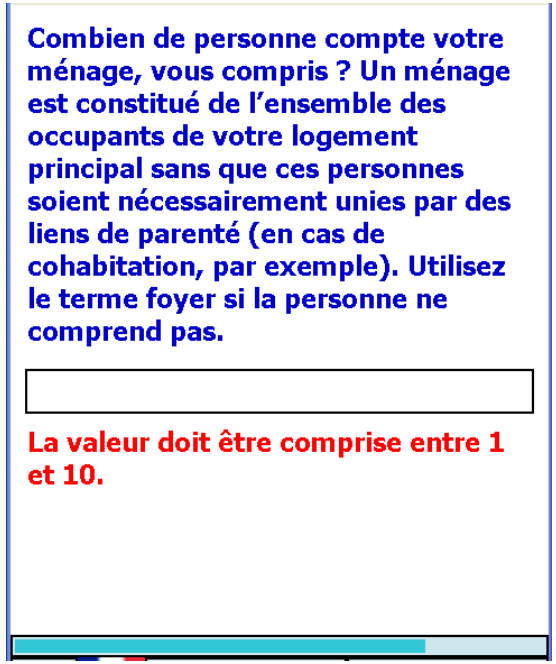
Fin | Précédent | Suivant

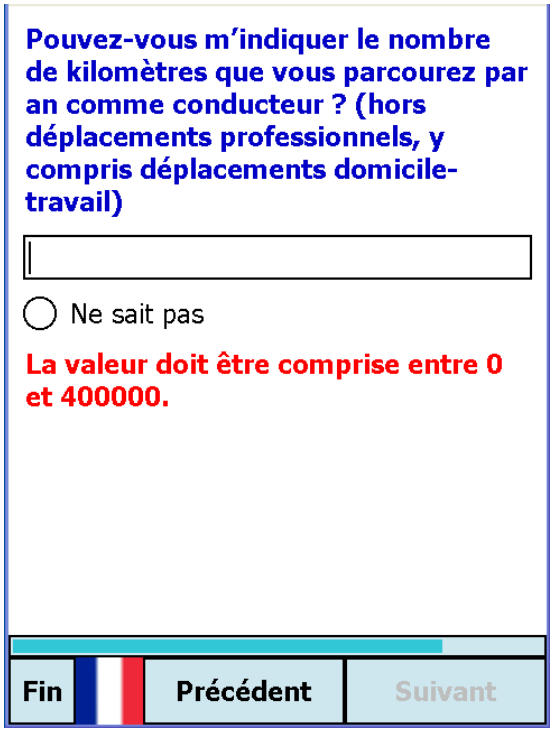
Comment qualifieriez-vous le temps de trajet ? Vous paraît-il :

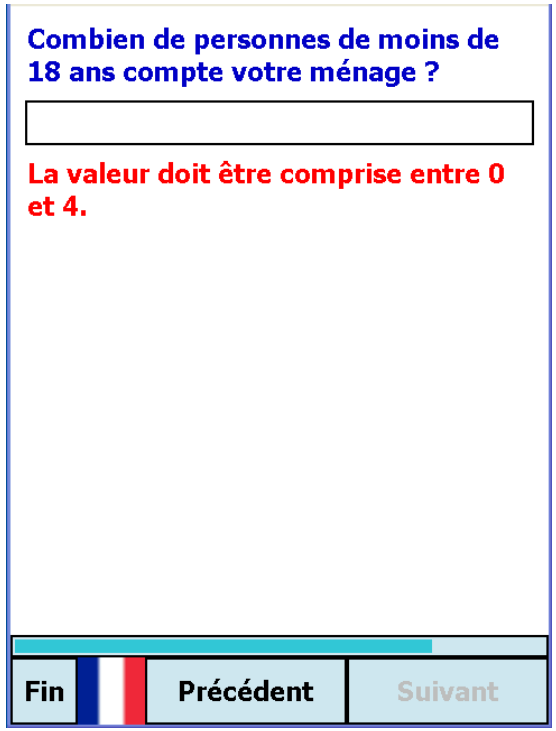
- ◯ Très long
- ◯ Assez long
- ◯ Correct
- ◯ Assez court
- ◯ Très court
- ◯ Sans opinion

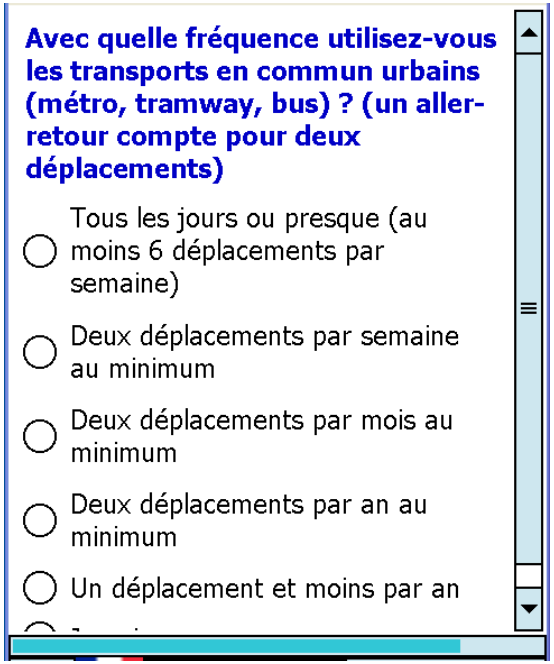
Fin | Précédent | Suivant

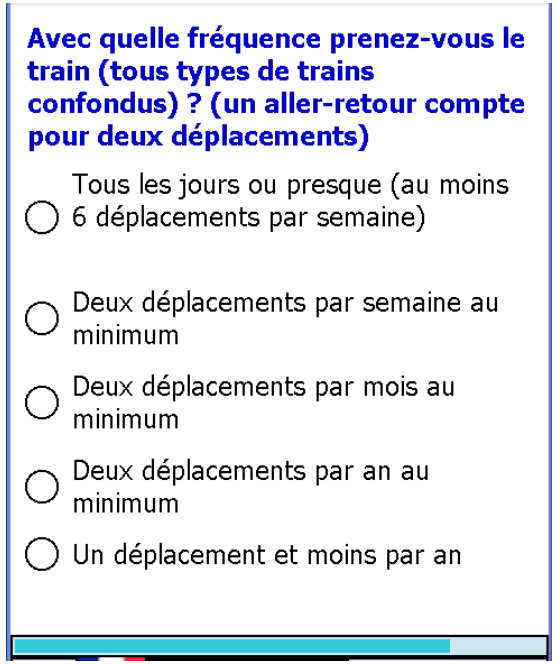

En semaine (du lundi au vendredi), avec quelle fréquence utilisez-vous la voiture en tant que conducteur ou passager ? Un aller-retour compte pour deux déplacements
Tous les jours ou presque (au moins 6 déplacements par semaine)
Deux déplacements par semaine au minimum
Deux déplacements par mois au minimum
Deux déplacements par an au minimum
Un déplacement et moins par an

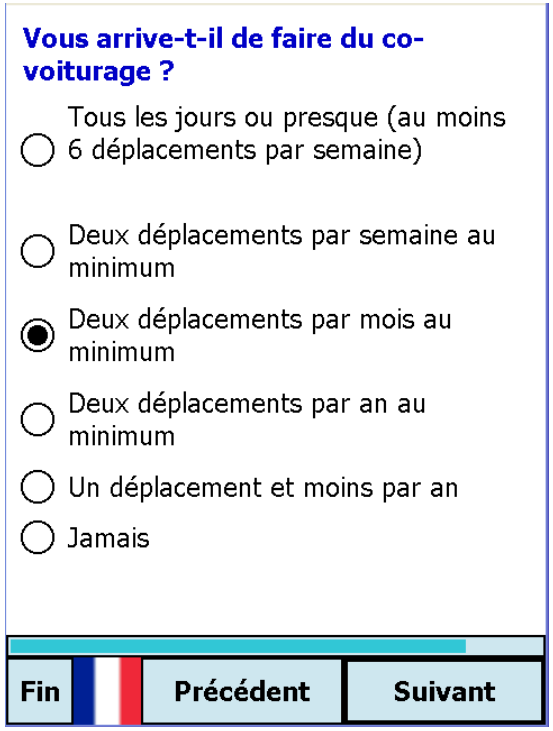

Combien de personne compte votre ménage, vous compris ? Un ménage est constitué de l'ensemble des occupants de votre logement principal sans que ces personnes soient nécessairement unies par des liens de parenté (en cas de cohabitation, par exemple). Utilisez le terme foyer si la personne ne comprend pas.
La valeur doit être comprise entre 1 et 10.

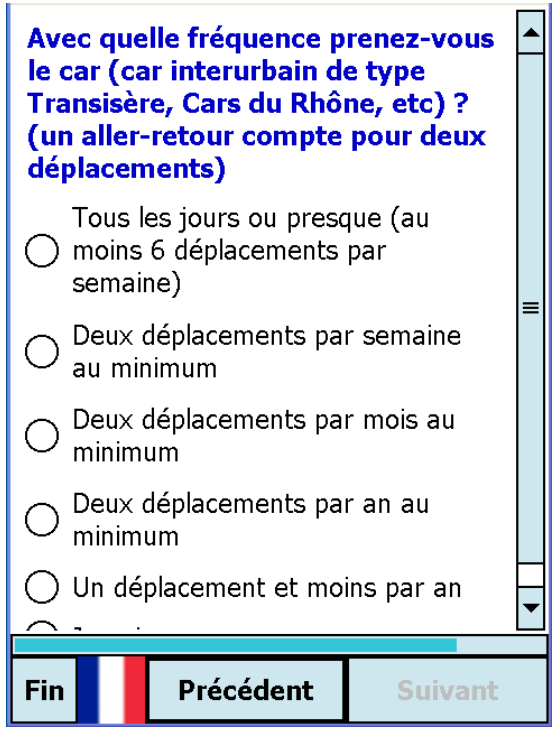

Pouvez-vous m'indiquer le nombre de kilomètres que vous parcourez par an comme conducteur ? (hors déplacements professionnels, y compris déplacements domicile-travail)
Ne sait pas
La valeur doit être comprise entre 0 et 400000.
Fin
Précédent
Suivant

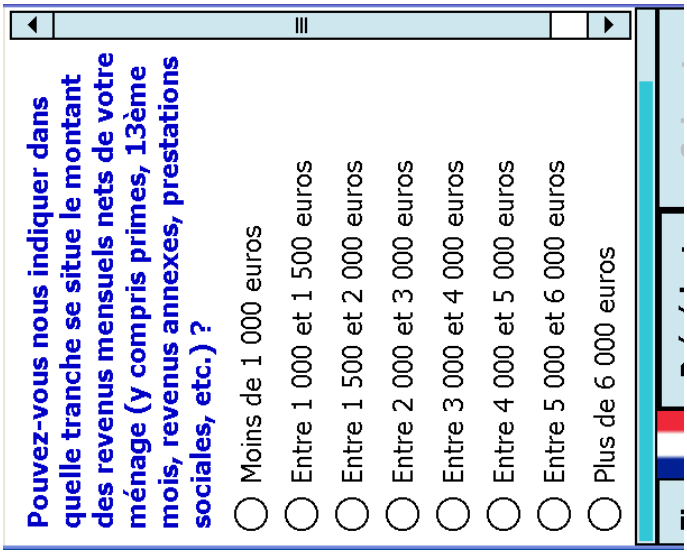

Combien de personnes de moins de 18 ans compte votre ménage ?
La valeur doit être comprise entre 0 et 4.
Fin
Précédent
Suivant


Avec quelle fréquence utilisez-vous les transports en commun urbains (métro, tramway, bus) ? (un aller-retour compte pour deux déplacements)
Tous les jours ou presque (au moins 6 déplacements par semaine)
Deux déplacements par semaine au minimum
Deux déplacements par mois au minimum
Deux déplacements par an au minimum
Un déplacement et moins par an

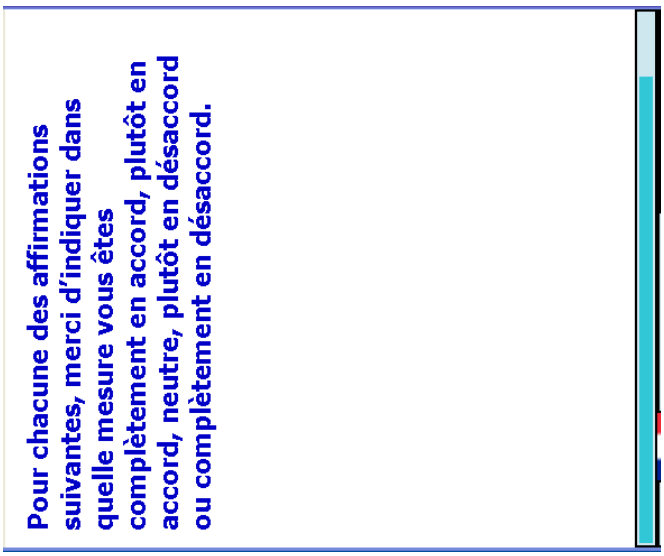

Avec quelle fréquence prenez-vous le train (tous types de trains confondus) ? (un aller-retour compte pour deux déplacements)
Tous les jours ou presque (au moins 6 déplacements par semaine)
Deux déplacements par semaine au minimum
Deux déplacements par mois au minimum
Deux déplacements par an au minimum
Un déplacement et moins par an

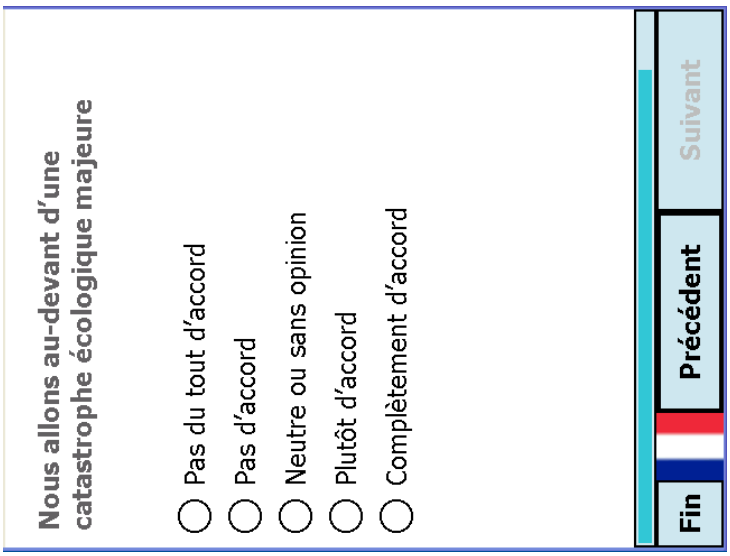

Vous arrive-t-il de faire du co-voiturage ?
Tous les jours ou presque (au moins 6 déplacements par semaine)
Deux déplacements par semaine au minimum
Deux déplacements par mois au minimum
Deux déplacements par an au minimum
Un déplacement et moins par an
Jamais
Fin
Précédent
Suivant


Avec quelle fréquence prenez-vous le car (car interurbain de type Transisère, Cars du Rhône, etc) ? (un aller-retour compte pour deux déplacements)
Tous les jours ou presque (au moins 6 déplacements par semaine)
Deux déplacements par semaine au minimum
Deux déplacements par mois au minimum
Deux déplacements par an au minimum
Un déplacement et moins par an
Fin
Précédent
Suivant



Partie « Opinion »





Pour chacune des affirmations suivantes, merci d'indiquer dans quelle mesure vous êtes d'accord ou non.

J'aime conduire vite

- ◯ Pas du tout d'accord
- ◯ Pas d'accord
- ◯ Neutre ou sans opinion
- ◯ Plutôt d'accord
- ◯ Complètement d'accord

Fin | **Précédent** | Suivant

Merci d'indiquer dans quelle mesure vous vous sentez concerné(e) par chacun des problèmes suivants causés par l'utilisation de la voiture :

L'insécurité routière

- ◯ Pas du tout concerné(e)
- ◯ Pas vraiment concerné(e)
- ◯ Neutre ou sans opinion
- ◯ Plutôt concerné(e)
- ◯ Complètement concerné(e)

Fin | **Précédent** | Suivant

Voici des phrases qu'on pourrait dire sur les trajets en train. Merci d'indiquer dans mesure vous êtes d'accord ou non

Ça me fait plaisir de voir des gens, d'être en contact avec les autres

- ◯ Pas du tout d'accord
- ◯ Pas d'accord
- ◯ Neutre ou sans opinion
- ◯ Plutôt d'accord
- ◯ Complètement d'accord

Fin | Précédent | Suivant

Merci d'indiquer comment vous vous sentez pendant vos trajets en train

Je me sens libre

- ◯ Oui, tout le temps
- ◯ Oui, souvent
- ◯ Oui, parfois
- ◯ Non, jamais
- ◯ Neutre ou sans opinion

Fin | Précédent | Suivant

Cette enquête est maintenant terminée.
Nous vous remercions de votre participation et vous souhaitons une bonne fin de voyage.
Fin
Précédent
Valider

APPENDIX D

Literature review on ICLV models: identification, estimation and structural model (Ch.4)

This appendix details the identification and estimation issues associated with ICLV models (see Chapter 4). It also details how the structural model associated with the "comfort" latent variable is explored in the literature. This case serves as an example to show the weakness of the structural models of the SEM component of the ICLV models.

D.1 Identification of ICLV models

In ICLV models, a preliminary step is to set the scale of the latent variables since the factor loadings contained in $\mathbf{D}$ (Equation 5.5, Chapter 4) can be identified only up to a scale. One solution is to set the variance of each latent variable by fixing the variance of the error term v_n to a constant, generally one. Another solution is, for each latent variable, to constrain the factor loading of an indicator - the one most strongly related to the latent variable - to a constant, generally one. The latent variable then has the same units as the indicator. If the location of the latent variable has not been set with a normalization around the mean, then a constant should be added in each measurement equation and set to zero for the indicator most strongly related to the latent variable at issue.

Using data simulating mode choice, Raveau et al. (2012) test these two methods of normalization with two variants for each method. In the first method, the variances of the latent variables are either set to their simulated value, or set to one. In the second method, the loading of one item for each latent variable is either set to its simulated value, or set to one. They found that the best results, with the latent variable parameters closest to the targets, are achieved when one loading per latent variable is

fixed to the simulated value. However, in practice, there is generally no *a priori* knowledge about the values of the loadings or variances. Since setting loadings to arbitrary values can even lead to wrong sign parameters in the discrete choice model, the second best solution is to normalize the structural equations' variances to one. Nevertheless, this does not seem to be the most common practice. Of the ICLV models on mode choice, around a third does not give any information on how the scale of the latent variables were set (e.g., Anwar et al., 2014; Córdoba and Jaramillo, 2012), another third set the scale through the measurement component of the SEM (e.g., Paulssen et al., 2014; Rieser-Schüssler and Axhausen, 2012) and the last third set the scale through the structural component of the SEM (e.g., Daziano, 2015; Scagnolarie et al., 2015).

Building on Ben-Akiva et al. (2002a), Vij and Walker (2014) discuss the three-step identification rules whose aim is to confirm that each component of the ICLV model is identified. For the two structural models, each explanatory variable is considered as perfectly measured. Furthermore, the items of the measurement model are used only for the estimation and identification of the parameters explaining the latent variables (Daziano and Rizzi, 2015; Yanez et al., 2010).

For the measurement component of the SEM, several simplified rules exist to simplify the resolution of complex equations (Vij and Walker, 2014). They depend on the number of indicators, the number of latent variables and the correlation between items or between latent variables. For instance, the three-indicator rule states that a model with at least one latent variable is identified when it has at least three indicators per latent variable, a factor complexity of one (i.e., each item load on only one latent variable) and uncorrelated measurement errors between the items. This situation corresponds, for instance, to the measurement model described in Johansson et al. (2006). For the structural component of the SEM, simplifying rules can also generally be applied (Vij and Walker, 2014) to ensure identification. For instance, a model in which endogenous variables [1] are not interrelated, that is, matrix $\mathbf{A}'$ in Equation 4.8 (Chapter 4) is null, is identified (see the models in Morikawa and Sasaki, 1998; Habib et al., 2011; Glerum et al., 2014). The recursive models, that is, models with only unidirectional effects (matrix A' is lower triangular), are also identified if the endogenous variables are uncorrelated.

Theoretical identification assumes an infinite number of observations. A lack of data or of variability in the data may lead to empirically unidentified models (Walker, 2001). A second source of non-identification is multicollinearity, which can advocate for the use of stated preferences surveys when possible. A third source is related to the violation of some hypotheses of the model, concerning

1. Generally, in ICLV models, all the latent variables are endogenous, since they are explained by observed variables.

linearity, additivity and normality in the SEM. Omission of variables or paths between variables may also result in unidentified models (Vij and Walker, 2014). Raveau et al. (2012) address the issue of variability in the data more specifically. Using simulated data, they introduce different variability scenarios. They show that the discrete choice part of the model is more robust than the SEM model. When variability is low, the parameters obtained in the structural component of the SEM are either underestimated or overestimated. Wrong sign problems may even occur, leading to counterintuitive relationships between variables. The sample size and sampling method are therefore critical to ensure empirical identification.

D.2 Estimation of ICLV models

For estimating ICLV models, the sequential approach is still sometimes preferred due to the high computational burden of the simultaneous approach. [2] Several studies compare the sequential and the simultaneous approaches with somewhat contradictory results. Morikawa and Sasaki (1998) notice that with the simultaneous approach, many parameters of the SEM model have different signs from the sequential approach. Their explanation is that, with the simultaneous approach, the parameters in the SEM model are influenced not only by the observed items but also by the choice indicator. Conversely, Raveau et al. (2010) compare both approaches using real and synthetic datasets. They find that, with a logit kernel, estimators produced by the sequential approach are not statistically different from those produced by the simultaneous approach. In the context of simulated data with large enough samples, both approaches seem capable of recovering the true parameters. However, Yanez et al. (2010) show that, with a logit kernel, a sequential approach produces inconsistent estimators for the latent variables in comparison to a mixed logit kernel. They attribute this result to the measurement error associated with introducing only the expected values of the latent variables into the choice model. This problem can be corrected by assuming that the latent variable coefficients are normally distributed over the individuals. According to Daziano and Rizzi (2015), "when using the sequential estimator, [...] estimates can be arbitrarily affected by the modeler's choice of normalization. When using the full-information estimator, alternative normalizations do not affect the estimates of the marginal utilities."

