Four essays in finance and macroeconomics : the contribution of nonlinear econometrics

Quentin Lajaunie

PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT

DE L'UNIVERSITÉ PSL

Préparée à Université Paris-Dauphine

Four essays in finance and macroeconomics: the contribution of nonlinear econometrics

Soutenue par

Quentin Lajaunie

Le 3 décembre 2020

École doctorale n°543

Ecole doctorale de Dauphine

Spécialité

Sciences économiques

Composition du jury :

Christophe HURLIN Université d'Orléans	*Rapporteur*
Valérie MIGNON Université de Paris Nanterre	*Rapporteure*
Jean-Baptiste HASSE Université d'Aix-Marseille	*Membre du Jury*
Benoît SÉVI Université de Nantes	*Président du Jury*
Yannick LE PEN Université Paris-Dauphine	*Directeur de thèse*

Dauphine UNIVERSITÉ PARIS | PSL

Verso l'Alto!

Abstract

This paper-based thesis is composed of four autonomous chapters and contributes to the field of nonlinear econometrics. The first chapter focuses on the contribution of nonlinear econometrics through the measurement of financial performance using a dichotomous variable as an independent variable. The next three chapters are based on nonlinear regression models where the dichotomous variable is the dependent variable in the equation. Given the links between financial risk and the macroeconomic context, this section is linked to the theme of optimal allocation through the study of crises and recessions. This class of model (probit/logit) is used in the second chapter to empirically study the role of financial development in the probability of the occurrence of banking crises. Then, the last two chapters focus on the methodological framework developed by Kauppi and Saikkonen (2008) and Candelon, Dumitrescu and Hurlin (2012; 2014) concerning the forecasting of business cycles using probit/logit models. Thus, the third chapter examines the empirical relationship linking the evolution of the interest rate spread and the future probability of expansion/recession in an extended data panel while testing the homogeneity of this relationship. Finally, the fourth chapter proposes a theoretical contribution by deriving the response functions of probit/logit models from the approach of Kauppi and Saikkonen (2008). These response functions are then used in an empirical framework to estimate the impact of an exogenous shock on the expansion/recession cycle.

Résumé

Cette thèse sur articles est composée de quatre chapitres autonomes, contribuant au domaine de l'économétrie non-linéaire. Le premier chapitre s'intéresse à l'apport de l'économétrie non-linéaire à travers la mesure de la performance financière en utilisant une variable dichotomique comme variable indépendante. Les trois chapitres suivants sont basés sur les modèles de régression non-linéaire où la variable dichotomique est la variable dépendante de l'équation. Compte tenu des liens entre le risque financier et le contexte macroéconomique, cette partie est liée au thème de l'allocation optimale via l'étude des crises et récessions. Cette classe de modèle (probit/logit) est utilisée dans le second chapitre pour étudier empiriquement le rôle du développement financier dans la probabilité d'occurrence de crises bancaires. Ensuite, les deux derniers chapitres se concentrent sur le cadre méthodologique développé par Kauppi et Saikkonen (2008) et Candelon, Dumitrescu et Hurlin (2012 ; 2014) au sujet de la prévision des cycles économiques à partir de modèles probit/logit. Ainsi, le troisième chapitre étudie la relation empirique liant l'évolution du spread de taux et la probabilité future d'expansion / récession dans un panel de données élargi tout en testant l'homogénéité de cette relation. Enfin, le quatrième chapitre propose une contribution théorique en dérivant les fonctions de réponse des modèles probit/logit à partir de l'approche de Kauppi et Saikkonen (2008). Ces fonctions de réponse sont ensuite utilisées dans un cadre empirique afin d'estimer l'impact d'un choc exogène sur le cycle expansion / récession.

Remerciements

Tout d'abord, je tiens à exprimer ma sincère gratitude à mon directeur de thèse, Yannick Le Pen. Son soutien continu, sa disponibilité, et ses multiples conseils m'auront permis de m'épanouir personnellement et intellectuellement dans mon travail de thèse. Je tiens également à le remercier pour la confiance qu'il a su m'accorder en acceptant d'encadrer ma thèse, ainsi que pour la liberté qu'il m'a permis d'avoir dans mes travaux de recherche. Après m'avoir accompagné en Master, il a su me soutenir quelques années de plus pour réaliser cette thèse, et je lui en suis extrêment reconnaissant.

Aussi, j'adresse mes remerciements au Professeur Christophe Hurlin pour ses précieux commentaires lors de ma pré-soutenance, et au Professeur Valérie Mignon pour avoir accepté d'être rapporteure pour ma thèse.

J'aimerais aussi exprimer ma reconnaissance toute particulière à mon parrain de thèse, le Professeur Bertrand Candelon ainsi qu'à mon tuteur de thèse le Docteur Sylvain Benoît qui ont su me suivre et m'accompagner sur mes travaux de recherche.

Je remercie également le Docteur Jean-Baptiste Hasse pour les travaux que nous avons pu réaliser ensemble, pour son soutien infaillible et pour la confiance qu'il m'a témoignée.