With regard to the estimation method, a large majority of studies use the simulated maximum likelihood estimation method (e.g., Walker et al., 2013; Paulssen et al., 2014; Scagnolarie et al., 2015)

2. A third approach is to integrate the fitted latent variables with their distribution in the DCM. Estimates are then consistent but inefficient (Walker, 2001). However, this method has no clear advantage over the simultaneous approach since it still involves an integral.

since, for a large number of latent variables, the computational complexity rises exponentially and numerical integration is not feasible. Glerum et al. (2014) use an exogenous sample maximum likelihood adapted to non-random samples.

D.3 Structural model in ICLV models: an example with comfort

This section explores in detail the structural models of the literature associated with the "comfort" latent variable. Need for comfort may be explained by values as well as by socio-economic factors (Table D.1). Need for comfort seems to increase with education (Johansson et al., 2006), car ownership, working full time and having no having children (Anwar et al., 2014). Hedonism, power and safety values also increase the need for comfort (Paulssen et al., 2014). Nevertheless, the three studies concerned with this variable (Johansson et al., 2006; Anwar et al., 2014; Temme et al., 2007) are not enough to enable us to draw conclusions about how the need for comfort is formed. The variables included are, indeed, either included in only one study or the results are divergent (see Table D.1). For example, age and gender are each found to be associated with a parameter which is either negative, positive or not significantly different from zero. Income has either a positive or a non-significant effect on need for comfort.

Perceptions of comfort for different modes (mainly public transport modes) are explained by socio-economic factors as well as by factors related to the trip. Age, education and being a man improve the perception of public transport modes (except for shared bus and taxi in Yanez et al. (2010). But living in a French-speaking area, working full time, income and car ownership all induce a poorer perception of public transport modes. Perception of comfort also depends on activities performed on board. Daziano and Rizzi (2015) find that performing an intellectual activity has a positive impact on the perception of train comfort but a negative one on the perceptions of bus comfort. On the the other hand, a desire to engage in social activities impacts negatively on the perception of comfort in both modes, although more so for buses than trains. They also found that travelling in a group increases the perception of comfort. According to Morikawa and Sasaki (1998), travel time has a negative impact on perception of comfort but travelling first-class by train has no significant impact on its perception.

These results suggest we should be cautious when interpreting structural models, not only because of the contradictory results, but also because of heterogeneity in the definition of latent variables, the small number of studies and the redundancy of datasets.

TABLE D.1 – Variables explaining need for comfort and perception of comfort on public transport

	Internal mental states (need for comfort)			**Perceived alternative-specific variables (perception of comfort)**			**Total**
	Negative impact	No impact	Positive impact	Negative impact	No impact	Positive impact	
Age	1	1	1	0	0	3	6
No children	0	0	1	0	0	0	1
Home ownership	1	0	0	0	0	0	1
Gender (male)	1	1	1	0	0	1	4
Income	0	1	1	2	0	0	4
Car ownership	0	0	1	2	0	0	3
Education	0	0	1	0	0	2	3
Full-time worker	0	0	1	1	0	0	2
French-speaking area	0	0	0	2	0	0	2
Intellectual activity	0	0	0	1(bus)	0	1(train)	1
Social activity	0	0	0	1	0	0	1
Travel in a group	0	0	0	0	0	1	1
Travel time	0	0	0	1	0	0	1
First-class train travel	0	0	0	0	1	0	1
Values	0	0	1	0	0	0	1

Notes: The numbers in the cells indicate the number of papers in which such a relationship was found. For instance, one paper found that age has a negative impact on the need for comfort.

APPENDIX E

Determinants of value of time (Ch.5)

Heterogeneity in VOT is determined by individual, trip and survey variables. Among individual variables, income is a key determinant. Elasticities relative to gross demand product per capita are in the range 0.47 to 0.67 (depending on the trip purpose) in an international meta-analysis (Shires and De Jong, 2009), in the range 0.7 to 0.85 in the Europan Union (Wardman et al., 2012) and about 0.9 in the U.K. (Abrantes and Wardman, 2011). For car users, VOT increases unambiguously with income (Abou-Zeid and Ben-Akiva, 2011; Hossan et al., 2016), but for public transport users cross-sectional income elasticity is typically less than one (Wardman and Lyons, 2016). An explanation is that people on higher incomes benefit more from the digital revolution. They undertake productive activities during their journey, which "reduce its disutility and hence dampen the positive effect of income on variations in VTTS [Value of Travel Time Savings] across income groups" (Wardman and Lyons, 2016, p.35). Age is also found to influence VOT, with older people having lower VOT. The effect of age on VOT is observed either through its direct effect on cost and time perception (Hossan et al., 2016) or through its effect on car-lover attitudes (Abou-Zeid and Ben-Akiva, 2011). Using different methodologies (ICLV model and a DCM using a multiple indicator solution to correct for endogeneity), both articles converge in the finding that the higher the car-loving attitude, the smaller the VOT.

VOT also depends on journey distance and journey purpose (Wardman et al., 2004; Abrantes and Wardman, 2011; Wardman et al., 2012). For public transport, different VOTs may be derived depending on the stage of the journey (walk, access, wait, schedule delay, and so on). They are usually converted into an equivalent in-vehicle time. The official VOTs used in the appraisal of transport infrastructures are mainly dependent on the transport mode (e.g., in France, Quinet et al., 2014). The VOT estimated for each mode is mostly estimated for the corresponding type of users, yielding the car's VOT for car users and public transport's VOT for public transport users. The finding of Algers et al. (1998) is that VOTs for the alternative mode are generally higher than for the mode actually used. However, two different effects have to be disentangled. Firstly, the user effect is related to the fact that users of some modes may have different socio-economic characteristics from users of other modes. Results on this

self-selection effect are divergent, with car drivers having either a higher VOT (Fosgerau et al., 2010) or a lower VOT (Gunn and Rohr, 1996) than train users. Secondly, the mode effect translates the fact that traveling by certain modes may be more productive (e.g., working on a train) or less unpleasant than traveling by other modes (e.g., traveling alone in a car). According to Gunn and Rohr (1996), train as a mode is found to have a higher value than the car. The findings of Fosgerau et al. (2010) support this result, but only for respondents with a high VOT. Indeed, for the respondents having the lowest VOT (current or potential bus users), no significant mode effects could be found.

VOT is finally highly dependent on the characteristics of the survey used to derive VOT (Shires and De Jong, 2009; Abrantes and Wardman, 2011; Wardman et al., 2012), including the numeraire (road pricing, cost fuels or all charges), the study aim, the type of survey (stated preferences *versus* revealed preferences) or the number of replications for choice questions in stated preferences surveys.

APPENDIX F

Goodness-of-fit indicators for the ICLV models (Ch.5)

This appendix sets out the details of the ICLV models (see Chapter 5). Table F.1 details the goodness-of-fit indicators 1) for the sequential estimation, which first estimates the SEM and then incorporates the simulated latent variables in the DCM, and 2) the simultaneous estimation (ICLV model).

The estimation methodology described in Chapter 5 for the ICLV models has been adopted because the "ICLV Full" Model did not converge with initial values set to zero or one (according to the parameters). However, this methodology does not allow for an insight into the goodness-of-fit indicators, since the initial loglikelihood is unknown. Therefore, the three models with only one latent variable ("ICLV Pbc","ICLV Ptime" and "ICLV Feelings") are also estimated, with all parameters set to zero, except the δ's (the thresholds in the measurement model), the σ's (the standard-deviations of the error terms in the structural model), the $\sigma^{\star}$'s (the standard-deviations of the error terms in the measurement model) and λ's (the loadings in the measurement model), which were set to one. This gives us the initial loglikelihood for the ICLV models and makes it possible to calculate the $\bar{\rho}^2$ for all models, except the one with all three latent variables. Moreover, for all models (including the two MNL models and excluding the "ICLV Full" Model), the likelihood-ratio tests show that they outperform the null models.

To allow for a comparison between models and display at least partial indicators for the "ICLV Full" Model, the results of a sequential estimation are also shown in Table F.1. Firstly, the SEM components of the models are estimated. $\bar{\rho}^2$ are comprised between 0.213 ("ICLV Ptime" Model) and 0.333 ("ICLV Full" Model). Secondly, the DCM components of the models are estimated. The initial likelihoods are always equal to -4,895, that is, $-Nln(J)$ with $N = 4,456$ the size of the sample and $J = 3$ the number of alternatives. According to the likelihood-ratio tests, the "ICLV Full" Model outperforms the three other ICLV models ("ICLV Pbc","ICLV Ptime" and "ICLV Feelings"). However, the premium, in terms

of $\bar{\rho}^2$, is low. Indeed, $\bar{\rho}^2$ are comprised between 0.253 ("ICLV Feelings" Model) and 0.255 ("ICLV Full" Model).

The three methods of estimation (sequential, simultaneous with null initial values and simultaneous with initial values calculated step by step) all offer robust results.

TABLE F.1 – Estimated Likelihood and $\bar{\rho}^2$ for the sequential and simultaneous estimation of the ICLV models

		ICLV Pbc	**ICLV Ptime**	**ICLV Feelings**	**ICLV Full**
SEM	Initial $\mathcal{L}$	-26,223	-42,924	-49,236	-118,648
	Final $\mathcal{L}$	-17,471	-33,761	-33,490	-85,623
	K	16	25	30	71
	$\bar{\rho}^2$	0.333	0.213	0.319	0.278
DCM	Initial $\mathcal{L}$	-4,895	-4,895	-4,895	-4,895
	Final $\mathcal{L}$	-3,643	-3,643	-3,647	-3,634
	K	10	10	10	14
	$\bar{\rho}^2$	0.254	0.254	0.253	0.255
ICLV	Initial $\mathcal{L}$	-31,120	-47,810	-53,038	
	Final $\mathcal{L}$	-21,061	-37,349	-37,083	-88,326
	K	26	35	40	85
	$\bar{\rho}^2$	0.322	0.218	0.300	

Notes: K is the number of parameters in the model.

APPENDIX G

Estimation results and goodness-of-fit indicators for the M3 models (Ch.6)

This appendix sets out the details of the M3 models (see Chapter 6). Table G.1 shows the estimation results and Table G.2 details the goodness-of-fit indicators.

Label	M3a	M3b	M3c	M3d	M3e	M3f	M3g	M3h	M3i
b1		0.499 (0.077) ***	0.233 (0.086) ***		0.487 (0.074) ***	0.223 (0.085) ***		0.488 (0.074) ***	0.221 (0.085) ***
b2d	0.792 (0.125) ***		0.438 (0.113) ***	0.782 (0.123) ***		0.439 (0.117) ***	0.783 (0.123) ***		0.445 (0.117) ***
b2i		0.334 (0.056) ***	0.152 (0.057) ***		0.35 (0.057) ***	0.156 (0.06) ***		0.35 (0.057) ***	0.154 (0.06) ***
b2t			0.59 (0.096) ***			0.595 (0.097) ***			0.599 (0.097) ***
b3	-0.091 (0.016) ***	-0.071 (0.015) ***	-0.082 (0.015) ***	-0.091 (0.016) ***	-0.072 (0.015) ***	-0.082 (0.015) ***	-0.091 (0.016) ***	-0.072 (0.015) ***	-0.082 (0.015) ***
b4	0.001 (0) ***	0.001 (0) ***	0.001 (0) ***	0.001 (0) ***	0.001 (0) ***	0.001 (0) ***	0.001 (0) ***	0.001 (0) ***	0.001 (0) ***
b5	0.099 (0.075) NS	0.035 (0.068) NS	0.073 (0.069) NS	0.099 (0.075) NS	0.036 (0.068) NS	0.073 (0.069) NS	0.099 (0.075) NS	0.036 (0.068) NS	0.074 (0.07) NS
b6	0.46 (0.136) ***	0.366 (0.119) ***	0.385 (0.123) ***	0.456 (0.135) ***	0.365 (0.119) ***	0.385 (0.123) ***	0.457 (0.135) ***	0.366 (0.119) ***	0.385 (0.123) ***
c1				0.557 (0.085) ***	0.646 (0.096) ***	0.646 (0.096) ***	0.111 (0.071) NS	0.092 (0.07) NS	0.091 (0.07) NS
c2d	-0.677 (0.09) ***	-0.562 (0.072) ***	-0.563 (0.073) ***				-0.599 (0.092) ***	-0.52 (0.077) ***	-0.521 (0.077) ***
c2i				-0.264 (0.043) ***	-0.265 (0.041) ***	-0.265 (0.041) ***	-0.048 (0.031) NS	-0.034 (0.027) NS	-0.034 (0.027) NS
c2t							-0.647 (0.087) ***	-0.554 (0.072) ***	-0.556 (0.072) ***
c3	0.056 (0.011) ***	0.057 (0.011) ***	0.057 (0.011) ***	0.049 (0.011) ***	0.053 (0.011) ***	0.053 (0.011) ***	0.055 (0.011) ***	0.057 (0.011) ***	0.057 (0.011) ***
c4	-0.001 (0) ***	-0.001 (0) ***	-0.001 (0) ***	0 (0) ***	-0.001 (0) ***	-0.001 (0) ***	-0.001 (0) ***	-0.001 (0) ***	-0.001 (0) ***
c5	0.057 (0.055) NS	0.057 (0.055) NS	0.057 (0.055) NS	0.021 (0.052) NS	0.01 (0.055) NS	0.01 (0.055) NS	0.046 (0.054) NS	0.048 (0.055) NS	0.048 (0.055) NS
c6	-0.251 (0.079) ***	-0.254 (0.081) ***	-0.254 (0.081) ***	-0.225 (0.071) ***	-0.232 (0.074) ***	-0.233 (0.074) ***	-0.245 (0.077) ***	-0.249 (0.079) ***	-0.249 (0.079) ***
c7	0.158 (0.025) ***	0.155 (0.025) ***	0.155 (0.025) ***	0.173 (0.024) ***	0.17 (0.024) ***	0.17 (0.024) ***	0.163 (0.025) ***	0.16 (0.025) ***	0.16 (0.025) ***
d0	-0.267 (0.032) ***	-0.252 (0.031) ***	-0.233 (0.03) ***	-0.276 (0.032) ***	-0.249 (0.031) ***	-0.246 (0.031) ***	-0.259 (0.032) ***	-0.246 (0.031) ***	-0.229 (0.03) ***
e3	0.02 (0.006) ***	0.024 (0.008) ***	0.024 (0.008) ***	0.02 (0.006) ***	0.024 (0.008) ***	0.024 (0.008) ***	0.02 (0.007) ***	0.024 (0.008) ***	0.024 (0.008) ***
e4	0 (0) ***	0 (0) ***	0 (0) ***	0 (0) ***	0 (0) ***	0 (0) ***	0 (0) ***	0 (0) ***	0 (0) ***
e5	-0.051 (0.027) *	-0.06 (0.033) *	-0.06 (0.033) *	-0.052 (0.027) *	-0.061 (0.033) *	-0.061 (0.033) *	-0.052 (0.027) *	-0.061 (0.033) *	-0.061 (0.033) *
m1				-0.474 (0.064) ***	-0.411 (0.054) ***	-0.409 (0.054) ***	-0.433 (0.061) ***	-0.377 (0.052) ***	-0.376 (0.051) ***
m3				-0.014 (0.006) **	-0.014 (0.006) **	-0.014 (0.006) **	-0.015 (0.006) **	-0.015 (0.006) **	-0.015 (0.006) **
m4				0 (0) NS	0 (0) NS	0 (0) NS	0 (0) NS	0 (0) NS	0 (0) NS
m5				0.102 (0.03) ***	0.103 (0.03) ***	0.102 (0.03) ***	0.104 (0.029) ***	0.105 (0.03) ***	0.105 (0.03) ***
p1		0.669 (0.065) ***	0.651 (0.064) ***		0.719 (0.066) ***	0.7 (0.065) ***		0.717 (0.066) ***	0.698 (0.065) ***
p3		-0.006 (0.008) NS	-0.006 (0.008) NS		-0.007 (0.008) NS	-0.007 (0.008) NS		-0.007 (0.008) NS	-0.007 (0.008) NS
p4		0 (0) NS	0 (0) NS		0 (0) *	0 (0) *		0 (0) *	0 (0) *
p5		0.078 (0.038) **	0.077 (0.038) **		0.081 (0.04) **	0.08 (0.04) **		0.081 (0.04) **	0.08 (0.04) **
Chi2	577	1771	1734	1248	2542	2512	1121	2402	2370
Df	165	593	592	271	779	778	270	778	777
Cfi	0.95	0.903	0.906	0.903	0.874	0.876	0.916	0.884	0.886
Rmsea	0.049	0.043	0.043	0.058	0.046	0.046	0.055	0.044	0.044

TABLE G.1 – Estimation results for the M3 models

	M3a	**M3b**	**M3c**	**M3d**	**M3e**	**M3f**	**M3g**	**M3h**	**M3i**
chisq.scaled	576.68	1770.75	1734.36	1247.90	2541.65	2511.96	1120.73	2401.89	2369.81
df.scaled	165.00	593.00	592.00	271.00	779.00	778.00	270.00	778.00	777.00
gfi	0.97	0.95	0.96	0.94	0.93	0.94	0.95	0.94	0.94
pgfi	0.63	0.75	0.75	0.65	0.75	0.75	0.66	0.76	0.76
agfi	0.95	0.94	0.94	0.92	0.92	0.92	0.93	0.92	0.93
cfi.scaled	0.95	0.90	0.91	0.90	0.87	0.88	0.92	0.88	0.89
tli.scaled	0.94	0.89	0.90	0.89	0.86	0.86	0.90	0.87	0.88
nfi.scaled	0.93	0.86	0.86	0.88	0.83	0.83	0.89	0.84	0.84
rmsea.scaled	0.05	0.04	0.04	0.06	0.05	0.05	0.05	0.04	0.04

TABLE G.2 – Goodness-of-fit indicators for the M3 models

APPENDIX H

Les modèles de choix discret : un cas particulier de modèles linéaires généralisés (Ch.8)

H.1 Modèles linéaires généralisés : intérêt et concepts clés

Un modèle linéaire classique explique la valeur espérée d'une variable continue Y à l'aide d'une fonction linéaire de p variables explicatives : $\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$. Pour l'observation n, on a donc :

$$E(y_n) = \beta_0 + \beta_1 x_{n,1} + \cdots + \beta_p x_{n,p}, \quad n = 1, \ldots, N. \tag{H.1}$$

Le modèle estimé prend en compte des termes d'erreur ε_n, indépendants et identiquement distribués, avec $\varepsilon_n \rightsquigarrow \mathcal{N}(0, \sigma^2)$. La structure du modèle linéaire est

$$y_n = \beta_0 + \beta_1 x_{n,1} + \cdots + \beta_p x_{n,p} + \varepsilon_n, \quad n = 1, \ldots, N. \tag{H.2}$$

Il en résulte que les y_n sont identiquement distribués selon une loi normale : $y_n \rightsquigarrow \mathcal{N}(\mu_n, \sigma^2)$. μ_n est donné par la combinaison linéaire $\beta_0 + \beta_1 x_{n,1} + \cdots + \beta_p x_{n,p}$. L'équation (H.2) peut donc se ré-écrire (H.3).

$$\mu_n = E(y_n) = \beta_0 + \beta_1 x_{n,1} + \cdots + \beta_p x_{n,p} \tag{H.3}$$

Ce cadre théorique rencontre cependant plusieurs limites, en particulier lorsque :

- la variance de Y dépend de sa moyenne ;
- les valeurs que Y peut prendre sont limitées (exemples : variables de comptage, variables catégorielles)

Les modèles linéaires généralisés (*Generalized Linear Models* ; GLM), proposés pour la première fois par Nelder et Wedderburn (1972), permettent de pallier ces deux limites. Ils étudient le lien entre une variable réponse (ou variable dépendante) et une combinaison linéaire de variables explicatives (le prédicteur). Ces modèles se caractérisent par trois composantes : la composante aléatoire à laquelle est associée une loi de probabilité, le prédicteur linéaire et la fonction de lien.

- La variable réponse Y est la composante aléatoire du modèle à laquelle est associée une loi de probabilité. Dans le cas des modèles des choix discrets avec deux options, Y est la variable dichotomique prenant la valeur 1 si l'option considérée est choisie et la valeur 0 sinon. L'échantillon statistique est composé de N variables aléatoires $(Y_1, \cdots, Y_N)$ indépendantes admettant des distributions issues d'une structure exponentielle ;
- Le prédicteur linéaire, noté η_n, est semblable à celui des modèles linéaires classiques. Il s'agit d'une combinaison linéaire des variables explicatives : $\eta_n = \beta_0 + \beta_1 x_{n,1} + \cdots + \beta_p x_{n,p}$. C'est la composante déterministe du modèle ;
- La fonction de lien g établit la relation fonctionnelle entre le prédicteur et l'espérance mathématique de la variable de réponse Y.

Par la suite, les équations sont décrites au niveau individuelle par souci de simplicité, ce qui permet d'omettre les indices n.

H.1.1 Loi de la variable réponse

La loi de probabilité associée à la variable réponse doit appartenir à la famille exponentielle qui inclut notamment la loi normale. Dans le cas d'une variable univariée, la fonction de densité s'exprime sous la forme suivante (Nelder et Wedderburn, 1972) :

$$f(y;\theta) = exp\left(\frac{y\theta - b(\theta)}{\phi}\omega + c(y,\phi)\right) \tag{H.4}$$

La fonction de densité fait intervenir deux fonctions (b et c) et trois paramètres (θ, ϕ et ω).

- L'expression des fonctions b et c dépend de la loi considérée ;
- ϕ est le paramètre de dispersion et il est supposé connu;
- ω est un poids, également supposé connu ;
- Le paramètre θ est le paramètre canonique ou naturel.

Pour chaque distribution de la famille exponentielle, θ est une fonction de la moyenne μ. La fonction

de densité peut donc s'écrire en fonction de y et μ (voir équation (H.5), Dobson et Barnett, 2008). Cette équation fait intervenir quatre fonctions (a, b, c et θ) et un paramètre μ. Si a est la fonction identité ($a(y) = y$), alors la distribution est dite canonique et $\theta(\mu)$ est appelé le paramètre naturel.

$$f(y;\mu) = \exp\left(a(y)\theta(\mu) + b(\mu) + c(y)\right) \tag{H.5}$$

H.1.2 Fonction de lien

Dans l'équation (H.3), la combinaison linéaire des variables explicatives, i.e. le prédicteur linéaire η, est égal à l'espérance de la variable réponse. Les variables η et Y doivent donc avoir le même espace de définition, en l'occurrence, $\mathbb{R}$. Par l'intermédiaire de la fonction de lien g, les GLM permettent à Y, et en particulier son espérance μ, de prendre des valeurs limitées (par exemple, des valeurs discrètes). La fonction de lien est définie par :

$$\begin{aligned} g : \mathcal{M} &\longrightarrow \mathbb{R} \\ \mu &\longmapsto \eta, \end{aligned}$$

avec $\mathcal{M}$ l'espace de définition de μ. Cet espace prend des formes différentes selon la loi de Y. En particulier, quand Y suit une loi normale, $\mathcal{M} = \mathbb{R}$ et $g = id$.

La fonction de lien qui associe μ au paramètre naturel θ est appelée la fonction de lien canonique. Le prédicteur η est alors égal au paramètre naturel :

$$\theta(\mu) = \theta = \eta \tag{H.6}$$

Tous les GLM définis avec la fonction de lien canonique possèdent des propriétés intéressantes et sont faciles à estimer car leur vraisemblance est strictement concave (McCullagh et Nelder, 1989).

H.2 GLM appliqués aux variables réponses qualitatives

Les GLM sont en particulier adaptés à la modélisation des variables réponses qualitatives non ordonnées. Deux distributions se distinguent selon le nombre d'alternatives possibles : deux ou plus de deux. Dans le cas de deux alternatives, alors la loi de probabilité est la loi de Bernoulli. Dans le cas de plus de deux alternatives, alors la loi de probabilité est la loi catégorielle.

H.2.1 Loi de Bernoulli

Une variable aléatoire Y suit une loi de Bernoulli si elle prend la valeur 1 (succès) avec la probabilité π et la valeur 0 (échec) avec la probabilité $1-\pi$, avec $\pi \in [0,1]$. La détention d'une voiture ou d'un permis de conduire ou encore le fait d'avoir ou non réalisé un déplacement dans la journée sont des variables qui suivent une loi de Bernoulli. La fonction de probabilité s'écrit

$$\Pr(Y=y) = \pi^y (1-\pi)^{1-y}, \tag{H.7}$$

avec $y \in \{0,1\}$. Et son espérance est $\mathrm{E}(Y) = \pi$.

En arrangeant l'équation (H.7), nous obtenons l'équation (H.8) (Peyhardi, 2013) qui peut s'écrire sous la forme (H.4) et sous la forme (H.5) :

$$\begin{aligned} \Pr(Y=y) &= exp\left(\log\left(\pi^y(1-\pi)^{(1-y)}\right)\right) \\ &= exp\left(y\,\log\left(\frac{\pi}{1-\pi}\right) + \log(1-\pi)\right), \end{aligned} \tag{H.8}$$

avec les définitions suivantes

(i) $\phi = \omega = 1$ (forme (H.4)),

(ii) $c(y,\phi) = 0$ (forme (H.4)) et $c(y) = 0$ (forme (H.5)),

(iii) $a(y) = y$ (forme (H.5)),

(iv) $\theta = \log\left(\frac{\pi}{1-\pi}\right)$ (forme (H.4)) et $\theta(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$ (forme (H.5)) (H.9) ,

(v) $b(\theta) = \log(1+exp(\theta)) = -\log(1-\pi)$ (forme (H.4)) et $b(\pi) = \log(1-\pi)$ (forme (H.5)).

H.2.2 Loi catégorielle

Nous avons vu jusqu'à présent l'écriture des GLM avec une variable réponse univariée. Pour le cas multivarié, il est nécessaire d'adapter l'écriture des équations (H.4) et (H.5). Considérons une variable réponse $Y \in \mathbb{R}^{J-1}$ dont la distribution dépend du paramètre $\theta \in \mathbb{R}^{J-1}$.

La forme (H.4) se généralise de la façon suivante pour les distributions multivariés

$$f(y;\theta) = exp\left(\frac{y^T\theta - b(\theta)}{\phi}\omega + c(y,\phi)\right), \tag{H.10}$$

avec a, b et c des fonctions connues, ϕ le paramètre de dispersion et θ le paramètre naturel.

La forme (H.5) se généralise de la façon suivante pour les distributions multivariées

$$f(y;\theta) = exp\left(a(y)^T\theta(\mu) + b(\mu) + c(y)\right), \tag{H.11}$$

avec a, b, c et θ des fonctions connues.

Soit Y qui suit une loi catégorielle et $\pi_1, \ldots, \pi_J$ les probabilités associées à chacune des J alternatives, avec $\sum_{j=1}^{J} \pi_j = 1$.

Les paramètre à estimer sont les probabilités π_j. La probabilité que Y prenne la valeur j $(j = 1, \ldots, J)$ s'écrit

$$\Pr(Y = j) = \pi_j = \prod_{j=1}^{J} \pi_j^{y_j}, \tag{H.12}$$

avec $y_j = 1$ si $Y = j$ et 0 sinon. Donc une seule des composantes du vecteur $y^T = (y_1, \ldots, y_J)$ vaut 1 et $\sum_{j=1}^{J} y_j = 1$. L'équation (H.12) peut donc se ré-écrire :

$$\begin{aligned} \pi_j &= f(y;\pi) \\ &= \prod_{j=1}^{J-1} \pi_j^{y_j} \pi_J^{y_J} \\ &= \prod_{j=1}^{J-1} \pi_j^{y_j} \left(1 - \sum_{j=1}^{J-1} \pi_j\right)^{1-\sum_{j=1}^{J-1} y_j}. \end{aligned} \tag{H.13}$$

Comme pour la loi de Bernoulli, il est possible d'arranger l'équation (H.13) pour la ré-écrire sous les formes (H.10) et (H.11) (Peyhardi, 2013)

$$\begin{aligned} f(y;\pi) &= exp\left(\log\left[\prod_{j=1}^{J-1} \pi_j^{y_j} \left(1 - \sum_{j=1}^{J-1} \pi_j\right)^{1-\sum_{j=1}^{J-1} y_j}\right]\right) \\ &= exp\left(\sum_{j=1}^{J-1} y_j \log\left[\frac{\pi_j}{1 - \sum_{j=1}^{J-1} \pi_j}\right] + \log\left[1 - \sum_{j=1}^{J-1} \pi_j\right]\right) \end{aligned} \tag{H.14}$$

avec les définitions suivantes :

(i) $\phi = \omega = 1$ (forme (H.10)),

(ii) $c(y, \phi) = 0$ (forme (H.10)) et $c(y) = 0$ (forme (H.11)),

(iii) $a(y) = y$ (forme (H.11)),

(iv) $\theta = \theta(\pi) = \left(\log\left(\frac{\pi_1}{1 - \sum_{j=1}^{J-1} \pi_j} \right), \ldots, \log\left(\frac{\pi_{J-1}}{1 - \sum_{j=1}^{J-1} \pi_j} \right) \right)^T$ (formes (H.10) et (H.11)) (H.15),

(v) $b(\theta) = \log\left(1 + \sum_{j=1}^{J-1} exp(\theta_j)\right) = \log\left(1 + \sum_{j=1}^{J-1} \frac{\pi_j}{1 - \sum_{j=1}^{J-1} \pi_j}\right) = -\log\left(1 - \sum_{j=1}^{J-1} \pi_j\right)$ (forme (H.10))

et $b(\pi) = \log\left(1 - \sum_{j=1}^{J-1} \pi_j\right)$ (forme (H.11)).

H.2.2.1 Fonction de lien canonique

La fonction de lien canonique pour les GLM ayant une variable réponse catégorielle s'écrit

$$\begin{aligned} g : \mathcal{M} &\longrightarrow \mathbb{R}^{J-1} \\ \pi &\longmapsto \eta, \end{aligned} \tag{H.16}$$

avec

$$g_j(\pi) = \log\left(\frac{\pi_j}{1 - \sum_{i=1}^{J-1} \pi_i} \right), \forall j = 1, \ldots, J-1, \tag{H.17}$$

et $\mathcal{M} = \{\pi \in]0,1[^J \sum_{j=1}^{J} \pi_j = 1\}$.

L'équation (H.17) permet bien de retrouver l'équation (H.15) pour $J > 2$ et le cas particulier de l'équation (H.9) pour $J = 2$.

H.2.2.2 Lien avec le modèle logit multinomial

Notons η_j $(j = 1, \cdots, J)$ chacune des composantes du prédicteur $\eta = (\eta_1, \cdots, \eta_J)^T$. Soit Z la matrice de design, c'est-à-dire la matrice qui contient la structure des variables explicatives. Par exemple, si le choix de l'alternative j est conditionné par une constante, un vecteur de k variables individuelles x et un vecteur de l attributs ω_j qui se définit comme la différence entre les attributs de l'alternative j et ceux de l'alternative de référence J, alors la matrice de design Z, de dimension $(J-1) \times p(J-1)$, avec $p = 1 + l + k$, s'écrit

$$Z = \begin{pmatrix} 1 & & x^T & & \omega_1^T & & \\ & \ddots & & \ddots & & \ddots & \\ & & 1 & & x^T & & \omega_{J-1}^T \end{pmatrix}.$$

Par l'adjonction de contraintes, d'autres matrices de design peuvent être construites.

En notant β un vecteur de $p(J-1)$ paramètres, on a alors $\eta = Z\beta$. Et, avec z_j^T le vecteur défini par la j$^{\text{ème}}$ ligne de la matrice Z, on peut écrire $\eta_j = z_j^T\beta$. Notons que dans le cas où $J = 2$, alors le design est compris dans le vecteur z de dimension p. Avec ces définitions, et à l'aide de la fonction de lien canonique (équation (H.17)), il est possible d'arranger l'équation (H.12) pour définir la probabilité de choix d'une option (équation (H.18)) telle qu'on la trouve dans la littérature dédiée aux modèles de choix discret (Train, 2009). Remarquons d'abord que, par définition de la fonction de lien canonique :

$$g_j(\pi) = \eta_j, \forall j \neq J, \text{ et donc } \frac{\pi_j}{1 - \sum_{i=1}^{J-1} \pi_i} = \exp(\eta_j), \forall j \neq J.$$

La fonction de probabilité peut donc s'écrire, $\forall j \neq J$:

$$\begin{aligned} \pi_j &= \frac{\pi_j}{\sum_{i=1}^{J} \pi_i} \\ &= \frac{\pi_j/(1 - \sum_{k=1}^{J-1} \pi_k)}{\sum_{i=1}^{J} \pi_i/(1 - \sum_{k=1}^{J-1} \pi_k)} \\ &= \frac{\exp(\eta_j)}{\sum_{i=1}^{J} \exp(\eta_i)} \\ &= \frac{\exp(z_j^T\beta)}{\sum_{i=1}^{J} \exp(z_i^T\beta)}. \end{aligned} \tag{H.18}$$

H.3 Synthèse

Le tableau H.1 fait un parallèle entre l'approche statistique et l'approche économique des modèles de choix discret. Le tableau H.2 fait le lien entre distribution, exemples d'application, fonction de probabilité, fonction de lien canonique et moyenne.

TABLEAU H.1 – Parralèle entre approche statistique et économie des modèles de choix discret

Notation	**Signification statistique**	**Signification économique**
Y	Variable réponse. Suit une distribution faisant partie de la famille exponentielle. Loi binomiale pour la régression logistique et loi catégorielle pour la régression logistique multinomiale.	Variable dépendante, variable que l'on cherche à expliquer. Dans les modèles de choix discret, elle prend des valeurs discrètes. $Y \in \mathbb{N}^*$. La distribution de Y est généralement non mentionnée. .
β	Vecteur de coefficients/paramètres	
$Z\beta$ou η	Prédicteur. Sa structure est donnée par la matrice de design Z.	Combinaison linéaire des variables explicatives.
g	Fonction de lien. Intuition : c'est la fonction qui permet à la variable réponse d'avoir un espace de définition différent de $\mathbb{R}$. Pour chaque distribution, plusieurs fonctions de liens sont possibles. L'une d'entre elles, la fonction de lien canonique permet de faciliter les estimations.	Généralement non mentionnée par les économistes. Dans la plupart des modèles de choix discret, la fonction de lien canonique est telle que $\theta(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$.

TABLEAU H.2 – Exemples de fonctions de distribution

Distribution	**Exemple type**	**Exemple transport**	**Support de distribution**	**Fonction de probabilité**	**Fonction de lien canonique**	**Fonction moyenne**
Bernoulli	Un lancer d'une pièce. Probabilité d'avoir pile.	Probabilité qu'une personne ait fait au moins un déplacement dans la journée.	Entier : $Y \in \{0, 1\}$	$\Pr(Y = y) = \pi_1^y (1 - \pi)^{1-y}$ $E(y) = \pi$	Fonction logit : $z^T \beta = \log\left(\frac{\pi}{1-\pi}\right)$	$\pi = \frac{1}{1+\exp(-z^T \beta)}$
Catégorielle	Un lancer d'un dé à J faces. Probabilité d'avoir la face j.	Probabilité qu'une personne utilise le mode de déplacement j parmi les J modes possibles.	Entier : $Y \in \{1, \ldots, J\}$ ou Vecteur Y de dimension J : $Y = (Y_1, \ldots, Y_j)$, avec $Y_j \in \{0, 1\}$	$\Pr(Y = j) = \pi_j = \prod_{i=1}^{J} \pi_i^{y_i}$ $E(Y) = \pi_i$	Fonction logit multivariée: $z_j^T \beta = \log\left(\frac{\pi_j}{1 - \sum_{i=1}^{J-1} \pi_i}\right)$	$\pi_j = \frac{\exp(z_j^T \beta)}{\sum_{i=1}^{J} \exp(z_i^T \beta)}$

APPENDIX I

Value of time, time equivalent and marginal effects in reference models (Ch.8)

I.1 Specific willingness to pay

For a RUM model, the Value Of Time (VOT) is the change in cost that keeps the utilities (deterministic part) unchanged after increasing the time by one unit. It can be shown that no change in utilities implies no change in choice probabilities. The VOT can therefore be defined as the change in cost that keeps the choice probabilities unchanged after increasing the time by one unit. This is the value Δc_{j_0} such that $\pi'_j = \pi_j$ for all alternatives $j = 1, \dots, J$ after the transformation of the time $t'_{j_0} = t_{j_0} + 1$ for the specific alternative j_0 (the cost is transformed as $c'_{j_0} = c_{j_0} + \Delta c_{j_0}$). As the link function of a reference model is a surjective map from the simplex to $\mathbb{R}^{J-1}$, for any cdf $F \in \mathfrak{F}$, the equalities between the J probabilities $\pi'_j = \pi_j$ are equivalent to the equalities between the $J-1$ predictors $\eta'_j = \eta_j$. If $j_0 \neq J$, only the predictor η'_{j_0} is impacted by the change $t'_{j_0} = t_{j_0} + 1$ and we have (the linear predictor contains the interaction between time and the binary variable comfort)

$$
\begin{aligned}
\eta'_{j_0} &= \eta_{j_0} \\
\gamma^{\text{time}} t'_{j_0} + \gamma^{\text{time}\times\text{comf}} t'_{j_0} \mathbf{1}_{\text{comf}=1} + \gamma^{\text{cost}} c'_{j_0} &= \gamma^{\text{time}} t_{j_0} + \gamma^{\text{time}\times\text{comf}} t_{j_0} \mathbf{1}_{\text{comf}=1} + \gamma^{\text{cost}} c_{j_0} \\
\gamma^{\text{time}} + \gamma^{\text{time}\times\text{comf}} \mathbf{1}_{\text{comf}=1} + \gamma^{\text{cost}} \Delta c_{j_0} &= 0
\end{aligned}
$$

If $j_0 = J$ all the $J-1$ predictors are identically impacted by the change t'_{j_0} and we obtain the same result. Finally, the VOT is given by

$$\text{VOT} = -\frac{\gamma^{\text{time}} + \gamma^{\text{time}\times\text{comf}}\mathbf{1}_{\text{comf}=1}}{\gamma^{\text{cost}}}. \tag{I.1}$$

TEC is the amount of time Δt_{j_0} necessary to compensate a transition from guarantee of a seat (comfort $= 1$) to no guarantee of a seat. After a similar computation it can be shown that

$$\text{TEC} = -\frac{\gamma^{\text{comf}} + \gamma^{\text{time}\times\text{comf}} t_{j_0}}{\gamma^{\text{time}}}. \tag{I.2}$$

The VOT and TEC formulas obtained for reference models are those obtained with the canonical multinomial logit model.

I.2 Marginal effects

I.2.1 Marginal effects computation

For the sake of simplicity, we denote by ω_j the value of only one alternative-specific variable, such as the time for instance. The marginal effect of a change in ω_j on the probability of choosing the alternative i is given by

$$\begin{aligned}
\text{me}_i(\omega_j) &= \frac{\partial \pi_i}{\partial \omega_j}, \\
&= \frac{\partial \eta}{\partial \omega_j}\frac{\partial \pi_i}{\partial \eta}, \\
\text{me}_i(\omega_j) &= \sum_{k \neq J} \frac{\partial \eta_k}{\partial \omega_j}\frac{\partial \pi_i}{\partial \eta_k},
\end{aligned} \tag{I.3}$$

for all $i \in \{1, \ldots, J\}$. On the one hand, we have $\eta_k = \alpha_k + x^T\delta_k + (\omega_k - \omega_J)\gamma$ for $k \neq J$ and thus

$$\frac{\partial \eta_k}{\partial \omega_j} = \gamma\left\{\mathbf{1}_{(j=k)} - \mathbf{1}_{(j=J)}\right\},$$

for $k \neq J$ and all $j = 1, \ldots, J$. On the other hand, recall that for $i \neq J$ and $j \neq J$

$$\begin{aligned}\frac{\partial \pi_i}{\partial \eta_j} &= d_j \text{Cov}(Y_i, Y_j),\\ \frac{\partial \pi_i}{\partial \eta_j} &= d_j \pi_i \left\{\mathbf{1}_{(i=j)} - \pi_j\right\},\end{aligned} \tag{I.4}$$

where $d_j = f(\eta_j)/[F(\eta_j)\{1 - F(\eta_j)\}]$ and Y_j is the indicator variables (equal to 1 if j is the chosen alternative and 0 otherwise) for $j \neq J$; see the supplementary materials of Peyhardi et al. (2015) for more details. Because of the constraint $\pi_J = 1 - \sum_{i \neq J} \pi_i$, it follows that for $j \neq J$

$$\begin{aligned}\frac{\partial \pi_J}{\partial \eta_j} &= -\sum_{i \neq J} \frac{\partial \pi_i}{\partial \eta_j},\\ &= -d_j \left\{\pi_j - \pi_j \sum_{i \neq J} \pi_i\right\},\\ &= -d_j \left\{\pi_j - \pi_j(1 - \pi_J)\right\},\\ \frac{\partial \pi_J}{\partial \eta_j} &= -d_j \pi_j \pi_J,\end{aligned}$$

which means that Equation (I.4) holds for $j \neq J$ and all $i = 1, \ldots, J$. Finally, the equality (I.3) becomes

$$\begin{aligned}\text{me}_i(\omega_j) &= \gamma \left\{\mathbf{1}_{(j \neq J)} \frac{\partial \pi_i}{\partial \eta_j} - \mathbf{1}_{(j=J)} \sum_{k \neq J} \frac{\partial \pi_i}{\partial \eta_k}\right\}\\ \text{me}_i(\omega_j) &= \gamma \pi_i \left[\mathbf{1}_{(j \neq J)} d_j \left\{\mathbf{1}_{(i=j)} - \pi_j\right\} - \mathbf{1}_{(j=J)} \left\{\mathbf{1}_{(i \neq j)} d_i - \sum_{k \neq J} d_k \pi_k\right\}\right]\end{aligned}$$

for all $i = 1, \ldots, J$ and all $j = 1, \ldots, J$ and thus the desired result is proved.

I.2.2 Invariance of marginal effects under normalization

Let us show that marginal effects are not impacted by a normalization of the cdf. First, the probabilities π_j stay unchanged since $F_{m,s}(\eta'_j) = F(\eta_j)$ for $j \neq J$ by definition. Now, let us note that

$$\begin{aligned}f_{m,s}(z) &= \frac{\partial}{\partial z} F_{m,s}(z)\\ &= \frac{\partial}{\partial z} F(\frac{z-m}{s})\\ f_{m,s}(z) &= \frac{1}{s} f(\frac{z-m}{s})\end{aligned}$$

and thus $f_{m,s}(\eta'_j) = f(\eta_j)/s$, which implies that $d'_j = d_j/s$ for $j \neq J$. After normalization the marginal effects become:

$$\begin{aligned}
\text{me}'_i(\omega_j) &= \gamma' \pi_i \left[\mathbf{1}_{(j \neq J)} d'_j \left\{ \mathbf{1}_{(i=j)} - \pi_j \right\} - \mathbf{1}_{(j=J)} \left\{ \mathbf{1}_{(i \neq j)} d'_i - \sum_{k \neq J} d'_k \pi_k \right\} \right], \\
&= \frac{\gamma'}{s} \pi_i \left[\mathbf{1}_{(j \neq J)} d_j \left\{ \mathbf{1}_{(i=j)} - \pi_j \right\} - \mathbf{1}_{(j=J)} \left\{ \mathbf{1}_{(i \neq j)} d_i - \sum_{k \neq J} d_k \pi_k \right\} \right], \\
\text{me}'_i(\omega_j) &= \text{me}_i(\omega_j),
\end{aligned}$$

since $\gamma' = s\gamma$.