Ayant eu l'occasion de faire partie du programme de recherche au sein de la chaire Risk Management & Investment Strategies (RMIS) issue de la collaboration entre l'Institut Louis Bachelier et d'insti7, je souhaiterais remercier l'ensemble des personnes avec qui j'ai eu l'occasion de travailler pendant ma thèse.

Je remercie aussi l'ensemble des doctorants du LEDa, spécialement CB, LB, PB, NC, EJ, et AR du bureau P157 avec qui j'aurai passé ces années.

Merci enfin à ma famille et mes amis qui auront toujours su trouver les mots et faire en sorte que je garde le moral. Merci d'avoir été là et de m'avoir permis d'avancer.

Merci à ceux que j'ai pu oublier.

Contents

List of Figures

List of Tables

General introduction

Introduction

Until the early 1990s, econometrics was essentially based on linear models. Indeed, macroeconomics was largely based on the hypothesis that the underlying dynamics of economic systems are themselves linear: this is Frisch-Slutsky's paradigm (Pesaran and Potter, 1992). However, the asymmetry in the evolution of economic cycles observed in the United States (Neftci, 1984; Falk, 1986), the United Kingdom (Burgess, 1992) and all OECD member countries (Terasvirta and Anderson, 1992) shows the limits of a linear approach. Similarly, in finance, LeBaron (1994), Mizrach (1992), and Cao and Tsay (1992) reject the hypothesis of the linearity of exchange rate and equity return series.

These questions on the relevance of using linear regressions in economics are contemporaneous with the development of another part of the econometric literature, which is interested in the development of tests of stability hypotheses. Indeed, the 1990s were also a period rich in important contributions in this field: the development of the structural change tests of Andrews (1993, 2003), Bai and Perron (1998) and Hansen (1996, 1999) contributed to the emergence of nonlinear econometrics.

Nonlinear econometrics is interested in stochastic processes with nonlinear characteristics such as asymmetric cycles, thresholds and breaks. This branch of econometrics is marked by pioneering contributions: Markov-switching regressions (Hamilton, 1989), smooth transition autoregressions (STAR) (Chan and Tong, 1986), threshold vector autoregressions (T-VAR) (Tong, 1990) and probit/logit regressions (Estrella and Hardouvelis, 1991).

The literature on the contribution of nonlinear econometrics to economics and finance has grown considerably since this seminal work. The analysis in this thesis focuses on dichotomous models. These models have been the subject of both theoretical and empirical studies. Although initially developed during the first part of the 20th century (Bliss, 1934; Berkson, 1944), probit and logit models and the use of dichotomous variables more generally were introduced into the economic and financial literature later. The occurrence of successive crises since the end of the 1990s has highlighted the advantages of this approach for predicting boom and bust phases and for measuring the impact of financial crises and the risk of certain asset classes. In less than 30 years, nonlinear econometrics has become an important part of the literature. The branch of nonlinear econometrics dealing with dichotomous variables has provided valuable tools to better understand and predict the evolution of macroeconomic and financial mechanisms. Starting in the 2000s, the work of Chauvet and Potter (2005), Kauppi and Saikonnen (2008) and Candelon, Dumitrescu and Hurlin (2013; 2014) provides a methodological framework for better anticipating the economic cycles of a given country or panel of countries by using several specifications of dichotomous models.

However, there are still questions about the properties and limitations of the proposed models. First, the effects of an exogenous shock on the expansion/recession cycle are not yet well known: the impulse-response functions of the approach of Kauppi and Saikonnen (2008) remain to be determined. Then, apart from the work of Candelon, Dumitrescu and Hurlin (2014), which partially addresses this question, the empirical validity of the dichotomous model of Kauppi and Saikkonen (2008) extended to a panel of other countries, and the homogeneity of the results has not been studied. On the other hand, the use of dichotomous models contributes to a better understanding of economic issues. Thus, we apply these models to measure the role of financial development in the occurrence of banking crises. Finally, the use of a dichotomous variable to evaluate the financial performance of an asset, conditional on an extrafinancial category of the asset under consideration, could be a modeling solution. This method would provide some answers to practitioners' concerns about the difference between ESG funds and conventional funds.

Dissertation problem statement

This thesis proposes four original contributions, in the form of autonomous papers, to the literature on nonlinear regression models with dichotomous variables. These four articles are linked by the following question: « How does the use of dichotomous models allow for concrete and innovative answers to current questions in macroeconomics and finance? »

Objectives of the thesis

This thesis proposes methodological, theoretical and empirical innovations in nonlinear regressions based on dichotomous variables. The objective of the first chapter is to propose a new methodological approach to measure the financial performance of an asset portfolio, conditional on a predefined extrafinancial classification. The aim of the second chapter is to empirically study the role of financial development in the occurrence of banking crises. Finally, based on the approach of Kauppi and Saikkonen (2008), the objective in chapters 3 and 4 is to contribute to the literature on economic cycles. Chapter 3 proposes an empirical study to test the validity of the link between interest rate spreads and the future probability of a phase change by testing the homogeneity of this relationship in a panel of OECD member countries. Finally, the fourth chapter is intended to provide a better understanding of the impact of an exogenous shock on economic cycles by introducing the response functions of dichotomous models.

Chapter 1

The first chapter examines the impact of socially responsible investment (SRI) on the financial performance of equity funds. Assets under management in this asset class have increased significantly in recent years. Much of the literature has focused on the impact that extrafinancial criteria can have on the financial performance of mutual funds. Most of these studies are based on the dichotomy between conventional mutual funds and mutual funds that present themselves as incorporating ethical investment criteria. In recent contributions, Borgers (2015) and El Ghoul and Karoui (2017) proposed an approach in which funds are compared without any distinction being made between conventional funds and funds presenting themselves as ethical. In these studies, fund performance is generally estimated using the CAPM (Sharpe, 1964; Fama and French, 1993; Carhart, 1997. The results of these studies do not converge easily. However, a consensus seems to be emerging that the impact of an ethical criterion on financial performance would be either negative or insignificant.

In this chapter, Bertrand Candelon, Jean-Baptiste Hasse and I question whether the decisions of open funds presenting themselves as ethical are in accordance with the announced principles. To answer this question, we study a panel of more than 600 European funds and nearly 900 American funds. Fund returns are extracted from the Morningstar database and cover the period 2013-2018 with a monthly frequency. Survivor bias was corrected for using the method of Elton et al. (1996). Concerning the extrafinancial criteria, we use two different databases: Morningstar Sustainability Rating and MSCI ESG Fund Metrics. The two data providers rate the funds based on their investments and on environmental, social, and governance (ESG) criteria. [1] Having two quantitative indicators calculated using different criteria allows us to ensure the robustness of the results obtained. Indeed, it seemed necessary to verify the robustness of our results following the criticism of Berg, Koelbel and Rigobon (2019). In their paper, they show that many discrepancies exist within the providers of ethical data.

Our approach consists first of identifying funds that present themselves as ethical. To do so, we build a dictionary of ethical words by extending the list of Nofsinger and Varma (2014). Then, we search for funds that contain these words in their presentation, in their KIID (Key Investor Information Document), or in their name. This allows us to differentiate between two groups: conventional funds and funds that present themselves as ethical. We compared the distribution of ESG ratings for these two groups. Conventional funds and ethical funds should theoretically have two distinct, or significantly different, distributions. Indeed, the scores of ethical funds should be on average higher than the scores of conventional funds. However, the overlap between these two distributions confirms that some funds that present themselves as ethical have lower scores than some conventional funds. This first result suggests that funds that present themselves as ethical are not necessarily ethical. To further quantitatively investigate these results,

[1] The Morningstar Sustainability Rating constructs its ratings using data from Sustainalytics, which calculates scores for each action based on 163 different indicators. These scores are then concatenated for each portfolio. MSCI scores are calculated in a similar manner, but the data source is internal and based on 68 indicators.

we estimate the performance of funds in a panel with the CAPM with 4 factors (Fama-French, 1993; Carhart, 1997), [2] including a dichotomous variable a la Hansen (2000). The dichotomous variable takes a value of 1 when the fund is advertised as ethical and 0 otherwise. This model makes it possible to estimate whether ethical and conventional funds are significantly different. The ESG score is also included as an explanatory variable in our regression.

Our results indicate that there is no significant difference in the performance of these two types of funds. The real ESG determinant of a fund is therefore not its marketing display but its investment choices. An investor who wishes to invest by integrating ethical constraints cannot limit him or herself to the KIID, the fund description, or the fund label. On the other hand, our results indicate that the impact of ethical criteria on financial performance is negative and significant. Indeed, the coefficient associated with the ethical rating of funds is negative regardless of the behavior displayed by conventional and ethical funds. This confirms that, on average, a fund that is more exposed to ESG constraints will have lower financial performance. These results show that investing in ESG companies restricts the investment universe, thereby reducing financial performance.

Chapter 2

The second chapter examines the relationship between financial development and the occurrence of banking crises. After a large part of the literature has focused on the positive impact of financial development on growth,[3] numerous empirical studies have shown that financial development can also be a source of banking crises (Demirguc-Kunt and Detragiache, 2005; Cihák, 2007; Davis et al., 2011; Schularick and Taylor, 2012; and Duca and Peltonen, 2013). However, these empirical studies do not consider different aspects of financial development. Čihák et al. (2012) decompose the development of financial institutions and financial markets in terms of depth, access, and efficiency, resulting in a final

[2]The 4-factor CAPM contains the market factor, the size factor (small minus big – SMB), the growth Factor (high minus low - HML), and the momentum factor (MOM).

[3]Schumpeter (1934), Robinson (1952), Goldsmith (1970), McKinnon and Shaw (1973), Shaw (1973), Lucas Jr. (1988), Greenwood and Jovanovic (1990), Bencivenga and Smith (1991), Saint Paul (1998), Rousseau and Wachtel (2000), Levine (2005).

financial development index. In addition, these studies use static probit/logit models. In these models, the explanatory variable corresponds to a binary variable equal to 1 when the country experiences a banking crisis and 0 otherwise. However, Kauppi and Saikkonen (2008) and Candelon, Dumitrescu and Hurlin (2014) have shown that this static form can lead to misleading estimates when the variable studied is persistent. Since banking crises are of a persistent nature, sometimes lasting several quarters, we propose to analyze the different aspects of financial development with a dynamic panel. Our empirical study covers a panel of approximately 100 countries with three levels of income: developed countries, emerging countries, and low-income countries. For the index that reflects the financial development of the countries in our panel, we use the database constructed by Čihák et al. (2012) and Sahay et al. (2015) and extended by Svirydzenka (2016). Financial development is thus decomposed into six indices that summarize the development of financial institutions and financial markets according to their depth, access and efficiency. Finally, for banking crises, we use the database of Laeven and Valencia (2013) extended with data from Candelon et al. (2020).