APPENDIX J

Synthèse des indicateurs économiques (Conclusion)

Cette annexe synthétise et discute l'ensemble des consentements à payer et équivalents temps calculés dans les différents chapitres. La compilation des valeurs du temps générées dans la thèse donne notamment l'opportunité de les comparer plus précisément avec les valeurs officielles du rapport Quinet et al. (2014).

J.1 Valeur du temps

Les valeurs du temps calculées à partir de l'enquête de préférences déclarées sont comprises entre 4,6 €/heure et 30,4 €/heure, selon l'exercice de choix, le type de modèle de choix discret, les variables explicatives, le type d'usager et le niveau de confort (voir tableau J.1). En excluant les valeurs aberrantes estimées avec le modèle de référence ne prenant en compte que les attributs, alors la valeur du temps minimale est de 10,2 €/heure. Ces valeurs sont toutes dans l'ordre de grandeur de celles du rapport Quinet et al. (2014) qui sont les valeurs de référence pour l'évaluation des projets de transport en France. Dans le rapport Quinet, les valeurs du temps tous modes pour des déplacements urbains sont comprises entre 6,8 et 22,3 €/heure selon la région (Ile de France ou France entière) et le motif. Pour tous types de motifs à l'échelle de l'ensemble de la France, la valeur est de 7,9 €/heure. Pour des déplacements interurbains, les valeurs du temps (hors avion) sont comprises entre 6,8 et 43,3 €/heure, selon le mode, le motif et la distance parcourue. Même la valeur la plus élevée estimée pour les déplacements en voiture à partir de l'enquête de préférences déclarées (30,4 €/heure) ne semble donc pas être aberrante.

En revanche, le classement des valeurs du temps dans le rapport Quinet n'est pas le même que celui réalisé avec l'enquête de préférences déclarées. Tous motifs confondus, les valeurs du temps sont les mêmes quel que soit le mode de transport pour les déplacements de 20 km. Pour les déplacements de plus de 20 km, la valeur du temps en voiture est la plus faible, celle en train est intermédiaire et celle

en car est la plus élevée. L'enquête de préférences déclarées confirme que le car, et le train sans garantie d'une place assise, génèrent la valeur du temps la plus élevée (environ 14 €/heure). Mais la voiture est associée à une valeur du temps intermédiaire (environ 13 €/heure) et non à la valeur la plus faible qui est générée par train (environ 10 €/heure), du fait de l'intégration de la dimension confort avec la garantie d'une place assise.

Le rapport Quinet prend en compte le confort en estimant que les valeurs du temps présentées doivent être augmentées de 25% (plus un paramètre fonction de la charge du véhicule). A l'aide d'une méta-analyse, Wardman et Whelan (2011) trouvent une augmentation de l'ordre de 50%. Ce calcul s'applique certainement davantage aux transports en commun urbains qu'au train puisqu'elles prennent en compte la charge du véhicule. D'après les estimations réalisées à partir de l'exercice 3 dans les chapitres 3 et 5, la valeur du temps en train augmente d'environ 40 % lors du passage d'une place assise garantie à une place assise non garantie.

Par ailleurs, dans le rapport Quinet, la valeur du temps du train, du car et de la voiture augmentent avec la distance de déplacement jusqu'à 400 km. Au contraire, les valeurs du temps estimées à l'aide du modèle prenant en compte l'aversion au risque (chapitre 7) génère des valeurs du temps qui diminuent avec le temps de déplacement.

J.2 Confort

Concernant le confort, deux méthodes de calcul différentes sont utilisées dans les chapitre 3 et 8 (voir tableau J.2). Dans le chapitre 3, le confort n'est intégré que dans l'utilité du train (et est confondu avec la constante dans les utilités du car et de la voiture). Le consentement à payer pour le confort est compris entre 2,1 et 6,3 € selon le temps de déplacement et selon qu'on considère l'exercice 3 seul ou les trois exercices de choix combinés. Le confort peut également être traduit en équivalent-temps à partir de la valeur du temps. Les valeurs générées sont comprises entre 9,0 et 28,4 minutes pour avoir la garantie d'une place assise, selon le temps de déplacement et le modèle. Dans le chapitre 8, le confort n'est pas différencié entre les alternatives, ce qui conduit à l'introduire dans l'utilité de l'ensemble des modes. Comme cela est discuté dans le chapitre correspondant, les valeurs calculées dans le modèle ne comprenant que les attributs sont globalement aberrantes. Celles calculées avec le modèle complet sont plus réalistes et comprises entre 14,5 et 26,3 minutes.

J.3 Temps de retard, fréquence de passage des trains et des cars et cadencement [1]

Le tableau J.3 montre que le consentement à payer pour éviter une heure de retard ayant une probabilité d'occurrence de 0,05 ou 0,1 (respectivement 0,2) est d'environ 20 €/heure (respectivement environ 36 €/heure). Les voyageurs sont prêts à rallonger leur temps de déplacement d'environ 1,4 à 2,5 minutes (respectivement 2,6 à 2,9) minutes pour éviter une minute de retard ayant une probabilité d'occurrence de 0,05 ou 0,1 (respectivement 0,2).

Le consentement à payer pour que la fréquence de passage des trains augmente de un train par heure est de 2,8 à 3,6 € selon l'exercice de choix considéré. Les usagers sont prêts à passer de 13 à 17 minutes supplémentaires dans le train ou le car pour bénéficier de cette hausse de fréquence.

Le consentement à payer pour bénéficier d'horaires cadencés en train ou en car est de 4,3 € et les usagers sont prêts à passer environ 21 minutes de plus à voyager pour bénéficier du cadencement.

1. Afin d'être exhaustif, cette section rappelle brièvement les valeurs déjà présentées et discutées dans le chapitre 3. Voir le chapitre 3 pour une comparaison avec la littérature.

TABLEAU J.1 – Valeurs du temps (en €/heure) estimées à l'échelle de la Région Rhône-Alpes

Chapitre	**Modèle**	**Exercice**	**Car/Train (confort=0)**	**Train (confort=1)**	**Voiture**
Chapitre 3	Logit multinomial	1, 2, 3	[12,0;14,3] selon l'exercice	[9,9;10,2]	[13,0;14,4] selon l'exercice
Chapitre 5	Logit multinomial	3	15.4	14,2 10.8	13,9 14,4
	Modèle de choix hybride		[13,5;15,4] selon les valeurs prises par la variable latente	[9,4;11,3] selon les valeurs prises par la variable latente	[13,1;13,7] selon la variable latente du modèle
Chapitre 7	Nested logit	3	[12,5;12,8] selon la prise en compte des probabilités de retard		[13,4;15,4]
	Latent class model		[12,1;14,7] selon le type d'usager		[14,5;30,4] selon le type d'usager
Chapitre 8	Modèles de référence uniquement attributs	3	[4,6;7,4] selon modèle	[5,9;7,1] selon modèle	voir (confort=0)
	Reference models attributs+variables individuelles		[12,5;14,6] selon modèle	[11,5;12,7] selon modèle	voir (confort=0)

Notes : (confort=0) : sans garantie de place assise ; (confort=1) : garantie de place assise. Les fourchettes de valeur indiquent les valeurs minimales et maximales estimées.

TABLEAU J.2 – Consentements à payer et équivalents temps du confort estimés à l'échelle de la Région Rhône-Alpes

Chapitre	**Modèle**	**Indicateur**	**Temps de déplacement**		
			30 min	**60 min**	**90 min**
Chapitre 3	Logit multinomial	Consentement à payer	[2,1;3,3]	[4,2;4,6]	[6,0;6,3]
		Equivalent temps	[9,0;15,6]	[17,6;22,0]	[26,2;28,4]
Chapitre 8	Reference models uniquement attributs	Equivalent temps	[19,0;75,2]	[20,3;64,9]	[21,6;54,6]
	Reference models attributs+variables individuelles		[14,5;20,1]	[17,0;23,2]	[19,6;26,3]

Notes : min = minutes. Consentement à payer en €. Équivalent temps en minutes. Les fourchettes de valeur indiquent les valeurs minimales et maximales estimées selon le modèle.

TABLEAU J.3 – Consentements à payer et équivalents temps du temps de retard, de la fréquence et du cadencement estimés à l'échelle de la Région Rhône-Alpes

Chapitre	**Modèle**	**Indicateur**	**Temps de retard** (/heure)		**Fréquence** (/train/heure)	**Cadencement**
			Prob.=0,05/0,1	Prob.=0,2		
Chapitre 3	Multinomial logit	Consentement à payer	[19,3;20,1]	[35,1;36,2]	[2,8;3,6]	4,3
		Équivalent temps	[1,4;2,5]selon mode+modèle	[2,6;2,9] selon mode+modèle	[13,5;16,7]	[20,3;21,6]

Notes : Consentement à payer en €. Équivalent temps en minutes. Prob. = probabilité. Les fourchettes de valeur indiquent les valeurs minimales et maximales estimées selon le modèle.

Bibliography

Abou-Zeid, M. and Ben-Akiva, M. (2011). The effect of social comparisons on commute well-being. Transportation Research Part A: Policy and Practice, 45(4):345–361. (Cited on pages 109, 130, 134, 136, 137, and 352.)

Abou-Zeid, M. and Ben-Akiva, M. (2012). Travel mode switching: Comparison of findings from two public transportation experiments. Transport Policy, 24:48–59. (Cited on page 289.)

Abou-Zeid, M., Ben-Akiva, M., Bierlaire, M., Choudhury, C., and Hess, S. (2010). Attitudes and value of time heterogeneity. (Cited on page 149.)

Abou-Zeid, M., Witter, R., Bierlaire, M., Kaufmann, V., and Ben-Akiva, M. (2012). Happiness and travel mode switching: findings from a swiss public transportation experiment. Transport Policy, 19(1):93–104. (Cited on page 109.)

Abrantes, P. A. and Wardman, M. R. (2011). Meta-analysis of uk values of travel time: An update. Transportation Research Part A: Policy and Practice, 45(1):1–17. (Cited on pages 148, 261, 352, and 353.)

Aïchi, L. (2015). Rapport au nom de la commission d'enquête sur le coût économique et financier de la pollution de l'air. Technical report, Sénat. (Cited on page 3.)

Agresti, A. (2013). Categorical Data Analysis. Wiley, New York, third edition. (Cited on pages 237 and 241.)

Ajzen, I. (1985). From Intentions to Actions: A Theory of Planned Behavior. SSSP Springer Series in Social Psychology. Springer Berlin Heidelberg. (Cited on page 154.)

Ajzen, I. (1991). The theory of planned behavior. Organizational behavior and human decision processes, 50(2):179–211. (Cited on pages 10, 135, 176, and 295.)

Algers, S., Bergström, P., Dahlberg, M., and Lindqvist Dillén, J. (1998). Mixed logit estimation of the value of travel time. Technical report, Working Paper, Department of Economics, Uppsala University. (Cited on pages 148 and 352.)

Allais, M. (1953). La psychologie de l'homme rationnel devant le risque: la théorie et l'expérience. Journal de la Société de Statistique de Paris, 94:47–73. (Cited on pages 13, 206, and 217.)

Allcott, H., Mullainathan, S., and Taubinsky, D. (2014). Energy policy with externalities and internalities. Journal of Public Economics, 112:72–88. (Cited on page 3.)

Allcott, H. and Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. The American Economic Review, 104(10):3003–3037. (Cited on page 11.)

Amemiya, T. (1981). Qualitative response models: A survey. Journal of economic literature, 19(4):1483–1536. (Cited on page 239.)

Anable, J. (2005). 'complacent car addicts' or 'aspiring environmentalists'? identifying travel behaviour segments using attitude theory. Transport Policy, 12(1):65–78. (Cited on page 109.)

Anderson, M. W. (2012). New ecological paradigm (nep) scale. The Berskshire Encyclopedia of Sustanability: Measurements, Indicators, and research Methods for Sustainability, pages 260–262. (Cited on page 179.)

Anderson, S. P., de Palma, A., and Thisse, J. F. (1992). Discrete choice theory of product differentiation. MIT press. (Cited on page 15.)

Andersson, H., Jonsson, L., and Ögren, M. (2010). Property prices and exposure to multiple noise sources: Hedonic regression with road and railway noise. Environmental and resource economics, 45(1):73–89. (Cited on page 3.)

Anwar, A. M., Tieu, A. K., Gibson, P., Berryman, M. J., Win, K. T., McCusker, A., and Perez, P. (2014). Temporal and parametric study of traveller preference heterogeneity using random parameter logit model. International Journal for Traffic and Transport Engineering, 4(4):437–455. (Cited on pages 125, 129, 132, 136, 138, 139, 141, 142, 145, 157, 348, and 350.)

Arentze, T. A. and Molin, E. J. (2013). Travelers' preferences in multimodal networks: Design and results of a comprehensive series of choice experiments. Transportation Research Part A: Policy and Practice,

58:15–28. (Cited on pages 21, 47, 50, 62, 81, 168, and 261.)
Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R., and Schuman, H. (1993). Report of the noaa panel on contingent valuation. (Cited on page 39.)
Atasoy, B., Glerum, A., and Bierlaire, M. (2011). Attitudes towards mode choice in switzerland. Technical Report TRANSP-OR 110502, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne. (Cited on pages 154, 185, 186, and 187.)
Atasoy, B., Glerum, A., and Bierlaire, M. (2013). Attitudes towards mode choice in switzerland. disP - The Planning Review, 49(2):101–117. (Cited on pages 124, 129, 139, 141, and 144.)
Aulong, S., Rinaudo, J.-D., and Bouscasse, H. (2006). Assessing the costs and benefits of groundwater quality improvement in the upper rhine valley quaternary aquifer (france). Technical report, BRGM - BRIDGE European Project (FP6). (Cited on page 36.)
Avineri, E. (2012). On the use and potential of behavioural economics from the perspective of transport and climate change. Journal of Transport Geography, 24:512–521. (Cited on page 12.)
Bahamonde-Birke, F. J., Kunert, U., Link, H., and de Dios Ortúzar, J. (2017). About attitudes and perceptions: finding the proper way to consider latent variables in discrete choice models. Transportation, 44(3):475–493. (Cited on page 115.)
Balliet, D., Mulder, L. B., and Van Lange, P. A. (2011). Reward, punishment, and cooperation: a meta-analysis. (Cited on page 119.)
Baltas, G. and Doyle, P. (2001). Random utility models in marketing research: a survey. Journal of Business Research, 51(2):115–125. (Cited on pages 19 and 239.)
Bamberg, S., Ajzen, I., and Schmidt, P. (2003a). Choice of travel mode in the theory of planned behavior: The roles of past behavior, habit, and reasoned action. Basic and applied social psychology, 25(3):175–187. (Cited on pages 10 and 135.)
Bamberg, S., Hunecke, M., and Blöbaum, A. (2007). Social context, personal norms and the use of public transportation: Two field studies. Journal of Environmental Psychology, 27(3):190–203. (Cited on page 135.)
Bamberg, S., Rölle, D., and Weber, C. (2003b). Does habitual car use not lead to more resistance to change of travel mode? Transportation, 30(1):97–108. (Cited on pages 11, 109, and 178.)
Bamberg, S. and Schmidt, P. (2003). Incentives, morality, or habit? predicting students' car use for university routes with the models of ajzen, schwartz, and triandis. Environment and behavior, 35(2):264–285. (Cited on pages 10, 114, 135, and 178.)
Baron, R. M. and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology, 51(6):1173. (Cited on page 188.)
Barr, S. (2007). Factors influencing environmental attitudes and behaviors a uk case study of household waste management. Environment and behavior, 39(4):435–473. (Cited on page 176.)
Batarce, M. and Ivaldi, M. (2014). Urban travel demand model with endogenous congestion. Transportation Research Part A: Policy and Practice, 59:331–345. (Cited on page 214.)
Bateman, I., Munro, A., Rhodes, B., Starmer, C., and Sugden, R. (1997). A test of the theory of reference-dependent preferences. The quarterly journal of economics, 112(2):479–505. (Cited on pages 272 and 292.)
Bates, J., Polak, J., Jones, P., and Cook, A. (2001). The valuation of reliability for personal travel. Transportation Research Part E: Logistics and Transportation Review, 37(2):191–229. (Cited on pages 67, 69, 70, and 214.)
Batley, R. and Ibáñez, J. N. (2012). Randomness in preference orderings, outcomes and attribute tastes: An application to journey time risk. Journal of choice modelling, 5(3):157–175. (Cited on pages 66, 67, 68, 69, and 70.)
Baumeister, R. F., Bratslavsky, E., Finkenauer, C., and Vohs, K. D. (2001). Bad is stronger than good. Review of general psychology, 5(4):323. (Cited on page 216.)
Bayart, C. (2009). Le potentiel du web pour les enquêtes de mobilité urbaine. PhD thesis, Université de Lyon. (Cited on page 40.)
Bayart, C. and Bonnel, P. (2010). L'impact du mode d'enquête sur la mesure des comportements de mobilité. Economie et statistique, 437(1):47–70. (Cited on page 38.)

Beaud, M., Blayac, T., and Stéphan, M. (2016). The impact of travel time variability and travelers' risk attitudes on the values of time and reliability. Transportation Research Part B: Methodological, 93:207–224. (Cited on pages 207 and 226.)

Beaumais, O., Casabianca, A., Pieri, X., and Prunetti, D. (2016). Why not allow individuals to rank freely? a scaled rank-ordered logit approach applied to waste management in corsica. Annals of Economics and Statistics/Annales d'Économie et de Statistique, (121/122):187–212. (Cited on page 37.)

Becker, G. (1965). A Theory of the Allocation of Time. The Economic Journal, 75(299):493–517. (Cited on pages 148 and 159.)

Beirão, G. and Cabral, J. S. (2007). Understanding attitudes towards public transport and private car: A qualitative study. Transport policy, 14(6):478–489. (Cited on pages 109 and 287.)

Bekhor, S. and Albert, G. (2014). Accounting for sensation seeking in route choice behavior with travel time information. Transportation research part F: traffic psychology and behaviour, 22:39–49. (Cited on page 111.)

Ben-Akiva, M. and Boccara, B. (1987). Integrated framework for travel behavior analysis. In International Association of Travel Behavior Research (IATBR) Conference, Aix-en-Provence, France. (Cited on page 109.)

Ben-Akiva, M., de Palma, A., McFadden, D., Abou-Zeid, M., Chiappori, P.-A., de Lapparent, M., Durlauf, S. N., Fosgerau, M., Fukuda, D., Hess, S., et al. (2012). Process and context in choice models. Marketing Letters, 23(2):439–456. (Cited on pages 14 and 208.)

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Börsch-Supan, A., Brownstone, D., Bunch, D. S., and others (2002a). Hybrid choice models: progress and challenges. Marketing Letters, 13(3):163–175. (Cited on pages 19 and 348.)

Ben-Akiva, M. and Morikawa, T. (2002). Comparing ridership attraction of rail and bus. Transport Policy, 9(2):107–116. (Cited on pages 4 and 21.)

Ben-Akiva, M. and Swait, J. (1986). The akaike likelihood ratio index. Transportation Science, 20(2):133–136. (Cited on pages 225, 226, 254, 255, and 256.)

Ben-Akiva, M., Walker, J., Bernardino, A. T., Gopinath, D. A., Morikawa, T., and Polydoropoulou, A. (2002b). Integration of choice and latent variable models. Perpetual motion: Travel behaviour research opportunities and application challenges, pages 431–470. (Cited on pages 4, 5, 19, 110, 133, 150, and 282.)

Ben-Akiva, M. E. and Lerman, S. R. (1985). Discrete choice analysis: theory and application to travel demand, volume 9. MIT press. (Cited on page 35.)

Ben-Elia, E. and Ettema, D. (2009). Carrots versus sticks: Rewarding commuters for avoiding the rush-hour—a study of willingness to participate. Transport policy, 16(2):68–76. (Cited on page 119.)

Bergdahl, M., Ehling, M., Elvers, E., Földesi, E., Körner, T., Kron, A., Lohauß, P., Mag, K., Morais, V., Nimmergut, A., et al. (2007). Handbook on data quality assessment methods and tools. Technical report, Eurostat. (Cited on page 39.)

Berliner, R. M., Malokin, A., Circella, G., and Mokhtarian, P. L. (2015). Travel-based multitasking: modeling the propensity to conduct activities while commuting. In Transportation Research Board 94th - Annual Meeting. (Cited on pages 119 and 181.)

Bhat, C. R. (1995). A heteroscedastic extreme value model of intercity mode choice. Transportation Research Part B, 29:471–483. (Cited on page 4.)

Bhat, C. R. (2015). A new generalized heterogeneous data model (ghdm) to jointly model mixed types of dependent variables. Transportation Research Part B: Methodological, 79:50–77. (Cited on page 289.)

Bhat, C. R. and Castelar, S. (2002). A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the san francisco bay area. Transportation Research Part B: Methodological, 36(7):593–616. (Cited on page 289.)

Bhat, C. R. and Dubey, S. K. (2014). A new estimation approach to integrate latent psychological constructs in choice modeling. Transportation Research Part B: Methodological, 67:68–85. (Cited on pages 115, 116, 118, 122, 124, 129, 130, 132, and 290.)

Bhat, C. R. and Sardesai, R. (2006). The impact of stop-making and travel time reliability on commute mode choice. Transportation Research Part B: Methodological, 40(9):709–730. (Cited on page 214.)