Kauppi and Saikkonen (2008) propose four specifications: the first is a static probit or logit model; the second refers to a dynamic probit or logit model including a lagged binary variable; the third is also dynamic and includes the lagged underlying index of the model as an explanatory variable; and, finally, the fourth takes into account both the lagged binary variable and the lagged underlying index of the model. Among these four specifications developed by Kauppi and Saikkonen (2008), we select the second one that minimizes the AIC criterion. We then follow Candelon, Dumitrescu and Hurlin (2014) and implement the correction á la Carro (2007) to correct for possible biases related to our fixed effects in our panel analysis.

The results show that access to financial institutions, the depth of financial institutions and the depth of financial markets increase the frequency of banking crises for developed countries. Conversely, the efficiency of financial institutions tends to reduce the occurrence of future banking crises. For low-income countries, access to financial institutions and financial markets reduces the likelihood of banking crises occurring, while the depth of financial institutions and financial markets and the efficiency of financial institutions tend to increase it. It can therefore be seen that the impact of financial development is not homogeneous and varies according to the countries studied and the indicators tested. These results have implications for macroprudential policies, suggesting that financial regulation should take into account the specificities of emerging markets compared to advanced countries. For example, higher capital requirements should be imposed on banks in advanced economies to offset increased access to and depth of funding, while this should not be the case for emerging market financial institutions.

Chapter 3

In the third chapter, Jean-Baptiste Hasse and I study the relationship between yield spread and the phases of the business cycle. The role of yield spread variations as a predictor of the business cycle is assumed to be a stylized fact in the literature. Many studies have examined this relationship from different perspectives. Most of them have focused on cycles in the United States, such as the research of Estrella and Hardouvelis (1991), Wright (2006), or Kauppi and Saikkonen (2008). Some of them have also studied European countries, such as Duarte, Venetis and Paya (2005) and Moneta (2005). Recently, Chinn and Kucko (2015) proposed a univariate approach to estimate this relationship across 9 countries. In this chapter, we extend these latter results by studying 13 OECD countries over the period 1975-2020 using a panel approach. [4]

This stylized fact had not previously been studied in a panel. This approach allows us to test this relationship over the period 1975-2020 while controlling for numerous macroeconomic variables established in the literature: the central bank rate (Wright, 2006), stock market returns (Nyberg, 2010), oil prices (Engemann, et al., 2011; Kilian and Vigfusson, 2017), real estate (Ng, 2012), a sentiment index (Christiansen, Eriksen, and Molleret, 2014), credit spread (Ponka, 2017), liquidity spread (Ng, 2012; Erdogan, Bennett, and Ozylidirim, 2015), uncertainty (Karnizova and Li, 2014), and volatility (Adrian, Estrella, and Shin, 2010). The panel approach provides sufficient data to test all of these variables while controlling for the persistence of the relationship between the yield spread and business cycles.

In this chapter, we use a dynamic panel model developed by Candelon, Dumitrescu and Hurlin (2014). We prefer a logistic function to a Gaussian function. Indeed, the logit model is more appropriate when we want to study extreme events such as economic

[4] The panel countries are Australia, Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, New Zealand, Sweden, Switzerland, the United States and the United Kingdom.

crises. The results confirm that the yield spread is robust to and significant under various control variables cited above over the period 1975-2020. A subsection is also devoted to the 2000-2020 segment to show the persistence of the relationship over the last 20 years.

Finally, in the second part, we develop a cluster approach to establish the groups of countries for which the relationship between the yield spread, monetary policy and the economic cycle is homogeneous. This econometric approach consists of estimating the fixed effects of each country in a first step and then the coefficients of the control variables in a second step by testing all the possible country partitions. This two-stage estimation thus makes it possible to extend the work of Zhang, Wang and Zhu (2019) to dichotomous panels. Our results confirm partial homogeneity in our panel: 2 main groups of countries stand out. In the first group, the yield spread appears to be significant, confirming its ability to predict economic cycles, contrary to the second group. [5] Finally, the link between conventional and unconventional monetary policy is discussed to understand why the coefficient of the central bank policy rate is not significant for all groups.

Chapter 4

In the fourth chapter, we determine the impulse-response function of the dichotomous models. The exact shape adapted to the four specifications of Kauppi and Saikkonen (2008)[6] is established and thus allows us to study the propagation mechanisms of an exogenous shock over a binary time series. Sims (1980) proposed a method for analyzing the behavior of a time series following an exogenous shock in the framework of a linear model. Koop, Pesaran and Potter (1996) defined the framework of the generalized response function to fit both linear and nonlinear models.[7] In this chapter, we propose to

[5] Belgium, Canada, Japan, the Netherlands, Sweden and the United States are in the first group, while Australia, Germany, New Zealand and the United Kingdom are in the second group. Finally, France, Italy and Switzerland appear in different groups.