Bhattacharya, T., Brown, J., Jaroszynski, M., and Batuhan, T. (2014). The effects of perception vs."reality" on travel behavior after a major transit service change: The case of Tallahassee, Florida. Journal of Public Transportation, 17(2):1. (Cited on pages 8, 144, and 181.)

Bierlaire, M. (2016). Pythonbiogeme: a short introduction. Technical report, Transp-OR, Ecole Polytechnique Federale de Lausanne. (Cited on pages 93, 160, and 224.)

Bliemer, M. C. and Rose, J. M. (2010). Serial choice conjoint analysis for estimating discrete choice models. In Choice Modelling: The State-of-the-art and the State-of-practice: Proceedings from the Inaugural International Choice Modelling Conference, pages 137–161. Emerald Group Publishing Limited. (Cited on page 59.)

Bliemer, M. C., Rose, J. M., and Hensher, D. A. (2009). Efficient stated choice experiments for estimating nested logit models. Transportation Research Part B: Methodological, 43(1):19–35. (Cited on page 56.)

Block, H. and Marschak, J. (1960). Random OrdOrder and Sochastic Theories of Reponse. Stanford University Press. (Cited on page 18.)

Bollen, K. and Hoyle, R. (2012). Latent variables in structural equation modeling. In Handbook of Structural Equation Modeling. Guilford Press. (Cited on page 130.)

Bollen, K. A. (2014). Structural equations with latent variables. John Wiley & Sons. (Cited on page 132.)

Bonnel, P. (2002). Prévision de la demande de transport. PhD thesis, Habilitation à Diriger la Recherche - Université Lumière-Lyon II. (Cited on page 29.)

Börjesson, M. and Eliasson, J. (2011). On the use of "average delay" as a measure of train reliability. Transportation Research Part A: Policy and Practice, 45(3):171–184. (Cited on pages 21, 38, 62, 66, 67, 68, 69, 70, and 99.)

Börjesson, M., Eliasson, J., and Franklin, J. P. (2012). Valuations of travel time variability in scheduling versus mean–variance models. Transportation Research Part B: Methodological, 46(7):855–873. (Cited on pages 62, 65, 67, 68, 69, 70, 148, and 214.)

Bouscasse, H., Joly, I., and Peyhardi, J. (2017). Comparison of link functions specifications of generalized linear models: application to travel mode choice. In Transportation Research Board - Annual Meeting. (Cited on page 233.)

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. Journal of public health, 27(3):281–291. (Cited on pages 38, 39, and 42.)

Brownstone, D. and Small, K. A. (2005). Valuing time and reliability: assessing the evidence from road pricing demonstrations. Transportation Research Part A: Policy and Practice, 39(4):279–293. (Cited on page 66.)

Bruhin, A., Fehr-Duda, H., and Epper, T. (2010). Risk and rationality: Uncovering heterogeneity in probability distortion. Econometrica, 78(4):1375–1412. (Cited on page 207.)

Buttel, F. H. (1979). Age and environmental concern: A multivariate analysis. Youth & Society, 10(3):237–256. (Cited on page 180.)

Cameron, A. C. and Trivedi, P. K. (1986). Econometric models based on count data. comparisons and applications of some estimators and tests. Journal of applied econometrics, 1(1):29–53. (Cited on page 238.)

Carlsson, F. (2010). Design of stated preference surveys: Is there more to learn from behavioral economics? Environmental and Resource Economics, 46(2):167–177. (Cited on page 41.)

Casals, J. (2012). Étude exploratoire sur les perceptions du temps et les activités dans les transports en commun - rapport de l'enquête quantitative. Technical report, Nova 7. (Cited on pages 149, 154, 181, 182, 186, and 187.)

Charness, G. and Gneezy, U. (2009). Incentives to exercise. Econometrica, 77(3):909–931. (Cited on page 11.)

Chateauneuf, A. and Cohen, M. (1994). Risk seeking with diminishing marginal utility in a non-expected utility model. Journal of Risk and Uncertainty, 9(1):77–91. (Cited on pages 207, 211, and 290.)

Chateauneuf, A., Cohen, M., and Meilijson, I. (2005). More pessimism than greediness: a characterization of monotone risk aversion in the rank-dependent expected utility model. Economic Theory, 25(3):649–667. (Cited on page 207.)

Cherchi, E. (2009). Modelling individual preferences: State of the art, recent advances and future directions. (Cited on page 12.)

Cherchi, E., T.H., R., Börjesson, M., and Bierlaire, M. (2017). Habitual latent behaviour and dynamic effect of inertia. lagged-based versus hazard-based hybrid mode choice model. In Transportation Research Board - Annual Meeting. (Cited on page 186.)

Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. The American Economic Review, 105(5):1–33. (Cited on page 108.)

Chorus, C. (2012a). What about behaviour in travel demand modelling? an overview of recent progress. Transportation Letters, 4(2):93–104. (Cited on page 12.)

Chorus, C. G. (2010). A new model of random regret minimization. European Journal of Transport and Infrastructure Research, 2(10):181–196. (Cited on pages 15 and 209.)

Chorus, C. G. (2012b). Random regret-based discrete choice modeling. Technical report, Delft University of Technology. (Cited on pages 209 and 211.)

Chorus, C. G. (2014). Capturing alternative decision rules in travel choice models: a critical discussion. Edward Elgar Publishing. (Cited on page 15.)

Chorus, C. G., Arentze, T. A., and Timmermans, H. J. (2008). A random regret-minimization model of travel choice. Transportation Research Part B: Methodological, 42(1):1–18. (Cited on pages 4, 15, and 209.)

Chorus, C. G. and Kroesen, M. (2014). On the (im-) possibility of deriving transport policy implications from hybrid choice models. Transport Policy, 36:217–222. (Cited on page 146.)

Cialdini, R.B. ad Kallgren, C. and Reno, R. (1991). A focus theory of normative conduct: a theoretical refinement and reevaluation of the role of norms in human behavior. Advances in Experimental Social Psychology, 24:201–235. (Cited on pages 12 and 180.)

Cohen, M. (2015). Risk perception, risk attitude, and decision: a rank-dependent analysis. Mathematical Population Studies, 22(1):53–70. (Cited on pages 13, 207, 211, 220, and 290.)

Collins, C. M. and Chambers, S. M. (2005). Psychological and situational influences on commuter-transport-mode choice. Environment and behavior, 37(5):640–661. (Cited on pages 109, 135, 179, and 291.)

Conte, A., Hey, J. D., and Moffatt, P. G. (2011). Mixture models of choice under risk. Journal of Econometrics, 162(1):79–88. (Cited on pages 20 and 207.)

Córdoba, J. and Jaramillo, G. (2012). Inclusion of the latent personality variable in multinomial logit models using the 16pf psychometric test. Procedia-Social and Behavioral Sciences, 54:169–178. (Cited on pages 125, 139, and 348.)

Croson, R. and Treich, N. (2014). Behavioral environmental economics: promises and challenges. Environmental and Resource Economics, 58(3):335–351. (Cited on pages 9, 177, 201, 202, and 287.)

Daly, A., Hess, S., and Train, K. (2012). Assuring finite moments for willingness to pay in random coefficient models. Transportation, 39(1):19–31. (Cited on page 148.)

Davies, A.-M. and Laing, R. (2002). Designing choice experiments using focus groups: Results from an aberdeen case study. In Forum Qualitative Sozialforschung/Forum: Qualitative Social Research, volume 3. (Cited on page 52.)

Daziano, R. A. (2015). Inference on mode preferences, vehicle purchases, and the energy paradox using a bayesian structural choice model. Transportation Research Part B: Methodological, 76:1–26. (Cited on pages 116, 118, 124, 125, 126, 129, 130, 133, 138, 145, and 348.)

Daziano, R. A. and Bolduc, D. (2013a). Covariance, identification, and finite-sample performance of the msl and bayes estimators of a logit model with latent attributes. Transportation, 40(3):647–670. (Cited on pages 124, 132, and 133.)

Daziano, R. A. and Bolduc, D. (2013b). Incorporating pro-environmental preferences towards green automobile technologies through a bayesian hybrid choice model. Transportmetrica A: Transport Science, 9(1):74–106. (Cited on page 111.)

Daziano, R. A. and Rizzi, L. I. (2015). Analyzing the impact of a fatality index on a discrete, interurban mode choice model with latent safety, security, and comfort. Safety science, 78:11–19. (Cited on pages 4, 111, 116, 125, 137, 138, 141, 149, 266, 348, 349, and 350.)

De Borger, B. and Fosgerau, M. (2008). The trade-off between money and travel time: A test of the theory of reference-dependent preferences. Journal of Urban Economics, 64(1):101–115. (Cited on page 209.)

De Borger, B. and Glazer, A. (2017). Support and opposition to a pigovian tax: Road pricing with reference-dependent preferences. Journal of Urban Economics, 99:31–47. (Cited on page 209.)

De Groot, J. and Steg, L. (2007). General beliefs and the theory of planned behavior: The role of environmental concerns in the TPB. Journal of Applied Social Psychology, 37(8):1817–1836. (Cited on pages 5, 10, 114, 135, 176, 177, and 178.)

De Jong, G. and Gunn, H. (2001). Recent evidence on car cost and time elasticities of travel demand in europe. Journal of Transport Economics and Policy, 35(2):137–160. (Cited on pages 104 and 266.)

de Lapparent, M., Axhausen, K., and Frei, A. (2013). Long Distance Mode Choice and Distributions of Values of Travel Time Savings in Three European Countries. European Transport, 53:pp.1–31. (Cited on page 148.)

de Lapparent, M. and Ben-Akiva, M. (2014). Risk aversion in travel mode choice with rank-dependent utility. Mathematical Population Studies, 21(4):189–204. (Cited on pages 14, 214, 215, 216, 220, 227, and 229.)

de Lapparent, M. and Koning, M. (2016). Analyzing Time Sensitivity to Discomfort in the Paris Subway: an Interval Data Model Approach. Transportation, 43(5):913–933. (Cited on pages 148 and 160.)

de Palma, A., Ben-Akiva, M., Brownstone, D., Holt, C., Magnac, T., McFadden, D., Moffatt, P., Picard, N., Train, K., Wakker, P., et al. (2008). Risk, uncertainty and discrete choice models. Marketing Letters, 19(3-4):269–285. (Cited on pages 214, 217, 218, and 220.)

de Palma, A. and Picard, N. (2005). Route choice decision under travel time uncertainty. Transportation Research Part A: Policy and Practice, 39(4):295–324. (Cited on pages 216 and 230.)

de Palma, A., Picard, N., and de Lapparent, M. (2015). Risk and uncertainty in urban and transport economics. Mathematical Population Studies, 22(1):1–3. (Cited on page 214.)

de Palma, A., Picard, N., and Inoa, I. (2014). Discrete Choice Decision-making with Multiple Decision-makers Within the Household, page 363. Edward Elgar Publishing. (Cited on page 9.)

de Palma, A., Picard, N., and Ziegelmeyer, A. (2011). Individual and couple decision behavior under risk: evidence on the dynamics of power balance. Theory and decision, 70(1):45–64. (Cited on page 9.)

de Palma, A. and Thisse, J.-F. (1989). Les modèles de choix discrets. Annales d'Economie et de Statistique, pages 151–190. (Cited on page 18.)

de Palma, A. and Zaouali, N. (2007). Monétarisation des externalités de transport : un état de l'art. Technical report, THEMA, Working Paper. (Cited on page 3.)

De Witte, A., Hollevoet, J., Dobruszkes, F., Hubert, M., and Macharis, C. (2013). Linking modal choice to motility: A comprehensive review. Transportation Research Part A: Policy and Practice, 49:329–341. (Cited on pages 7, 8, 122, 164, 183, 239, and 398.)

Denant-Boèmont, L. and L'Haridon, O. (2013). La rationalité à l'épreuve de l'économie comportementale. Revue française d'économie, 28(2):35–89. (Cited on pages 13 and 217.)

Denant-Boèmont, L. and Petiot, R. (2003). Information value and sequential decision-making in a transport setting: an experimental study. Transportation Research Part B: Methodological, 37(4):365–386. (Cited on page 289.)

Denant-Boèmont, L. and Hammiche, S. (2010). Flexibility of transport choice in a real-option setting: An experimental case study. Journal of Intelligent Transportation Systems, 14(3):140–153. (Cited on page 289.)

DeSerpa, A. C. (1971). A Theory of the Economics of Time. The Economic Journal, 81(324):828–846. (Cited on page 148.)

Di Ciommo, F., Monzón, A., and Fernandez-Heredia, A. (2013). Improving the analysis of road pricing acceptability surveys by using hybrid models. Transportation Research Part A: Policy and Practice, 49:302–316. (Cited on page 111.)

Di Tella, R. and MacCulloch, R. (2006). Some uses of happiness data in economics. The Journal of Economic Perspectives, 20(1):25–46. (Cited on page 109.)

Diecidue, E. and Wakker, P. P. (2001). On the intuition of rank-dependent utility. Journal of Risk and Uncertainty, 23(3):281–298. (Cited on pages 207 and 218.)

Dillman, D. A. and Bowker, D. K. (2001). The web questionnaire challenge to survey methodologists. Bemad Batinic, Ulf-Dietrich Reips, Michael Bosnjak & Andreas Werner: Online Social Sciences. Seattle: Hogrefe & Huber, pages 53–71. (Cited on pages 40 and 43.)

Dittmar, H. (1992). The social psychology of material possessions: to have is to be. Journal of Economic Psychology, 14(2):453–454. (Cited on page 12.)

Dobson, A. J. and Barnett, A. (2008). An introduction to generalized linear models. CRC press. (Cited on page 361.)

Donald, I. J., Cooper, S. R., and Conchie, S. (2014). An extended theory of planned behaviour model of the psychological factors affecting commuters' transport mode use. Journal of environmental psychology, 40:39–48. (Cited on page 10.)

Dugundji, E. and Walker, J. (2005). Discrete choice with social and spatial network interdependencies: an empirical example using mixed generalized extreme value models with field and panel effects. Transportation Research Record: Journal of the Transportation Research Board, (1921):70–78. (Cited on page 108.)

Dunlap, R., Liere, K., Mertig, A., and Jones, R. E. (2000). Measuring endorsement of the new ecological paradigm: A revised nep scale. Journal of Social Issues, 56(3):425–442. (Cited on pages 127, 180, 185, and 296.)

Dunlap, R. E. (2008). The New Environmental Paradigm Scale: from Marginality to Worldwide Use. The journal of environmental education, 40(1):3–18. (Cited on pages 179 and 180.)

Dunlap, R. E. and Van Liere, K. D. (1978). The "new environmental paradigm". The Journal of Environmental Education, 9(4):10–19. (Cited on page 179.)

Dussaix, A.-M. (2009). La qualité dans les enquêtes. revue Modulad, 137(39). (Cited on page 39.)

Eriksson, L., Garvill, J., and Nordlund, A. M. (2008). Interrupting habitual car use: The importance of car habit strength and moral motivation for personal car use reduction. Transportation Research Part F: Traffic Psychology and Behaviour, 11(1):10–23. (Cited on pages 11 and 186.)

Esposito Vinzi, V. and Trinchera, L. (2013). Modèles à équations structurelles, approches basées sur les composantes. In Modèles à variables latentes et modèles de mélange. Technip. (Cited on page 188.)

Eurostat (2016). Passenger Transport Statistics. http://ec.europa.eu/eurostat/statistics-explained/index.php/Passenger_transport_statistics. (Cited on page 176.)

Faburel, G. (2002). Évaluer les coûts sociaux: la nécessité de l'interdisciplinarité: Application au bruit des avions. In Annales des ponts et chaussées, volume 2002, pages 65–73. Elsevier. (Cited on page 36.)

Feng, D., Gourieroux, C., and Jasiak, J. (2008). The ordered qualitative model for credit rating transitions. Journal of Empirical Finance, 15(1):111–130. (Cited on page 238.)

Fernández-Antolín, A., Guevara-Cue, A., de Lapparent, M., and Bierlaire, M. (2016). Correcting for endogeneity due to omitted attitudes: Empirical assessment of a modified mis method using rp mode choice data. Journal of Choice Modelling, 20:1–15. (Cited on pages 112, 116, 124, 131, and 149.)

Festinger, L. (1954). A theory of social comparison processes. Human Relations, 7:117–140. (Cited on pages 12, 176, and 180.)

Fiebig, D., Keane, M., Louviere, J., and Wasi, N. (2010). The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. Marketing Science, 29:3:393–421. (Cited on page 93.)

Fishbein, M. and Ajzen, I. (1975). Belief, attitude, intention and behavior: An introduction to theory and research. Addison-Wesley Pub. Co. (Cited on pages 178 and 295.)

FitzRoy, F. and Smith, I. (1995). The demand for rail transport in european countries. Transport Policy, 2(3):153–158. (Cited on pages 4 and 63.)

Flynn, T. N., Louviere, J. J., Peters, T. J., and Coast, J. (2008). Estimating preferences for a dermatology consultation using best-worst scaling: comparison of various methods of analysis. BMC Medical Research Methodology, 8(1):1. (Cited on page 257.)

Folmer, H. and Johansson-Stenman, O. (2011). Does environmental economics produce aeroplanes without engines? on the need for an environmental social science. Environmental and Resource Economics, 48(3):337–361. (Cited on page 111.)

Fosgerau, M. (2006). Investigating the distribution of the value of travel time savings. Transportation Research Part B: Methodological, 40(8):688–707. (Cited on page 148.)

Fosgerau, M. (2007). Using nonparametrics to specify a model to measure the value of travel time. Transportation Research Part A: Policy and Practice, 41(9):842–856. (Cited on page 148.)

Fosgerau, M. and Bierlaire, M. (2007). A practical test for the choice of mixing distribution in discrete choice models. Transportation Research Part B: Methodological, 41(7):784–794. (Cited on page 148.)
Fosgerau, M. and De Borger, B. (2008). Hedonic preferences, symmetric loss aversion and the willingness to pay-willingness to accept gap. (Cited on page 209.)
Fosgerau, M., Hjorth, K., and Lyk-Jensen, S. V. (2010). Between-mode-differences in the value of travel time: Self-selection or strategic behaviour? Transportation research part D: transport and environment, 15(7):370–381. (Cited on pages 41 and 353.)
Fowkes, A. (1998). The development of stated preference techniques in transport planning. Technical report. (Cited on page 46.)
Fowkes, T. and Wardman, M. (1988). The design of stated preference travel choice experiments: with special reference to interpersonal taste variations. Journal of transport economics and policy, pages 27–44. (Cited on page 46.)
Fox, C. R. and Tversky, A. (1995). Ambiguity aversion and comparative ignorance. The quarterly journal of economics, 110(3):585–603. (Cited on page 215.)
Fransson, N. and Gärling, T. (1999). Environmental concern: Conceptual definitions, measurement methods, and research findings. Journal of environmental psychology, 19(4):369–382. (Cited on page 179.)
Franzen, A. and Vogl, D. (2013). Two decades of measuring environmental attitudes: A comparative analysis of 33 countries. Global Environmental Change, 23(5):1001–1008. (Cited on page 179.)
Frey, B. S. and Stutzer, A. (2014). Economic consequences of mispredicting utility. Journal of Happiness Studies, 15(4):937–956. (Cited on page 109.)
Fullerton, A. S. (2009). A conceptual framework for ordered logistic regression models. Sociological methods & research, 38(2):306–347. (Cited on page 239.)
Gaker, D., Zheng, Y., and Walker, J. (2010). Experimental economics in transportation: focus on social influences and provision of information. Transportation Research Record: Journal of the Transportation Research Board, (2156):47–55. (Cited on pages 12 and 289.)
Gärling, T., Fujii, S., Gärling, A., and Jakobsson, C. (2003). Moderating effects of social value orientation on determinants of proenvironmental behavior intention. Journal of environmental psychology, 23(1):1–9. (Cited on page 178.)
Gärling, T. and Schuitema, G. (2007). Travel demand management targeting reduced private car use: effectiveness, public acceptability and political feasibility. Journal of Social Issues, 63(1):139–153. (Cited on page 119.)
Garvill, J. (1999). Choice of transportation mode: Factors influencing drivers' willingness to reduce personal car use and support car regulations, pages 263–279. Taylor and Francis, Philadelphia. (Cited on pages 5 and 176.)
Gifford, R., Steg, L., and Reser, J. P. (2011). Environmental psychology, pages 440–470. Wiley Online Library. (Cited on pages 10, 109, 133, 178, and 296.)
Gigerenzer, G. and Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. Topics in Cognitive Science, 1(1):107–143. (Cited on page 49.)
Gill, D. and Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. The American economic review, 102(1):469–503. (Cited on page 292.)
Glerum, A., Atasoy, B., and Bierlaire, M. (2014). Using semi-open questions to integrate perceptions in choice models. Journal of Choice Modelling, 10:11–33. (Cited on pages 103, 104, 124, 125, 126, 130, 136, 141, 144, 145, 168, 261, 266, 348, and 350.)
Golob, T. F. (2003). Structural equation modeling for travel behavior research. Transportation Research Part B: Methodological, 37(1):1–25. (Cited on page 187.)
Goodman, L. A. (1983). The analysis of dependence in cross-classifications having ordered categories, using log-linear models for frequencies and log-linear models for odds. Biometrics, pages 149–160. (Cited on page 238.)
Green, C. and Tunstall, S. (1999). A psychological perspective. Oxford University Press. (Cited on page 42.)
Green, P. E. and Srinivasan, V. (1990). Conjoint analysis in marketing: new developments with implications for research and practice. The Journal of Marketing, pages 3–19. (Cited on page 50.)

Greene, W. H. (2003). Econometric Analysis. Pearson Education, New York, fifth edition. (Cited on pages 250 and 278.)

Greene, W. H. and Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B: Methodological, 37(8):681–698. (Cited on pages 20, 217, 221, 276, and 280.)

Gruber, J. and Kőszegi, B. (2004). Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes. Journal of Public Economics, 88(9):1959–1987. (Cited on page 3.)

Guagnano, G. A., Stern, P. C., and Dietz, T. (1995). Influences on attitude-behavior relationships a natural experiment with curbside recycling. Environment and behavior, 27(5):699–718. (Cited on pages 135, 179, and 295.)

Gunn, H. and Rohr, C. (1996). The 1985–1996 dutch value of time studies. In PTRC International Conference on the Value of Time. (Cited on page 353.)

Habib, K. M. N., Tian, Y., and Zaman, H. (2011). Modelling commuting mode choice with explicit consideration of carpool in the choice set formation. Transportation, 38(4):587–604. (Cited on pages 129 and 348.)

Habib, K. M. N. and Zaman, M. H. (2012). Effects of incorporating latent and attitudinal information in mode choice models. Transportation Planning and Technology, 35(5):561–576. (Cited on pages 125, 129, and 130.)

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., and Tatham, R. L. (2006). Multivariate data analysis, volume 6. Pearson Prentice Hall Upper Saddle River, NJ. (Cited on page 191.)

Hauser, R. M. and Goldberger, A. S. (1971). The treatment of unobservable variables in path analysis. Sociological methodology, 3:81–117. (Cited on page 109.)

Hayakawa, H. (2000). Bounded rationality, social and cultural norms, and interdependence via reference groups. Journal of Economic Behavior & Organization, 43(1):1–34. (Cited on page 9.)

Hayes, A. (2013a). Model templates for process for spss and sas. Technical report, The Guilford Press. (Cited on page 188.)

Hayes, A. F. (2013b). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. Guilford Press. (Cited on pages 117, 123, and 131.)

Haywood, L. and Koning, M. (2015). The distribution of crowding costs in public transport: New evidence from paris. Transportation Research Part A: Policy and Practice, 77:182–201. (Cited on pages 148 and 160.)

Heath, Y. and Gifford, R. (2002). Extending the theory of planned behavior: predicting the use of public transportation1. Journal of Applied Social Psychology, 32(10):2154–2189. (Cited on pages 135 and 178.)

Hensher, D., Louviere, J., and Swait, J. (1999). Combining sources of preference data. Journal of Econometrics, 89(1):197–221. (Cited on page 289.)

Hensher, D., Rose, J., and Greene, W. (2015). Applied Choice Analysis. Number 1188. Cambridge University Press. (Cited on pages 214, 215, 216, and 217.)

Hensher, D. A. (1994). Stated preference analysis of travel choices: the state of practice. Transportation, 21(2):107–133. (Cited on page 35.)

Hensher, D. A. (2006a). How do respondents process stated choice experiments? attribute consideration under varying information load. Journal of applied econometrics, 21(6):861–878. (Cited on page 48.)

Hensher, D. A. (2006b). Towards a practical method to establish comparable values of travel time savings from stated choice experiments with differing design dimensions. Transportation Research Part A: Policy and Practice, 40(10):829–840. (Cited on pages 48, 49, and 148.)

Hensher, D. A. (2007). Bus transport: economics, policy and planning, volume 18. Elsevier. (Cited on page 103.)

Hensher, D. A. (2010). Attribute processing, heuristics and preference construction in choice analysis. In Choice Modelling: The State-of-the-art and The State-of-practice: Proceedings from the Inaugural International Choice Modelling Conference, pages 35–69. Emerald Group Publishing Limited. (Cited on page 15.)

Hensher, D. A., Greene, W. H., and Li, Z. (2011). Embedding risk attitude and decision weights in non-linear logit to accommodate time variability in the value of expected travel time savings.

Transportation research part B: methodological, 45(7):954–972. (Cited on pages 208, 214, 216, and 217.)
Hensher, D. A. and Li, Z. (2012). Valuing travel time variability within a rank-dependent utility framework and an investigation of unobserved taste heterogeneity. Journal of Transport Economics and Policy, pages 293–312. (Cited on pages 14, 207, 214, 216, 227, and 229.)
Hensher, D. A., Rose, J. M., Leong, W., Tirachini, A., and Li, Z. (2013). Choosing public transport—incorporating richer behavioural elements in modal choice models. Transport Reviews, 33(1):92–106. (Cited on page 12.)
Herne, K. (1998). Testing the reference-dependent model: An experiment on asymmetrically dominated reference points. Journal of Behavioral Decision Making, 11(3):181–192. (Cited on page 272.)
Herrnstein, R. J., Loewenstein, G. F., Prelec, D., and Vaughan, W. (1993). Utility maximization and melioration: Internalities in individual choice. Journal of behavioral decision making, 6(3):149–185. (Cited on pages 3 and 108.)
Hess, S. (2014). Impact of unimportant attributes in stated choice surveys. European Journal of Transport and Infrastructure Research, 14(4):349–361. (Cited on pages 49 and 51.)
Hess, S. and Beharry-Borg, N. (2012). Accounting for latent attitudes in willingness-to-pay studies: the case of coastal water quality improvements in tobago. Environmental and Resource Economics, 52(1):109–131. (Cited on page 111.)
Hess, S., Bierlaire, M., and Polak, J. W. (2005). Estimation of value of travel-time savings using mixed logit models. Transportation Research Part A: Policy and Practice, 39(2):221–236. (Cited on page 148.)
Hilton, D., Charalambides, L., Demarque, C., Waroquier, L., and Raux, C. (2014). A tax can nudge: The impact of an environmentally motivated bonus/malus fiscal system on transport preferences. Journal of Economic Psychology, 42:17–27. (Cited on page 108.)
Hinkin, T. R., Tracey, J. B., and Enz, C. A. (1997). Scale construction: Developing reliable and valid measurement instruments. Journal of Hospitality & Tourism Research, 21(1):100–120. (Cited on page 127.)
Hjorth, K. and Fosgerau, M. (2011). Loss aversion and individual characteristics. Environmental and Resource Economics, 49(4):573–596. (Cited on page 209.)
Hjorth, K. and Fosgerau, M. (2012). Using prospect theory to investigate the low marginal value of travel time for small time changes. Transportation Research Part B: Methodological, 46(8):917–932. (Cited on page 217.)
Hoffmann, C., Abraham, C., White, M. P., Ball, S., and Skippon, S. M. (2017). What cognitive mechanisms predict travel mode choice? A systematic review with meta-analysis. Transport Reviews, (37):1–22. (Cited on pages 10 and 109.)
Holt, C. A., Laury, S. K., et al. (2002). Risk aversion and incentive effects. American economic review, 92(5):1644–1655. (Cited on page 219.)
Homer, P. M. and Kahle, L. R. (1988). A structural equation test of the value-attitude-behavior hierarchy. Journal of Personality and social Psychology, 54(4):638. (Cited on page 135.)
Hossan, M. S., Asgari, H., and Jin, X. (2016). Investigating preference heterogeneity in value of time (vot) and value of reliability (vor) estimation for managed lanes. Transportation Research Part A: Policy and Practice, 94:638–649. (Cited on pages 148 and 352.)
Hoyle, R. (2012). Handbook of Structural Equation Modeling. Guilford Press. (Cited on pages 12, 109, 156, 187, and 188.)
Hudík, M. (2014). Reference-dependence and marginal utility: Alt, samuelson, and bernardelli. History of Political Economy, 46(4):677–693. (Cited on page 272.)
Hurez, C. and Tébar, M. (2014). The Mobility of Rhône-Alpes Inhabitants: the Construction of a New Travel Survey Protocol. In Transport Research Arena-Paris. (Cited on page 151.)
Hurtubia, R., Nguyen, M., Glerum, A., and Bierlaire, M. (2014). Integrating psychometric indicators in latent class choice models. Transportation Research Part A: Policy and Practice, 64:135–146. (Cited on pages 112 and 185.)
Idris, A. O., Habib, K. M. N., and Shalaby, A. (2015). An investigation on the performances of mode shift models in transit ridership forecasting. Transportation Research Part A: Policy and Practice, 78:551–565. (Cited on pages 116, 124, 125, 126, 129, 130, 136, 138, and 143.)
Ingold, T. (2016). Lines: a brief history. Routledge. (Cited on page 107.)

Jäckle, A., Roberts, C., and Lynn, P. (2010). Assessing the effect of data collection mode on measurement. International Statistical Review, 78(1):3–20. (Cited on page 38.)

Jensen, A. F., Cherchi, E., Mabit, S. L., and Ortúzar, J. d. D. (2014). Assesing the Impact of Direct Experience on Individual Preferences and Attitudes for Electric Vehicles. PhD thesis, Pontificia Universidad Católica de ChilePontificia Universidad Católica de Chile. (Cited on page 111.)

Jensen, M. (1999). Passion and heart in transport—a sociological analysis on transport behaviour. Transport Policy, 6(1):19–33. (Cited on page 186.)

Johansson, M. V., Heldt, T., and Johansson, P. (2006). The effects of attitudes and personality traits on mode choice. Transportation Research Part A: Policy and Practice, 40(6):507–525. (Cited on pages 111, 116, 125, 127, 129, 132, 137, 138, 139, 140, 142, 144, 149, 150, 157, 348, and 350.)

John, O. P. and Benet-Martínez, V. (2000). Measurement: Reliability, construct validation, and scale construction. Cambridge University Press. (Cited on page 127.)

Joly, I. and Vincent-Geslin, S. (2016). Intensive travel time: an obligation or a choice? European Transport Research Review, 8(1):1–14. (Cited on page 181.)

Jones, R. E. and Dunlap, R. E. (1992). The social bases of environmental concern: Have they changed over time? 1. Rural sociology, 57(1):28–47. (Cited on page 180.)

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. Biometrika, 57(2):239–251. (Cited on page 109.)

Jöreskog, K. G. (1973). G.(1973). a general method for estimating a linear structural equation system. Structural equation models in the social sciences, pages 85–112. (Cited on page 109.)

Jou, R.-C. and Chen, K.-H. (2013). An application of cumulative prospect theory to freeway drivers' route choice behaviours. Transportation Research Part A: Policy and Practice, 49:123–131. (Cited on pages 208, 214, and 229.)

Jung, S.-Y. and Yoo, K.-E. (2016). A study on passengers' airport choice behavior using hybrid choice model: A case study of seoul metropolitan area, south korea. Journal of Air Transport Management, 57:70–79. (Cited on page 111.)

Kahn, L. M. and Morimune, K. (1979). Unions and employment stability: a sequential logit approach. International Economic Review, pages 217–235. (Cited on page 238.)

Kahneman, D. and Sugden, R. (2005). Experienced utility as a standard of policy evaluation. Environmental and resource economics, 32(1):161–181. (Cited on page 108.)

Kahneman, D. and Thaler, R. H. (2006). Anomalies: Utility maximization and experienced utility. The Journal of Economic Perspectives, 20(1):221–234. (Cited on page 108.)

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica: Journal of the Econometric Society, pages 263–291. (Cited on pages 13, 14, 208, 214, 217, 218, and 272.)

Kahneman, D., Wakker, P. P., and Sarin, R. (1997). Back to bentham? explorations of experienced utility. The quarterly journal of economics, 112(2):375–406. (Cited on page 108.)

Kaiser, F. G., Hübner, G., and Bogner, F. X. (2005). Contrasting the theory of planned behavior with the value-belief-norm model in explaining conservation behavior1. Journal of applied social psychology, 35(10):2150–2170. (Cited on page 296.)

Kamargianni, M., Ben-Akiva, M., and Polydoropoulou, A. (2014). Incorporating social interaction into hybrid choice models. Transportation, 41(6):1263–1285. (Cited on pages 126, 137, 138, and 143.)

Kamargianni, M., Dubey, S., Polydoropoulou, A., and Bhat, C. (2015). Investigating the subjective and objective factors influencing teenagers' school travel mode choice–an integrated choice and latent variable model. Transportation Research Part A: Policy and Practice, 78:473–488. (Cited on pages 108, 116, 125, 129, 130, 136, 138, 140, and 141.)

Kaufmann, V. (2002). Temps et pratiques modales. le plus court est-il le mieux? Recherche-Transports-Sécurité, 75:131–143. (Cited on pages 4, 24, 144, and 181.)

Kaufmann, V., Tabaka, K., Jean-Marie, G., and Louvet, N. (2010). Et si les français n" avaient plus seulement une voiture dans la tête? Technical report, Certu. (Cited on page 126.)

Kemel, E. and Paraschiv, C. (2013). Prospect theory for joint time and money consequences in risk and ambiguity. Transportation Research Part B: Methodological, 56:81–95. (Cited on page 215.)

Kerr, G. N. and Sharp, B. M. (2009). Efficient design for willingness to pay in choice experiments:

evidence from the field. (Cited on page 257.)
Khademi, E. and Timmermans, H. (2014). The long-term effectiveness of a reward scheme in changing daily travel choices. Procedia-Social and Behavioral Sciences, 111:380–389. (Cited on page 119.)
Knetsch, J. L. (2010). Values of gains and losses: reference states and choice of measure. Environmental and Resource Economics, 46(2):179–188. (Cited on page 272.)
König, A. and Axhausen, K. W. (2002). The reliability of the transportation system and its influence on the choice behaviour. In Swiss Transport Research Conference. (Cited on page 214.)
Koppelman, F. S. and Hauser, J. R. (1978). Destination choice behavior for non-grocery-shopping trips. Transportation Research Record, (673). (Cited on page 19.)
Koppelman, F. S. and Pas, E. I. (1980). Travel-choice behavior: Models of perceptions, feelings, preference, and choice. Transportation Research Record, (765). (Cited on page 19.)
Köszegi, B. and Rabin, M. (2006). A Model of Reference-dependent Preferences. The Quarterly Journal of Economics, 121(4):1133–1165. (Cited on pages 15, 272, 273, 275, and 292.)
Kottenhoff, K. and Lindh, C. (1995). The value and effects of introducing high standard train and bus concepts in blekinge, sweden. Transport policy, 2(4):235–241. (Cited on page 21.)
Kouwenhoven, M., de Jong, G. C., Koster, P., van den Berg, V. A., Verhoef, E. T., Bates, J., and Warffemius, P. M. (2014). New values of time and reliability in passenger transport in the netherlands. Research in Transportation Economics, 47:37–49. (Cited on page 66.)
Krinsky, I. and Robb, A. L. (1986). On approximating the statistical properties of elasticities. The Review of Economics and Statistics, pages 715–719. (Cited on page 254.)
Kroes, E., Kouwenhoven, M., Debrincat, L., and Pauget, N. (2013). On the value of crowding in public transport for Île-de-france. International Transport Forum Discussion Paper 2013-18, Paris. (Cited on pages 148 and 160.)
Kroesen, M., Handy, S., and Chorus, C. (2017). Do attitudes cause behavior or vice versa? an alternative conceptualization of the attitude-behavior relationship in travel behavior modeling. Transportation Research Part A: Policy and Practice, 101:190–202. (Cited on page 118.)
La Paix Puello, L. and Geurs, K. (2015). Modelling observed and unobserved factors in cycling to railway stations: application to transit-oriented-developments in the netherlands. European journal of transport and infrastructure research, 15(1):27–50. (Cited on pages 124, 127, 138, 142, and 144.)
Lancaster, K. J. (1966). A new approach to consumer theory. Journal of political economy, 74(2):132–157. (Cited on pages 16 and 36.)
Laurent, R.-A. (2007). Le modèle «élimination par attributs» de tversky. Revue économique, 58(3):545–554. (Cited on page 15.)
Le Borgne, G., Sirieix, L., and Costa, S. (2015). La sensibilité du consommateur au gaspillage alimentaire : proposition d'une échelle de mesure. Technical report, Montpellier Supagro. (Cited on page 179.)
Li, Y.-w. (2003). Evaluating the urban commute experience: A time perception approach. Journal of Public Transportation, 6(4):3. (Cited on page 149.)
Li, Z. and Hensher, D. (2011). Prospect theoretic contributions in understanding traveller behaviour: a review and some comments. Transport reviews, 31(1):97–115. (Cited on pages 5, 13, 214, and 226.)
Li, Z., Hensher, D. A., and Rose, J. M. (2010). Willingness to pay for travel time reliability in passenger transport: a review and some new empirical evidence. Transportation research part E: logistics and transportation review, 46(3):384–403. (Cited on pages 5 and 65.)
Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 22(140):1–55. (Cited on page 126.)
Lindhjem, H. and Navrud, S. (2011). Using internet in stated preference surveys: a review and comparison of survey modes. International Review of Environmental and Resource Economics (Forthcoming). (Cited on pages 38, 39, and 42.)
Little, T. D., Card, N. A., Bovaird, J. A., Preacher, K. J., and Crandall, C. S. (2007). Structural equation modeling of mediation and moderation with contextual factors, chapter 11, pages 207–230. Guilford Press. (Cited on page 131.)
Liu, X., Vedlitz, A., and Shi, L. (2014). Examining the determinants of public environmental concern: Evidence from national public surveys. Environmental Science & Policy, 39:77–94. (Cited on pages 180 and 200.)

Lois, D. and López-Sáez, M. (2009). The relationship between instrumental, symbolic and affective factors as predictors of car use: A structural equation modeling approach. Transportation Research Part A: Policy and Practice, 43(9-10):790–799. (Cited on pages 112, 180, and 186.)
Loomis, J. (2011). What's to know about hypothetical bias in stated preference valuation studies? Journal of Economic Surveys, 25(2):363–370. (Cited on page 41.)
Louviere, J. J. (1988). Analyzing decision making: Metric conjoint analysis. Number 67. Sage. (Cited on page 43.)
Louviere, J. J., Flynn, T. N., and Carson, R. T. (2010). Discrete choice experiments are not conjoint analysis. Journal of Choice Modelling, 3(3):57–72. (Cited on pages 37, 42, 43, 49, and 52.)
Louviere, J. J., Hensher, D. A., and Swait, J. D. (2000). Stated choice methods: analysis and applications. Cambridge University Press. (Cited on pages 35, 55, 103, 251, and 289.)
Louviere, J. J., Islam, T., Wasi, N., Street, D., and Burgess, L. (2008). Designing discrete choice experiments: do optimal designs come at a price? Journal of Consumer Research, 35(2):360–375. (Cited on page 50.)
Łowicki, D. and Piotrowska, S. (2015). Monetary valuation of road noise. residential property prices as an indicator of the acoustic climate quality. Ecological Indicators, 52:472–479. (Cited on page 3.)
Luce, R. D. (1959). Individual Choice Behavior: A Theoretical Analysis. Wiley, New York. (Cited on page 18.)
Lyons, G., Jain, J., and Holley, D. (2007). The use of travel time by rail passengers in great britain. Transportation Research Part A: Policy and Practice, 41(1):107–120. (Cited on page 119.)
Lyons, G., Jain, J., Susilo, Y., and Atkins, S. (2013). Comparing rail passengers' travel time use in great britain between 2004 and 2010. Mobilities, 8(4):560—-579. (Cited on pages 112, 149, 150, and 181.)
Lyons, G. and Urry, J. (2005). Travel time use in the information age. Transportation Research Part A: Policy and Practice, 39(2):257–276. (Cited on page 149.)
Mabit, S. L., Cherchi, E., Jensen, A. F., and Jordal-Jørgensen, J. (2015). The effect of attitudes on reference-dependent preferences: Estimation and validation for the case of alternative-fuel vehicles. Transportation Research Part A: Policy and Practice, 82:17–28. (Cited on page 209.)
Mabit, S. L. and Fosgerau, M. (2011). Demand for alternative-fuel vehicles when registration taxes are high. Transportation Research Part D: Transport and Environment, 16(3):225–231. (Cited on page 209.)
Mackie, P., Wardman, M., Fowkes, A., Whelan, G., Nellthorp, J., and Bates, J. (2003). Values of travel time savings uk. Technical report. (Cited on page 174.)
Mackie, P. and Worsley, T. (2013). International comparisons of transport appraisal practice: overview report. Technical report. (Cited on page 207.)
Maldonado-Hinarejos, R., Sivakumar, A., and Polak, J. W. (2014). Exploring the role of individual attitudes and perceptions in predicting the demand for cycling: a hybrid choice modelling approach. Transportation, 41(6):1287–1304. (Cited on pages 116, 129, 132, and 157.)
Malhotra, N. (2009). Order effects in complex and simple tasks. Public Opinion Quarterly, 73(1):180–198. (Cited on page 42.)
Malokin, A., Circella, G., and Mokhtarian, P. (2017). Do multitasking millennials value travel time differently? a revealed preference study of northern california commuters. In Transportation Research Board Annual Meeting. (Cited on pages 119 and 181.)
Malokin, A., Circella, G., and Mokhtarian, P. L. (2015). How do activities conducted while commuting influence mode choice? testing public transportation advantage and autonomous vehicle scenarios. In Transportation Research Board 94th Annual Meeting, number 15-1179. (Cited on page 172.)
Maness, M., Cirillo, C., and Dugundji, E. R. (2015). Generalized behavioral framework for choice models of social influence: Behavioral and data concerns in travel behavior. Journal of transport geography, 46:137–150. (Cited on page 108.)
Manski, C. F. (1977). The structure of random utility models. Theory and decision, 8(3):229–254. (Cited on page 18.)
Mariel, P., Meyerhoff, J., and Hess, S. (2015). Heterogeneous preferences toward landscape externalities of wind turbines–combining choices and attitudes in a hybrid model. Renewable and Sustainable Energy Reviews, 41:647–657. (Cited on page 111.)
Marquart-Pyatt, S. T. (2012). Contextual influences on environmental concerns cross-nationally: A

multilevel investigation. Social science research, 41(5):1085–1099. (Cited on page 180.)
Márquez, L., Cantillo, V., and Arellana, J. (2014). How are comfort and safety perceived by inland waterway transport passengers? Transport policy, 36:46–52. (Cited on page 111.)
Marschak, J. et al. (1960). Binary choice constraints on random utility indicators. Technical report, Symposium on Mathematical Methods in the Social Sciences, K. Arrow, Ed., Stanford University Press, Stanford. (Cited on page 18.)
Masters, G. N. (1982). A rasch model for partial credit scoring. Psychometrika, 47(2):149–174. (Cited on page 239.)
Maydeu-Olivares, A. (2001). Limited information estimation and testing of thurstonian models for paired comparison data under multiple judgment sampling. Psychometrika, 66(2):209–227. (Cited on page 191.)
Mazar, N., Koszegi, B., and Ariely, D. (2014). True context-dependent preferences? the causes of market-dependent valuations. Journal of Behavioral Decision Making, 27(3):200–208. (Cited on page 214.)
Mc Fadden, D. (1986). The choice theory approach to market research. Marketing Science (1986-1998), 5(4):275. (Cited on pages 19 and 109.)
McCullagh, P. (1980). Regression models for ordinal data. Journal of the royal statistical society. Series B (Methodological), pages 109–142. (Cited on page 238.)
McCullagh, P. and Nelder, J. A. (1989). Generalized linear models, no. 37 in monograph on statistics and applied probability. (Cited on page 361.)
McFadden, D. (1973). The measurement of urban travel demand. Journal of public economics, 3(4):303–328. (Cited on pages 18, 19, and 241.)
McFadden, D. (1978). Modeling the choice of residential location. Transportation Research Record, (673). (Cited on page 238.)
McFadden, D. (1981). Econometric models of probabilistic choice. Structural analysis of discrete data with econometric applications, 198272. (Cited on page 16.)
McFadden, D. (2007). The behavioral science of transportation. Transport policy, 14(4):269–274. (Cited on page 7.)
McFadden, D., Train, K., et al. (2000). Mixed mnl models for discrete response. Journal of applied Econometrics, 15(5):447–470. (Cited on pages 18 and 20.)
McKelvey, R. D. and Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. Journal of mathematical sociology, 4(1):103–120. (Cited on page 239.)
Metcalfe, R. and Dolan, P. (2012). Behavioural economics and its implications for transport. Journal of transport geography, 24:503–511. (Cited on page 12.)
Mezghani, M. (2001). Better mobility in urban areas. Technical report, Union Internationale des Transports Publics. (Cited on page 3.)
Michea, A. and Polak, J. (2006). Modelling risky choice behaviour: evaluating alternatives to expected utility theory. In 11th International Conference on Travel Behaviour Research, Kyoto. (Cited on page 14.)
Millock, K. and Nauges, C. (2010). Household adoption of water-efficient equipment: the role of socio-economic factors, environmental attitudes and policy. Environmental and Resource Economics, 46(4):539–565. (Cited on pages 111 and 176.)
Mokhtarian, P. L., Papon, F., Goulard, M., and Diana, M. (2015). What makes travel pleasant and/or tiring? an investigation based on the french national travel survey. Transportation, 42(6):1103–1128. (Cited on page 149.)
Mokhtarian, P. L. and Salomon, I. (2001). How derived is the demand for travel? some conceptual and measurement considerations. Transportation Research Part A: Policy and Practice, 35(8):695–719. (Cited on pages 23, 112, 127, 149, 176, and 180.)
Molin, E., Mokhtarian, P., and Kroesen, M. (2016). Multimodal travel groups and attitudes: A latent class cluster analysis of dutch travelers. Transportation Research Part A: Policy and Practice, 83:14–29. (Cited on pages 109 and 177.)
Morikawa, T., Ben-Akiva, M., and McFadden, D. (1996). Incorporating psychometric data in econometric choice models. Technical report, Massachusetts Institute of technology. (Cited on pages 154, 186, and 187.)