[6] Kauppi and Saikkonen (2008) define in their paper 4 shapes of the dichotomous model. The first model is a static model in which the binary variable is estimated by different exogenous variables. The three subsequent models are dynamic: the second model integrates the delayed binary variable as an explanatory variable to integrate the persistence of crises; the third model integrates the delayed latent variable in an autoregressive form that also allows for the integration of this persistence; and, finally, the fourth model integrates both the delayed binary variable and the delayed latent variable.

[7] Work on the generalized response function has also been carried out by Gallant Raussi and Tauchen (1993) and Potter (2000)

use the framework defined by Koop, Pesaran and Potter (1996) and to adapt it to the four specifications of Kauppi and Saikkonen (2008). After formalizing the response function, we use a block-bootstrap method to determine the confidence intervals.

The exact form of the response function is then applied to the study of the effect of an exogenous shock on the yield spread over the business cycle. In this study, the binary variable corresponds to the business cycle data in the United States as determined by the *National Bureau of Economic Research* (NBER). The yield spread as an explanatory variable is calculated by the difference between the 10-year rate and the 3-month rate. The quarterly data cover the period 1953-2020.

We first estimate the four specifications of Kauppi and Saikkonen (2008): the first in which only the yield spread is used as an explanatory variable, the second and third in which we include either the lagged binary variable or the underlying index of the lagged model, and finally the fourth in which the lagged binary variable and the lagged underlying index are taken into account. Next, we follow Koop, Pesaran and Potter (1996) and determine a shock for each specification from the historical distribution of innovations. On the other hand, to determine the consequences of the shock for our system of equations, we calculate the level of the underlying index of the model from which the binary variable will take the value of 1 or 0. [8] The shock is then applied to the underlying index of the model. Since the second and fourth specifications are the most relevant in view of the AIC and BIC criteria, we focus on the latter.

The response function of the second specification depends only on the delayed binary variable. If the shock is not sufficient for the underlying index of the model to cross the threshold, the binary variable will remain unchanged, and the shock will disappear and not propagate over subsequent periods. For the fourth specification, the response function takes into account both the lagged binary variable and the underlying index of the model. The autoregressive structure of the index will tend to accentuate the persistence phenomenon, which can be amplified by a change in the binary variable. Thus, if the index crosses the threshold, the risk of observing a recession over several periods will increase

[8] In an early warning system (EWS), two types of error exist: misidentified crises (type I) and false signals (type II). The threshold is calculated conditionally on these two errors. We use the threshold *accuracy measure* presented in Candelon, Dumitrescu and Hurlin (2014). Indeed, the latter takes into account both type I and type II errors.

considerably. In our empirical study, our results show that the shock considered will lead to a recession in the United States in the following quarter for the second specification. However, the persistence of this recession will be limited to one quarter. For the fourth specification, the forecast shows that without the shock, a recession would occur in all three quarters. The exogenous shock considered will extend the duration of this shock by two additional quarters.

Thesis plan

This thesis is organized as follows. Chapter I allows one to determine whether the investment decisions of funds presenting themselves as ethical are in accordance with the announced principles. Chapter II studies the relationship between financial development and the occurrence of banking crises for a panel of countries. Chapter III examines the role of yield spread variations as a predictor of the business cycle for a panel of countries and proposes an innovative cluster methodology. Finally, the last chapter (chapter IV) studies the propagation mechanisms of an exogenous shock for dichotomous models. An empirical application is proposed to illustrate its application to cycles in the United States.

Bibliography

Adrian, T., A. Estrella, and H. S. Shin (2010). “Monetary cycles, financial cycles and the business cycle”. In: *FRB of New York Staff Report* 421.

Andrews, D. W. (1993). “Tests for parameter instability and structural change with unknown change point”. In: *Econometrica: Journal of the Econometric Society*, pp. 821–856.

— (2003). “Tests for parameter instability and structural change with unknown change point: A corrigendum”. In: *Econometrica*, pp. 395–397.

Bai, J. and P. Perron (1998). “Estimating and testing linear models with multiple structural changes”. In: *Econometrica*, pp. 47–78.

Bencivenga, V. R. and B. D. Smith (1991). “Financial intermediation and endogenous growth”. In: *The review of economic studies* 58.2, pp. 195–209.

Berg, F., J. F. Koelbel, and R. Rigobon (2019). *Aggregate Confusion: The Divergence of ESG Ratings*. MIT Sloan Research Paper.

Berkson, J. (1944). “Application of the logistic function to bio-assay”. In: *Journal of the American statistical association* 39.227, pp. 357–365.

Bliss, C. I. (1934). “The method of probits.” In: *Science* 79.2034, pp. 38–39.

Borgers, A. et al. (2015). “Do social factors influence investment behavior and performance? Evidence from mutual fund holdings”. In: *Journal of Banking & Finance* 60, pp. 112–126.