Morikawa, T., Ben-Akiva, M., and McFadden, D. (2002). Discrete choice models incorporating revealed preferences and psychometric data. In Advances in Econometrics, pages 29–55. Emerald Group Publishing Limited. (Cited on page 110.)

Morikawa, T. and Sasaki, K. (1998). Discrete choice models with latent variables using subjective data. (Cited on pages 126, 141, 348, 349, and 350.)

Morris, E. A. (2015). Should we all just stay home? travel, out-of-home activities, and life satisfaction. Transportation Research Part A: Policy and Practice, 78:519–536. (Cited on page 177.)

Morris, E. A. and Guerra, E. (2015). Mood and mode: does how we travel affect how we feel? Transportation, 42(1):25–43. (Cited on pages 150 and 182.)

Munro, A. and Sugden, R. (2003). On the theory of reference-dependent preferences. Journal of Economic Behavior & Organization, 50(4):407–428. (Cited on page 292.)

Muthén, B. O. (1993). Goodness of fit with categorical and other nonnormal variables. SAGE Focus Editions, 154:205–205. (Cited on page 191.)

Nelder, J. and Wedderburn, R. (1972). Generalized linear models. Journal of the Royal Statistical Society. Series A (General), pages 370–384. (Cited on page 360.)

Niininen, O., Szivas, E., and Riley, M. (2004). Destination loyalty and repeat behaviour: an application of optimum stimulation measurement. International Journal of Tourism Research, 6(6):439–447. (Cited on page 127.)

Nilsson, M. and Küller, R. (2000). Travel behaviour and environmental concern. Transportation Research Part D: Transport and Environment, 5(3):211–234. (Cited on pages 176 and 177.)

Nordlund, A. M. and Garvill, J. (2003). Effects of values, problem awareness, and personal norm on willingness to reduce personal car use. Journal of environmental psychology, 23(4):339–347. (Cited on page 135.)

O'Brien, R. M. (1994). Identification of simple measurement models with multiple latent variables and correlated errors. Sociological methodology, 24:137–170. (Cited on page 132.)

O'Donoghue, T. and Rabin, M. (2006). Optimal sin taxes. Journal of Public Economics, 90(10):1825–1849. (Cited on page 3.)

OECD (2014). Greening household behaviour: Overview from the 2011 survey - revised edition. Technical report, OECD studies on environmental policy and household behaviour, OECD Publishing. (Cited on page 176.)

Ortuzar, J. d. D., Willumsen, L. G., et al. (2011). Modelling transport. Wiley New Jersey, 4th edition edition. (Cited on pages 35 and 103.)

Ory, D. T. and Mokhtarian, P. L. (2009). Modeling the structural relationships among short-distance travel amounts, perceptions, affections, and desires. Transportation Research Part A: Policy and Practice, 43(1):26–43. (Cited on page 187.)

Osgood, C. E. and Luria, Z. (1954). A blind analysis of a case of multiple personality using the semantic differential. The Journal of Abnormal and Social Psychology, 49(4p1):579. (Cited on page 126.)

Oskamp, S. (1977). Attitudes and Opinions. Englewood Cliffs N J: Prentice-Hall. (Cited on pages 137 and 295.)

Palmier, P. (2010). Mode d'emploi - musliw - logiciel de calcul d'accessibilité multimodal. Technical report, CETE Nord-Picardie. (Cited on page 223.)

Pan, X., Rasouli, S., and Timmermans, H. (2016). Modeling bounded rationality in choice behavior: relative utility vs. random regret models. In 13th International Conference on Design and Decision Support Systems in Architecture and Urban Planning. (Cited on pages 15 and 209.)

Pascal, M., de Crouy Chanel, P., Corso, M., Medina, S., Wagner, V., Goria, S., and al. (2016). Impact de l'exposition chronique aux particules fines sur la mortalité en france continentale et analyse des gains en santé de plusiplus scénarios de réduction de la pollution atmosphérique. Technical report, Santé publique France. (Cited on page 3.)

Paulssen, M., Temme, D., Vij, A., and Walker, J. L. (2014). Values, attitudes and travel behavior: a hierarchical latent variable mixed logit model of travel mode choice. Transportation, 41(4):873–888. (Cited on pages 111, 114, 116, 127, 129, 130, 132, 134, 135, 136, 139, 141, 145, 348, 349, and 350.)

Pearce, D., Atkinson, G., and Mourato, S. (2006). Cost-benefit analysis and the environment: recent developments. Organisation for Economic Co-operation and development. (Cited on pages 36

and 37.)
Pearmain, D., J., S., Kroes, E. P., and M., B. (1991). Stated preference techniques: a guide to practice, 2nd edition. Steer Davies Gleave and Hague Consulting Group. (Cited on pages 35, 44, 46, and 50.)
Peyhardi, J. (2013). A new GLM framework for analysing categorical data; application to plant structure and development. PhD thesis. (Cited on pages 254, 362, and 363.)
Peyhardi, J., Trottier, C., and Guédon, Y. (2015). A new specification of generalized linear models for categorical responses. Biometrika, page asv042. (Cited on pages 5, 25, 205, 233, 235, 239, 241, 242, and 370.)
Peyhardi, J., Trottier, C., and Guédon, Y. (2016). Partitioned conditional generalized linear models for categorical responses. Statistical Modelling, 16(4):297–321. (Cited on page 271.)
Picard, N., de Palma, A., and Dantan, S. (2013). Intra-household Discrete Choice Models of Mode Choice and Residential Location. International Journal of Transport Economics, 40(3):419–446. (Cited on page 9.)
Politis, I., Papaioannou, P., and Basbas, S. (2012). Integrated choice and latent variable models for evaluating flexible transport mode choice. Research in Transportation Business & Management, 3:24–38. (Cited on pages 127, 129, 130, 133, 140, 142, and 145.)
Pollak, R. A. (1976). Interdependent preferences. The American Economic Review, 66(3):309–320. (Cited on page 9.)
Polydoropoulou, A., Kamargianni, M., and Tsirimpa, A. (2014). Car use addiction vs. ecological consciousness: Which one prevails on mode choice behavior. Lulu Publishers, USA. (Cited on pages 136, 139, and 145.)
Pons, D. (2011). Mise en place d'enquêtes par préférences déclarées dans le cadre de projets d'étude relatifs au secteur des transports de personnes. PhD thesis, Université Lumière-Lyon II. (Cited on pages 39, 50, 54, 55, 62, 63, 64, 67, 69, 70, and 99.)
Prashker, J. N. (1979). Mode choice models with perceived reliability measures. Journal of Transportation Engineering, 105(3). (Cited on page 19.)
Prelec, D. (1998). The probability weighting function. Econometrica, pages 497–527. (Cited on page 220.)
Prochaska, J. O. and DiClemente, C. C. (1986). Toward a comprehensive model of change. In Treating addictive behaviors, pages 3–27. Springer. (Cited on page 291.)
Quiggin, J. (1982). A theory of anticipated utility. Journal of Economic Behavior & Organization, 3(4):323–343. (Cited on pages 13, 205, 214, and 218.)
Quinet, E. et al. (2014). L'évaluation socioéconomique des investissements publics. Technical report, HAL. (Cited on pages 22, 97, 168, 174, 230, 352, and 372.)
Rabin, M. (2002). A perspective on psychology and economics. European economic review, 46(4):657–685. (Cited on pages 12 and 13.)
Ramos, A., Labandeira, X., and Löschel, A. (2016). Pro-environmental households and energy efficiency in spain. Environmental and Resource Economics, 63(2):367–393. (Cited on pages 111 and 176.)
Rasouli, S. and Timmermans, H. (2014). Applications of theories and models of choice and decision-making under conditions of uncertainty in travel behavior research. Travel Behaviour and Society, 1(3):79–90. (Cited on pages 214 and 220.)
Rasouli, S. and Timmermans, H. (2015). Bounded Rational Choice Behaviour: Applications in Transport. Emerald Group Publishing Limited. (Cited on pages 4 and 12.)
Raveau, S., Álvarez-Daziano, R., Yáñez, M., Bolduc, D., and Ortúzar, J. d. D. (2010). Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. Transportation Research Record: Journal of the Transportation Research Board, (2156):131–139. (Cited on pages 116, 124, 136, 137, 138, 141, 142, 143, and 349.)
Raveau, S., Yáñez, M. F., and Ortúzar, J. d. D. (2012). Practical and empirical identifiability of hybrid discrete choice models. Transportation Research Part B: Methodological, 46(10):1374–1383. (Cited on pages 110, 124, 129, 132, 347, and 349.)
Razo, M. and Gao, S. (2013). A rank-dependent expected utility model for strategic route choice with stated preference data. Transportation Research Part C: Emerging Technologies, 27:117–130. (Cited on pages 214, 217, and 229.)
RFF (2013). Etude sur la perception et la valorisation de la régularité, du point de vue de l'usager.

Technical report, Réseau Ferré de France, étude réalisée par STRATEC. (Cited on pages 47, 50, 62, 65, 67, 68, 69, 70, 100, 261, and 264.)

Région Rhône-Alpes EDR (2016). Enquête Déplacements Régionale - les premiers résultats – vagues 1/2/3. Technical report, Cerema (Hurez, Christophe et Tebar, Maria) pour la Région Rhône-Alpes. (Cited on pages 30 and 31.)

Rhemtulla, M., Brosseau-Liard, P. E., and Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. Psychological Methods, 17(3):354–373. (Cited on page 191.)

Richter, C. and Keuchel, S. (2012). Modelling mode choice in passenger transport with integrated hierarchical information integration. Journal of Choice Modelling, 5(1):1–21. (Cited on pages 4, 21, 47, 62, 100, and 264.)

Rieser-Schüssler, N. and Axhausen, K. (2012). Investigating the influence of environmentalism and variety seeking on mode choice. Transportation Research Record: Journal of the Transportation Research Board, (2322):31–41. (Cited on pages 127, 138, 139, 143, and 348.)

Roberts, C. (2007). Mixing modes of data collection in surveys: A methodological review. Technical Report NCRM 008, ESRC National Centre for Research Methods. (Cited on page 41.)

Roberts, J., Popli, G., and Harris, R. J. (2014). Do environmental concerns affect commuting choices? hybrid choice modelling with household survey data. The Sheffield Economic Research Paper Series (SERPS). (Cited on pages 124, 125, 127, 129, 130, 133, 134, and 135.)

Rojas-Rueda, D., De Nazelle, A., Andersen, Z. J., Braun-Fahrländer, C., Bruha, J., Bruhova-Foltynova, H., Desqueyroux, H., Praznoczy, C., Ragettli, M. S., Tainio, M., et al. (2016). Health impacts of active transportation in europe. PloS one, 11(3):e0149990. (Cited on page 3.)

Rose, J. and Bliemer, M. C. (2013). Advances in the construction of efficient stated choice experimental designs. (Cited on page 54.)

Rose, J., Collins, A., Bliemer, M., and Hensher, D. (2009). Ngene 1.0 stated choice experiment design software. Technical report, University of Sydney. (Cited on pages 55, 56, and 57.)

Rose, J. M. and Bliemer, M. C. (2007). Stated preference experimental design strategies. In Handbook of Transport Modelling: 2nd Edition, pages 151–180. Emerald Group Publishing Limited. (Cited on pages 224 and 252.)

Rose, J. M., Bliemer, M. C., Hensher, D. A., and Collins, A. T. (2008). Designing efficient stated choice experiments in the presence of reference alternatives. Transportation Research Part B: Methodological, 42(4):395–406. (Cited on pages 55, 58, and 153.)

Rosseel, Y. (2012). lavaan: An r package for structural equation modeling. Journal of Statistical Software, 48(2):1–36. (Cited on page 191.)

Sallis, J. F. (2016). New evidence for the role of transportation in health. The Lancet Public Health, 1(2):e38–e39. (Cited on page 3.)

Sallis, J. F., Frank, L. D., Saelens, B. E., and Kraft, M. K. (2004). Active transportation and physical activity: opportunities for collaboration on transportation and public health research. Transportation Research Part A: Policy and Practice, 38(4):249–268. (Cited on page 3.)

Samuelson, W. and Zeckhauser, R. (1988). Status quo bias in decision making. Journal of risk and uncertainty, 1(1):7–59. (Cited on page 15.)

Savage, B., Knight, T., Bacon, J., Millington, A., Bullock, H., and Buckland, J. (2011). Behavioural insights toolkit. Technical report, Social Research and Evaluation Division, UK Department for Transport, London, UK. (Cited on pages 154 and 178.)

Scagnolarie, S., Walker, J., and Maggi, R. (2015). Young drivers' night-time mobility preferences and attitude toward alcohol consumption: A hybrid choice model. Accident Analysis & Prevention, 83:74–89. (Cited on pages 116, 125, 126, 127, 129, 130, 132, 133, 348, and 349.)

Scarpa, R. and Rose, J. M. (2008). Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why*. Australian journal of agricultural and resource economics, 52(3):253–282. (Cited on pages 54, 56, and 58.)

Schlenker, B. (1980). Impression management: The self-concept, social identity, and interpersonal relations. Brooks/Cole, Monterey, CA. (Cited on pages 12 and 180.)

Schleyer-Lindenmann, A., Dauvier, B., Ittner, H., and Piolat, M. (2014). Mesure des attitudes

environnementales : analyse structurale d'une version française de la NEPS (Dunlap et al., 2000). Psychologie Française. (Cited on pages 179 and 185.)
Schmeidler, D. (1989). Subjective probability and expected utility without additivity. Econometrica: Journal of the Econometric Society, pages 571–587. (Cited on page 207.)
Schwartz, S. H. (1977). Normative influences on altruism. Advances in experimental social psychology, 10:221–279. (Cited on pages 109, 135, 178, and 296.)
Schwartz, S. H., Melech, G., Lehmann, A., Burgess, S., Harris, M., and Owens, V. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. Journal of cross-cultural psychology, 32(5):519–542. (Cited on page 127.)
Senbil, M. and Kitamura, R. (2005). Heterogeneity in commuter departure time decision: A prospect theoretic approach. In Applied Research in Uncertainty Modeling and Analysis, pages 369–398. Springer. (Cited on page 227.)
Sheffi, Y. (1979). Estimating choice probabilities among nested alternatives. Transportation Research Part B: Methodological, 13(3):189–205. (Cited on page 240.)
Shires, J. D. and De Jong, G. C. (2009). An international meta-analysis of values of travel time savings. Evaluation and program planning, 32(4):315–325. (Cited on pages 148, 230, 352, and 353.)
Slovic, P. (1987). Perception of risk. 236:280–285. (Cited on page 207.)
Small, K. A. (1987). A discrete choice model for ordered alternatives. Econometrica: Journal of the Econometric Society, pages 409–424. (Cited on page 239.)
Small, K. A. (2012). Valuation of travel time. Economics of Transportation, 1(1):2 – 14. (Cited on page 147.)
Small, K. A. and Rosen, H. S. (1981). Applied welfare economics with discrete choice models. Econometrica: Journal of the Econometric Society, pages 105–130. (Cited on pages 100, 101, and 239.)
Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. Sociological methodology, 13:290–312. (Cited on page 188.)
SOeS (2017). Chiffres clés du transport, édition 2017. Technical report, Service de l'observation et des statistiques. (Cited on page 3.)
Sottile, E., Cherchi, E., and Meloni, I. (2014). Measuring soft measures within a stated preference survey: the effect of pollution and traffic stress on mode choice. volume 11, pages 434–451. Elsevier. (Cited on pages 116, 126, 127, 129, 135, 136, 138, 139, and 140.)
Sottile, E., Meloni, I., and Cherchi, E. (2013). A hybrid discrete choice model to assess the effect of awareness and attitude towards environmentally friendly travel modes. volume 5, pages 44–55. Elsevier. (Cited on pages 116, 125, 130, 140, and 150.)
Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. Journal of economic literature, pages 332–382. (Cited on page 208.)
Stathopoulos, A. and Hess, S. (2012). Revisiting reference point formation, gains–losses asymmetry and non-linear sensitivities with an emphasis on attribute specific treatment. Transportation research part A: policy and practice, 46(10):1673–1689. (Cited on page 209.)
Steg, L. (2003). Can public transport compete with the private car? IATSS Research, 27(2):27–35. (Cited on pages 181 and 185.)
Steg, L. (2005). Car use: lust and must. Instrumental, symbolic and affective motives for car use. Transportation Research Part A: Policy and Practice, 39(2-3):147–162. (Cited on pages 12, 109, 112, 178, 180, 181, 186, and 195.)
Steg, L., Vlek, C., and Slotegraaf, G. (2001). Instrumental-reasoned and symbolic-affective motives for using a motor car. Transportation Research Part F: Traffic Psychology and Behaviour, 4(3):151–169. (Cited on pages 12 and 186.)
Stern, P. C. (1992). Psychological dimensions of global environmental change. Annual Review of Psychology, 43(1):269–302. (Cited on page 178.)
Stern, P. C., Dietz, T., Abel, T. D., Guagnano, G. A., and Kalof, L. (1999). A value-belief-norm theory of support for social movements: The case of environmentalism. Human Ecology Review, 6(2):81. (Cited on pages 11, 109, 135, 178, and 296.)
Sun, Z., Arentze, T., and Timmermans, H. (2012). A heterogeneous latent class model of activity rescheduling, route choice and information acquisition decisions under multiple uncertain events.