Burgess, S. M. (1992). “Asymmetric employment cycles in Britain: evidence and an explanation”. In: *The Economic Journal* 102.411, pp. 279–290.

Candelon, B., A. Carare, et al. (2020). “Globalization and the New Normal”. In: *International Economics* 161.3, pp. 139–158.

Candelon, B., E.-I. Dumitrescu, and C. Hurlin (2014). “Currency crisis early warning systems: Why they should be dynamic”. In: *International Journal of Forecasting* 30.4, pp. 1016–1029.

Candelon, B., E.-I. Dumitrescu, C. Hurlin, and F. C. Palm (2013). “Multivariate dynamic probit models: an application to financial crises mutation”. In: *VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims*. Emerald Group Publishing Limited.

Cao, C. Q. and R. S. Tsay (1992). “Nonlinear time-series analysis of stock volatilities”. In: *Journal of Applied Econometrics* 7.S1, S165–S185.

Carhart, M. M. (1997). “On persistence in mutual fund performance”. In: *The Journal of finance* 52.1, pp. 57–82.

Carro, J. M. (2007). “Estimating dynamic panel data discrete choice models with fixed effects”. In: *Journal of Econometrics* 140.2, pp. 503–528.

Chan, K. S. and H. Tong (1986). “On estimating thresholds in autoregressive models”. In: *Journal of time series analysis* 7.3, pp. 179–190.

Chauvet, M. and S. Potter (2005). “Forecasting recessions using the yield curve”. In: *Journal of forecasting* 24.2, pp. 77–103.

Chinn, M. and K. Kucko (2015). “The predictive power of the yield curve across countries and time”. In: *International Finance* 18.2, pp. 129–156.

Christiansen, C., J. N. Eriksen, and S. V. Møller (2014). “Forecasting US recessions: The role of sentiment”. In: *Journal of Banking & Finance* 49, pp. 459–468.

Čihák, M. (2007). “Systemic loss: A measure of financial stability”. In: *Czech Journal of Economics and Finance* 57.1-2, pp. 5–26.

Čihák, M. et al. (2012). “Benchmarking financial systems around the world”. In: *World Bank Policy Research Working Paper* 6175.

Davis, E. P., D. Karim, and I. Liadze (2011). “Should multivariate early warning systems for banking crises pool across regions?” In: *Review of World Economics* 147.4, pp. 693–716.

Demirgüç-Kunt, A. and E. Detragiache (2005). “Cross-country empirical studies of systemic bank distress: a survey”. In: *National Institute Economic Review* 192.1, pp. 68–83.

Duarte, A., I. A. Venetis, and I. Paya (2005). "Predicting real growth and the probability of recession in the Euro area using the yield spread". In: *International Journal of Forecasting* 21.2, pp. 261–277.

Duca, M. L. and T. A. Peltonen (2013). "Assessing systemic risks and predicting systemic events". In: *Journal of Banking & Finance* 37.7, pp. 2183–2195.

El Ghoul, S. and A. Karoui (2017). "Does corporate social responsibility affect mutual fund performance and flows?" In: *Journal of Banking & Finance* 77, pp. 53–63.

Elton, E. J., M. J. Gruber, and C. R. Blake (1996). "Survivor bias and mutual fund performance". In: *The review of financial studies* 9.4, pp. 1097–1120.

Engemann, K. M., K. L. Kliesen, and M. T. Owyang (2011). "Do oil shocks drive business cycles? Some US and international evidence". In: *Macroeconomic Dynamics* 15.S3, pp. 498–517.

Erdogan, O., P. Bennett, and C. Ozyildirim (2015). "Recession prediction using yield curve and stock market liquidity deviation measures". In: *Review of Finance* 19.1, pp. 407–422.

Estrella, A. and G. A. Hardouvelis (1991). "The term structure as a predictor of real economic activity". In: *The journal of Finance* 46.2, pp. 555–576.

Falk, B. (1986). "Further evidence on the asymmetric behavior of economic time series over the business cycle". In: *Journal of Political Economy* 94.5, pp. 1096–1109.

Fama, E. F. and K. R. French (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of financial economics* 33.1, pp. 3–56.

Gallant, A. R., P. E. Rossi, and G. Tauchen (1993). "Nonlinear dynamic structures". In: *Econometrica: Journal of the Econometric Society*, pp. 871–907.

Goldsmith, R. W. (1970). "Financial structure and development." In: *The Economic journal* 80, pp. 365–367.

Greenwood, J. and B. Jovanovic (1990). "Financial development, growth, and the distribution of income". In: *Journal of political Economy* 98.5, Part 1, pp. 1076–1107.

Hamilton, J. D. (1989). "A new approach to the economic analysis of nonstationary time series and the business cycle". In: *Econometrica: Journal of the Econometric Society*, pp. 357–384.

Hansen, B. E. (1996). "Inference when a nuisance parameter is not identified under the null hypothesis". In: *Econometrica: Journal of the econometric society*, pp. 413–430.

Hansen, B. E. (1999). “Threshold effects in non-dynamic panels: Estimation, testing, and inference”. In: *Journal of econometrics* 93.2, pp. 345–368.