Transportation research part C: emerging technologies, 25:46–60. (Cited on page 217.)
Sunstein, C. R. and Reisch, L. A. (2013). Automatically green: Behavioral economics and environmental protection. Havard Environmental Law Review, 38:127. (Cited on page 287.)
Swait, J., Adamowicz, W., Hanemann, M., Diederich, A., Krosnick, J., Layton, D., Provencher, W., Schkade, D., and Tourangeau, R. (2002). Context dependence and aggregation in disaggregate choice analysis. Marketing Letters, 13(3):195–205. (Cited on page 272.)
Szczepańska, A., Senetra, A., and Wasilewicz-Pszczółkowska, M. (2015). The effect of road traffic noise on the prices of residential property–a case study of the polish city of olsztyn. Transportation Research Part D: Transport and Environment, 36:167–177. (Cited on page 3.)
Temme, D., Paulssen, M., Dannewald, T., et al. (2007). Integrating latent variables in discrete choice models–how higher-order values and attitudes determine consumer choice. Berlin, Germany: Humboldt University. (Cited on pages 108, 126, 137, 138, 141, 143, 149, and 350.)
Thisse, J.-F. and Billot, A. (1995). Modèles de choix individuels discrets: théorie et applications à la micro-économie. Revue économique, pages 921–931. (Cited on pages 15 and 238.)
Thorhauge, M., Haustein, S., and Cherchi, E. (2016). Accounting for the theory of planned behaviour in departure time choice. Transportation Research Part F: Traffic Psychology and Behaviour, 38:94–105. (Cited on pages 111, 134, and 150.)
Thurstone, L. L. (1927). A law of comparative judgment. Psychological review, 34(4):273. (Cited on pages 18 and 126.)
Tillema, T., Ben-Elia, E., Ettema, D., and van Delden, J. (2013). Charging versus rewarding: A comparison of road-pricing and rewarding peak avoidance in the netherlands. Transport Policy, 26:4–14. (Cited on page 119.)
Timmermans, H. (2010). On the (ir) relevance of prospect theory in modelling uncertainty in travel decisions. EJTIR, 4(10):368–384. (Cited on page 215.)
Timmermans, H., Heuden, R., and Westerveld, H. (1982). The identification of factors influencing destination choice: an application of the repertory grid methodology. Transportation, 11(2):189–203. (Cited on pages 43 and 52.)
Tinsley, H. E. and Brown, S. D. (2000). Handbook of applied multivariate statistics and mathematical modeling. Academic Press. (Cited on page 126.)
Toepoel, V., Vis, C., Das, M., and Van Soest, A. (2008). Design of web questionnaires: An information-processing perspective for the effect of response categories. Sociological Methods & Research. (Cited on page 42.)
Train, K. E. (2009). Discrete choice methods with simulation. Cambridge university press. (Cited on pages 19, 132, 154, 226, and 365.)
Triandis, H. C. (1977). Interpersonal behavior. Brooks/Cole Pub. Co. (Cited on pages 109, 135, 179, and 296.)
Tsirimpa, A. and Polydoropoulou, A. (2015). The impact of traffic information acquisition on the traffic conditions of the athens greater area. Transportation Systems and Engineering: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications, page 174. (Cited on page 111.)
Tudela, A., Habib, K., Carrasco, J., and Osman, A. (2011). Incorporating the explicit role of psychological factors on mode choice: A hybrid mode choice model by using data from an innovative psychometric survey. In 2nd International Choice Modelling Conference, Leeds. (Cited on pages 125, 126, 127, 129, 134, 135, and 143.)
Tutz, G. (1991). Sequential models in categorical regression. Computational Statistics & Data Analysis, 11(3):275–295. (Cited on pages 238 and 243.)
Tutz, G. (2012). Regression for Categorical Data. Cambridge University Press, New York. (Cited on pages 237, 241, and 245.)
Tversky, A. (1972a). Choice by elimination. Journal of mathematical psychology, 9(4):341–367. (Cited on page 15.)
Tversky, A. (1972b). Elimination by aspects: A theory of choice. Psychological review, 79(4):281. (Cited on page 15.)
Tversky, A. and Fox, C. R. (1995). Weighing risk and uncertainty. Psychological review, 102(2):269.

(Cited on page 215.)
Tversky, A. and Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. The quarterly journal of economics, 106(4):1039–1061. (Cited on pages 6, 14, 15, 26, 205, 208, 272, 273, 276, 281, and 286.)
Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and uncertainty, 5(4):297–323. (Cited on pages 13, 14, 25, 67, 211, 214, 218, 219, 220, 229, 230, 232, and 272.)
Tversky, A. and Simonson, I. (1993). Context-dependent preferences. Management science, 39(10):1179–1189. (Cited on pages 14, 208, 214, and 272.)
Verplanken, B., Aarts, H., van Knippenberg, A., and van Knippenberg, C. (1994). Attitude versus general habit: Antecedents of travel mode choice. Journal of Applied Social Psychology, 24(4):285 – 300. (Cited on pages 11, 126, and 186.)
Vij, A. and Walker, J. (2014). Hybrid choice models: The identification problem. Edward Elgar Publishing Limited. (Cited on pages 116, 122, 132, 348, and 349.)
Vij, A. and Walker, J. L. (2016). How, when and why integrated choice and latent variable models are latently useful. Transportation Research Part B: Methodological, 90:192–217. (Cited on pages 116, 122, 139, 143, 145, and 146.)
Vincent, S. (2008). Les «altermobilités»: analyse sociologique d'usages de déplacements alternatifs à la voiture individuelle. Des pratiques en émergence? PhD thesis, Université René Descartes-Paris V. (Cited on pages 12 and 296.)
Viney, R., Savage, E., and Louviere, J. (2005). Empirical investigation of experimental design properties of discrete choice experiments in health care. Health Economics, 14(4):349–362. (Cited on page 257.)
Von Neumann, J. and Morgenstern, O. (1944). (Cited on page 206.)
Wakker, P. (1994). Separating marginal utility and probabilistic risk aversion. Theory and decision, 36(1):1–44. (Cited on page 207.)
Walker, J. and Ben-Akiva, M. (2002). Generalized random utility model. Mathematical social sciences, 43(3):303–343. (Cited on page 111.)
Walker, J., Li, J., Srinivasan, S., and Bolduc, D. (2013). Travel demand models in the developing world: Correcting for measurement errors. Transportation Letters. (Cited on pages 125, 132, and 349.)
Walker, J., Wang, Y., Thorhauge, M., and Ben-Akiva, M. (2016). D-efficient or deficient? A robustness analysis of stated choice experimental designs. In Transportation Research Board 94th - Annual Meeting. (Cited on pages 53, 59, 60, and 61.)
Walker, J. L. (2001). Extended discrete choice models: integrated framework, flexible error structures, and latent variables. PhD thesis, Massachusetts Institute of Technology. (Cited on pages 110, 111, 116, 132, 154, 348, and 349.)
Wall, R., Devine-Wright, P., and Mill, G. A. (2007). Comparing and combining theories to explain proenvironmental intentions: The case of commuting-mode choice. Environment and behavior, 39(6):731–753. (Cited on page 135.)
Walton, D., Thomas, J., and Dravitzki, V. (2004). Commuters' concern for the environment and knowledge of the effects of vehicle emissions. Transportation Research Part D: Transport and Environment, 9(4):335–340. (Cited on pages 176 and 177.)
Wardman, M. (1998). The value of travel time: a review of british evidence. Journal of transport economics and policy, pages 285–316. (Cited on page 148.)
Wardman, M. (2001). A review of british evidence on time and service quality valuations. Transportation Research Part E: Logistics and Transportation Review, 37(2–3):107 – 128. Advances in the Valuation of Travel Time Savings. (Cited on pages 81 and 148.)
Wardman, M. (2014). Valuing convenience in public transport. (Cited on pages 4 and 264.)
Wardman, M., Chintakayala, P., de Jong, G., and Ferrer, D. (2012). European wide meta-analysis of values of travel time. Technical report, ITS, University of Leeds, Paper prepared for EIB. (Cited on pages 148, 168, 230, 261, 352, and 353.)
Wardman, M. and Lyons, G. (2016). The digital revolution and worthwhile use of travel time: implications for appraisal and forecasting. Transportation, 43(3):507–530. (Cited on pages 149, 181, and 352.)

Wardman, M., Shires, J., Lythgoe, W., and Tyler, J. (2004). Consumer benefits and demand impacts of regular train timetables. International Journal of Transport Management, 2(1):39–49. (Cited on pages 21, 63, 64, 99, 148, and 352.)

Wardman, M. and Wheat, P. (2013). Meta-analysis of post 1994 values of non work travel time savings. Technical report, University of Leeds. (Cited on page 261.)

Wardman, M. and Whelan, G. (2011). Twenty years of rail crowding valuation studies: evidence and lessons from british experience. Transport reviews, 31(3):379–398. (Cited on pages 170, 261, and 373.)

Watts, L. and Lyons, G. (2010). Travel remedy kit, pages 189–213. (Cited on pages 112, 113, 120, 174, and 288.)

Waygood, E. O. D. and Avineri, E. (2016). Communicating transportation carbon dioxide emissions information: Does gender impact behavioral response? Transportation Research Part D: Transport and Environment, 48:187–202. (Cited on page 180.)

West, S. G., Taylor, A., and Wu, W. (2012). Model fit and model selection in structural equation modeling. In Handbook of Structural Equation Modeling. Guilford Press. (Cited on pages 188 and 191.)

Wu, G. and Gonzalez, R. (1999). Nonlinear decision weights in choice under uncertainty. Management science, 45(1):74–85. (Cited on pages 217 and 218.)

Xiao, Y. and Fukuda, D. (2015). On the cost of misperceived travel time variability. Transportation Research Part A: Policy and Practice, 75:96–112. (Cited on pages 66 and 215.)

Yaari, M. E. (1987). The dual theory of choice under risk. Econometrica: Journal of the Econometric Society, pages 95–115. (Cited on pages 13, 214, 218, 220, and 227.)

Yanez, M. F., Raveau, S., and Ortúzar, J. d. D. (2010). Inclusion of latent variables in mixed logit models: modelling and forecasting. Transportation Research Part A: Policy and Practice, 44(9):744–753. (Cited on pages 116, 124, 125, 129, 136, 141, 142, 144, 145, 348, 349, and 350.)

Ye, R. and Titheridge, H. (2016). Satisfaction with the commute: The role of travel mode choice, built environment and attitudes. Transportation Research Part D: Transport and Environment. (Cited on page 187.)

Zailani, S., Iranmanesh, M., Masron, T. A., and Chan, T.-H. (2016). Is the intention to use public transport for different travel purposes determined by different factors? Transportation Research Part D: Transport and Environment, 49:18–24. (Cited on page 187.)

Zhang, J., Timmermans, H., Borgers, A., and Wang, D. (2004). Modeling traveler choice behavior using the concepts of relative utility and relative interest. Transportation Research Part B: Methodological, 38(3):215–234. (Cited on pages 15 and 209.)

Zikmund, W. G., Babin, B. J., Carr, J. C., and Griffin, M. (2013). Business research methods. Cengage Learning. (Cited on page 126.)

List of Figures

1 Classification des déterminants du choix modal basée sur la revue de De Witte et al. (2013). 8
2 Processus de choix . . . 10
3 Répartition modale des déplacements des rhônalpins (de 11 ans et plus) . . . 30
4 Nombre de déplacements en trains ou cars régionaux réalisés par 1 000 rhônealpins de 11 ans et plus selon la zone de résidence . . . 31

1.1 Exemple de question de choix . . . 47
1.2 Choix multimodal pour un déplacement de plus de 65 km avec le train pour mode principal 50
1.3 Sources d'efficacité de l'enquête . . . 54
1.4 Robustesse des résultats avec zoom sur des valeurs du temps proches de l'*a priori* . . . 60
1.5 Robustesse des résultats sur l'ensemble des valeurs du temps . . . 61
1.6 Scénarios de grilles horaires utilisées dans les exercices de choix . . . 64
1.7 Prise en compte de la fiabilité dans les EPD . . . 69

2.1 Explications pour l'exercice 1 . . . 75
2.2 Exemple de choix pour l'exercice 1 . . . 75
2.3 Question de choix pour l'exercice 1 . . . 76
2.4 Explications pour l'exercice 2 . . . 76
2.5 Exemple de choix pour l'exercice 2 . . . 77
2.6 Question de choix pour l'exercice 2 . . . 77
2.7 Explications pour l'exercice 3 . . . 78
2.8 Exemple de choix pour l'exercice 3 . . . 78
2.9 Question de choix pour l'exercice 3 . . . 79
2.10 Articulation entre source de l'échantillon, mode d'enquête et EDR . . . 82
2.11 Lignes ferroviaires en Région Rhône-Alpes. En rouge, les lignes enquêtées pour l'EPD. . . 84

3.1 Coût de déplacement en fonction du temps de déplacement . . . 88
3.2 Temps de déplacement en voiture en fonction du temps de déplacement en train . . . 89
3.3 Coût de déplacement en voiture en fonction du coût de déplacement en train . . . 89
3.4 Activités pratiquées pendant le déplacement par type d'usager . . . 92
3.5 Evaluation du temps de déplacement par type d'usager . . . 92

3.6 The gap between basic Discrete Choice Models (left) and the Complexity of Behavior (right) 110

4.1 Distribution of the size of samples . . . 125
4.2 Mediation and moderation structures in SEMs . . . 131
4.3 Topics related to latent variables included in ICLV mode choice models . . . 136

5.1 Example of choice question in the choice experiment . . . 153
5.2 VOT heterogeneity according to mode, comfort and latent variables . . . 170
5.3 Boxplot of VOT estimated with the "ICLV Full" Model . . . 171

6.1 Theoretical model . . . 184
6.2 Model M1 . . . 190
6.3 Model M2 . . . 190
6.4 Distribution of answers on items . . . 193

7.1 Example of choice question for Exercise 1 . . . 223
7.2 Desutility associated with travel . . . 227
7.3 Probability weighting functions . . . 228
7.4 Value of time associated with the train and car modes . . . 231

8.1 Reference models position among discrete choice models . . . 240
8.2 Normalized cdf . . . 244
8.3 Loglikelihood of reference models . . . 248
8.4 Example of choice question for Exercise 3 . . . 252
8.5 Simulated value of time for the reference models . . . 262
8.6 Simulated time equivalent of comfort for the reference models . . . 265
8.7 Own-time elasticity for reference models . . . 267
8.8 An illustration of a value function which respects the three features of the reference-dependent theory . . . 273
8.9 Density functions of distributions used in reference models . . . 276
8.10 Reference-dependent utility functions on simulated data . . . 277
8.11 Normalized reference-dependent utility functions . . . 278
8.12 Probability of choosing option 2, according to η_2 and η_3 . . . 279

B.1 Critères de différenciation entre modes à l'aide de la grille de Kelly . . . 306

List of Tables

1.1 Variantes des méthodes de modélisation des choix . . . 37
1.2 Valeurs estimées des attributs en lien avec le cadencement et la fréquence . . . 64
1.3 Valeurs estimées des attributs en lien avec la fiabilité des trains . . . 70

2.1 Variables présentées par exercice . . . 79

3.1 Nombre de personnes ayant commencé et fini chaque exercice de choix . . . 85
3.2 Comparaison des échantillons . . . 86
3.3 Statistiques descriptives sur les attributs . . . 88
3.4 Statistiques descriptives sur l'âge (en années) . . . 90
3.5 Motif à l'origine et à la destination (part en %) . . . 90
3.6 Statistiques descriptives des variables individuelles . . . 91
3.7 Part modale (en %) . . . 91
3.8 Estimation des modèles . . . 95
3.9 Consentements à payer et équivalents temps des attributs de l'EPD . . . 98
3.10 Variation compensatoire (en €) pour la fréquence de passage, le cadencement et le confort 101
3.11 Elasticités propres et élasticités croisées (temps et coût) . . . 103

4.1 Behavioural theories used in ICLV models. . . . 135
4.2 Latent variables grid with examples . . . 138

5.1 Descriptive statistics . . . 152
5.2 Estimation results of the DCM component . . . 161
5.3 Estimation results of the measurement model of the SEM component . . . 162
5.4 Estimation results of the structural model of the SEM component . . . 163
5.5 Scenarios for simulation . . . 167
5.6 Value of time and market shares according to mode and comfort . . . 169

6.1 Items measuring environmental concern . . . 185
6.2 Items measuring motives for car use . . . 186
6.3 Items measuring perceptions and feelings towards public transport . . . 187

6.4 Estimated models . . . 189
6.5 Descriptive statistics . . . 192
6.6 Frequency of use of transport modes (in %) . . . 192
6.7 Exploratory factor analysis loadings . . . 194
6.8 Test of mediation: Model M1 . . . 196
6.9 Test of mediation: Model M2 . . . 197
6.10 Estimation results . . . 199
6.11 Summary of hypotheses and results . . . 200

7.1 Descriptive statistics . . . 224
7.2 Estimation results for the seven models . . . 225
7.3 Value of time (in €/hour) . . . 230

8.1 Characteristics of the cumulative distribution functions $F \in \mathfrak{F}^0$ . . . 243
8.2 Descriptive statistics . . . 253
8.3 Goodness-of-fit indicators of (reference, F, $Z^{(i)}$) models for $F \in \mathfrak{F}^0$ with $i = 1, 2, 3$ . . . 255
8.4 Upper bound of the p-values for the Ben-Akiva and Swait test . . . 256
8.5 Parameter estimates of (reference, $F_{m,s}$, $Z^{(1)}$) models for $F_{m,s} \in \mathfrak{F}^0_{q_{0.95}}$ . . . 258
8.6 Parameter estimates of (reference, $F_{m,s}$, $Z^{(2)}$) models for $F_{m,s} \in \mathfrak{F}^0_{q_{0.95}}$ . . . 259
8.7 Parameter estimates of (reference, $F_{m,s}$, $Z^{(3)}$) models for $F_{m,s} \in \mathfrak{F}^0_{q_{0.95}}$ . . . 260
8.8 Value of time and time equivalent of comfort for reference models . . . 261
8.9 Own-time and own-cost elasticities for (reference, F, $Z^{(i)}$) models for $F \in \mathfrak{F}^0$ with $i = 2, 3$. 267
8.10 Unweighted cross-elasticities for reference models . . . 269
8.11 Weighted cross-elasticities for reference models . . . 270
8.12 Reference models compliance with the reference-dependent theory . . . 278

D.1 Variables explaining need for comfort and perception of comfort on public transport . . . 351

F.1 Estimated Likelihood and $\bar{\rho}^2$ . . . 355

G.1 Estimation results for the M3 models . . . 357
G.2 Goodness-of-fit indicators for the M3 models . . . 358

H.1 Parralèle entre approche statistique et économie des modèles de choix discret . . . 366
H.2 Exemples de fonctions de distribution . . . 367

J.1 Valeurs du temps (en €/heure) estimées à l'échelle de la Région Rhône-Alpes 375
J.2 Consentements à payer et équivalents temps du confort 376
J.3 Consentements à payer et équivalents temps du temps de retard, fréquence et cadencement 376

Essays on travel mode choice modeling

A discrete choice approach to the interactions between economic and behavioral theories

Abstract: The objective of this thesis is to incorporate aspects of psychology and behavioral economics theories in discrete choice models to promote a better understanding of mode choice at regional level. Part II examines the inclusion of latent variables to explain mode choice. A literature review of integrated choice and latent variable models – that is, models combining a structural equation model and a discrete choice model – is followed by the estimation of an integrated choice and latent variable model to show how the heterogeneity of economic outputs (here, value of time) can be explained with latent variables (here, perceived comfort in public transport) and observable variables (here, the guarantee of a seat). The simulation of scenarios shows, however, that the economic gain (decrease in value of time) is higher when policies address tangible factors than when they address latent factors. On the basis of a mediation model, the estimation of a structural equation model furthermore implies that the influence of environmental concern on mode choice habits is partially mediated by the indirect utility derived from public transport use. Part III examines two utility formulations taken from behavioral economics: 1) rank-dependent utility to model risky choices, and 2) reference-dependent utility. Firstly, a rank-dependent utility model is included in discrete choice models and, in particular, a latent-class model, in order to analyze intra- and inter-individual heterogeneity when the travel time is subject to variability. The results show that the probability of a delay is over-estimated for train travel and under-estimated for car travel, especially for car users, as train users are more likely to take into account the expected travel time. In the models that account for risk aversion, the utility functions are convex, which implies a decrease in value of time. Secondly, a new family of discrete choice models generalizing the multinomial logit model, the reference models, is estimated. On my data, these models allow for a better selection of explanatory variables than the multinomial logit model and a more robust estimation of economic outputs, particularly in cases of high unobserved heterogeneity. The economic formulation of reference models shows that the best empirical models are also more compatible with Tversky et Kahneman's reference-dependent model.

Essais sur la modélisation du choix modal

Une approche par les choix discrets des interactions entre théories économiques et comportementales

Résumé : Cette thèse a pour objectif d'incorporer des éléments de théories de psychologie et d'économie comportementale dans des modèles de choix discret afin d'améliorer la compréhension du choix modal réalisé à l'échelle régionale. Les estimations se basent sur une enquête de type *choice experiment* présentée en première partie. Une deuxième partie s'intéresse à l'incorporation de variables latentes pour expliquer le choix modal. Après une revue de littérature sur les modèles de choix hybrides, c'est-à-dire des modèles combinant modèle d'équations structurelles et modèle de choix discret, un tel modèle est estimé pour montrer comment l'hétérogénéité d'outputs économiques (ici, la valeur du temps) peut être expliquée à l'aide de variables latentes (ici, le confort perçu dans les transports en commun) et de variables observables (ici, la garantie d'une place assise). La simulation de scénarios montre cependant que le gain économique (diminution de la valeur du temps) est plus élevé lorsque les politiques agissent sur des dimensions palpables que sur des dimensions latentes. S'appuyant sur un modèle de médiation, l'estimation d'un modèle d'équations structurelles montre par ailleurs que l'effet de la conscience environnementale sur les habitudes de choix modal est partiellement médié par l'utilité indirecte retirée de l'usage des transports en commun. Une troisième partie s'intéresse à deux formalisations de l'utilité issues de l'économie comportementale : 1) l'utilité dépendante au rang en situation de risque et 2) l'utilité dépendante à la référence. Dans un premier temps, un modèle d'utilité dépendante au rang est inséré dans des modèles de choix discret et, en particulier, un modèle à classes latentes, afin d'analyser l'hétérogénéité intra- et inter-individuelle lorsque le temps de déplacement n'est pas fiable. La probabilité de survenue d'un retard est sur-évaluée pour les déplacements en train et sous-évaluée pour les déplacements en voiture, en particulier pour les automobilistes, les usagers du train prenant d'avantage en compte l'espérance du temps de déplacement. Dans les modèles prenant en compte l'aversion au risque, les fonctions d'utilité sont convexes, ce qui implique une décroissance de la valeur du temps. Dans un deuxième temps, une nouvelle famille de modèles de choix discret généralisant le modèle logit multinomial, les modèles de référence, est estimée. Sur mes données, ces modèles permettent une meilleure sélection des variables explicatives que le logit multinomial et l'estimation d'outputs économiques plus robustes, notamment en cas de forte hétérogénéité inobservée. La traduction économique des modèles de référence montre que les meilleurs modèles empiriques sont également les plus compatibles avec le modèle de dépendance à la référence de Tversky et Kahneman.