— (2000). “Sample splitting and threshold estimation”. In: *Econometrica* 68.3, pp. 575–603.

Karnizova, L. and J. C. Li (2014). “Economic policy uncertainty, financial markets and probability of US recessions”. In: *Economics Letters* 125.2, pp. 261–265.

Kauppi, H. and P. Saikkonen (2008). “Predicting US recessions with dynamic binary response models”. In: *The Review of Economics and Statistics* 90.4, pp. 777–791.

Kilian, L. and R. J. Vigfusson (2017). “The role of oil price shocks in causing US recessions”. In: *Journal of Money, Credit and Banking* 49.8, pp. 1747–1776.

Koop, G., M. H. Pesaran, and S. M. Potter (1996). “Impulse response analysis in nonlinear multivariate models”. In: *Journal of econometrics* 74.1, pp. 119–147.

Laeven, L. and F. Valencia (2013). “Systemic banking crises database.” In: *IMF Economic Review* 61.2, pp. 225–270.

LeBaron, B. (1994). “Chaos and nonlinear forecastability in economics and finance”. In: *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences* 348.1688, pp. 397–404.

Lucas, R. E. (1988). “On the Mechanics of Economic Development”. In: *Journal of Monetary Economics* 22.1, pp. 3–42.

McKinnon, R. and E. Shaw (1973). *Financial deepening in economic development.* Washington, Brookings Institution.

Mizrach, B. (1992). “Multivariate nearest-neighbour forecasts of ems exchange rates”. In: *Journal of Applied Econometrics* 7.1, pp. 151–163.

Moneta, F. (2005). “Does the yield spread predict recessions in the Euro area?” In: *International Finance* 8.2, pp. 263–301.

Neftci, S. N. (1984). “Are economic time series asymmetric over the business cycle?” In: *Journal of Political Economy* 92.2, pp. 307–328.

Ng, E. C. (2012). “Forecasting US recessions with various risk factors and dynamic probit models”. In: *Journal of Macroeconomics* 34.1, pp. 112–125.

Nofsinger, J. and A. Varma (2014). “Socially responsible funds and market crises”. In: *Journal of Banking & Finance* 48, pp. 180–193.

Nyberg, H. (2010). “Dynamic probit models and financial variables in recession forecasting”. In: *Journal of Forecasting* 29.1-2, pp. 215–230.

Pesaran, M. H. and S. M. Potter (1992). “Nonlinear dynamics and econometrics: An introduction”. In: *Journal of Applied Econometrics* 7.1, pp. 1–7.

Ponka, H. (2017). “The role of credit in predicting US recessions”. In: *Journal of Forecasting* 36.5, pp. 469–482.

Potter, S. M. (2000). “Nonlinear impulse response functions”. In: *Journal of Economic Dynamics and Control* 24.10, pp. 1425–1446.

Robinson, J. (1952). *The Generalization of the General Theory.* The Rate of Interest and Other Essays (MacMillan, London).

Rousseau, P. L. and P. Wachtel (2000). “Equity markets and growth: cross-country evidence on timing and outcomes, 1980–1995”. In: *Journal of Banking & Finance* 24.12, pp. 1933–1957.

Sahay, R. et al. (2015). “Rethinking financial deepening: Stability and growth in emerging markets”. In: *Revista de Economía Institucional* 17.33, pp. 73–107.

Saint-Paul, G. (1992). “Technological choice, financial markets and economic development”. In: *European Economic Review* 36.4, pp. 763–781.

Schularick, M. and A. M. Taylor (2012). “Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008”. In: *American Economic Review* 102.2, pp. 1029–61.

Schumpeter, J. (1934). “the theory of Economic Development, tr”. In: *By Redvers Opie, Harvard Economic Studies* 40, p. 20.

Sharpe, W. F. (1964). “Capital asset prices: A theory of market equilibrium under conditions of risk”. In: *The journal of finance* 19.3, pp. 425–442.

Shaw, E. S. (1973). *Financial deepening in economic development.* New York, NY (USA) Oxford Univ. Press.

Sims, C. A. (1980). “Macroeconomics and reality”. In: *Econometrica: journal of the Econometric Society*, pp. 1–48.

Svirydzenka, K. (2016). *Introducing a new broad-based index of financial development.* International Monetary Fund.

Terasvirta, T. and H. M. Anderson (1992). “Characterizing nonlinearities in business cycles using smooth transition autoregressive models”. In: *Journal of applied econometrics* 7.S1, S119–S136.

Tong, H. (1990). *Non-linear time series: a dynamical system approach.* Oxford University Press.

Wright, J. H. (2006). *The yield curve and predicting recessions.* FEDs Working Paper.

Zhang, Y., H. J. Wang, and Z. Zhu (2019). “Quantile-regression-based clustering for panel data”. In: *Journal of Econometrics* 213.1, pp. 54–67.

Introduction générale

Introduction

Jusqu'au début des années 1990, l'économétrie est essentiellement basée sur des modèles linéaires. En effet, la macroéconomie est alors majoritairement fondée sur l'hypothèse que la dynamique sous-jacente des systèmes économiques est elle-même linéaire : c'est le paradigme de Frisch-Slutsky (Pesaran et Potter, 1992). Cependant, l'asymétrie de l'évolution des cycles économiques constatée aux Etats-Unis (Neftci, 1984 ; Falk, 1986) au Royaume-Uni (Burgess, 1992) et dans l'ensemble des pays membres de l'OCDE (Terasvirta et Anderson, 1992) montre les limites d'une approche linéaire. De même, en finance, LeBaron (1994), Mizrach (1992) et Cao et Tsay (1992) rejettent l'hypothèse de linéarité des séries de taux de change et de rendements d'actions.

Ces questionnements sur la pertinence de l'utilisation de régressions linéaires en économie sont contemporains au développement d'un autre pan de la littérature en économétrie, qui s'intéresse à l'élaboration de tests d'hypothèses de stabilité. En effet, les années 1990 ont aussi été dans ce domaine une période riche de contributions importantes : le développement des tests de changement structurel d'Andrews (1993, 2003), Bai et Perron (1998) et d'Hansen (1996, 1999) ont contribué à l'émergence de l'économétrie non-linéaire.
L'économétrie non-linéaire s'intéresse aux processus stochastiques qui présentent des caractéristiques non-linéaires comme par exemple des cycles asymétriques, ou des seuils et ruptures. Cette branche de l'économétrie est marquée par des contributions pionnières: Markov-switching regression (Hamilton, 1989), Smooth Transition Autoregressions (STAR) (Chan et Tong, 1986), Threshold Vector Autoregressions (T-VAR) (Tong, 1990) et Probit/Logit regressions (Estrella et Hardouvelis, 1991).

La littérature portant sur l'apport de l'économétrie non-linéaire en économie et en finance s'est considérablement enrichie depuis ces travaux fondateurs. Dans cette thèse, les travaux portent essentiellement sur les modèles dichotomiques. Ces modèles ont fait l'objet d'études aussi bien théoriques qu'empiriques. Bien qu'initialement développés pendant la première partie du XXème siècle (Bliss, 1934 ; Berkson, 1944), les modèles Probit et Logit, et l'utilisation de variables dichotomiques plus généralement, ont été introduits dans la littérature économique et financière plus tardivement. L'occurrence de crises successives depuis la fin des années 1990 a mis en avant les avantages de cette approche pour prévoir les phases d'expansion et de récession et pour mesurer l'impact des crises financières ainsi que le risque de certaines classes d'actifs. En moins de 30 ans, l'économétrie non-linéaire a pris une place importante dans la littérature. Et la branche de l'économétrie non-linéaire traitant des variables dichotomiques a fourni des outils précieux pour mieux comprendre et prévoir l'évolution de mécanismes macroéconomiques et financiers. A partir des années 2000, les travaux de Chauvet et Potter (2005), de Kauppi et Saikonnen (2008) et de Candelon, Dumitrescu et Hurlin (2013 ; 2014) fournissent un cadre méthodologique permettant de mieux anticiper les cycles économiques d'un pays ou d'un panel de pays donné en utilisant plusieurs spécifications de modèles dichotomiques.

Cependant, il reste encore des interrogations sur les propriétés et les limites des modèles proposés. En premier lieu, les effets d'un choc exogène sur le cycle expansion / récession sont à ce jour mal connus : les fonctions impulsions-réponse de l'approche de Kauppi et Saikonnen (2008) restent à déterminer. Ensuite, mis à part les travaux de Candelon, Dumitrescu et Hurlin (2014) qui traitent partiellement de cette question, la validité empirique du modèle dichotomique de Kauppi et Saikkonen (2008) étendue à un panel d'autres pays et l'homogénéité des résultats n'a fait état d'aucune étude. D'autre part, l'utilisation des modèles dichotomiques contribue à accroître la compréhension de problématiques économiques. Ainsi nous appliquons ces modèles pour mesurer le rôle du développement financier dans l'occurrence des crises bancaires. Enfin, l'utilisation d'une variable dichotomique pour évaluer la performance financière d'un actif, conditionnellement à une catégorie extra-financière de cet actif considéré pourrait être une solution de modélisation. Cette méthode permettrait d'apporter des éléments de réponses aux préoccupations des praticiens sur les questions traitant de la différence entre les fonds ESG et les fonds conventionnels.

Problématique de la thèse

Cette thèse propose quatre contributions originales, sous la forme d'articles autonomes, à la littérature sur les modèles de régressions non-linéaires avec des variables dichotomiques. Ces quatre articles sont liés par la problématique suivante : « En quoi l'utilisation des modèles dichotomiques permet-elle d'apporter des réponses concrètes et innovantes à des questions d'actualité en macroéconomie et en finance ? »

Objectifs de la thèse

Cette thèse propose des innovations méthodologiques, théoriques et empiriques quant aux régressions non-linéaires basées sur des variables dichotomiques. L'objectif du premier chapitre est de proposer une nouvelle approche méthodologique pour mesurer la performance financière d'un portefeuille d'actifs, et ce conditionnellement à une classification extra-financière préétablie. Le second chapitre a pour but d'étudier empiriquement le rôle du développement financier dans l'occurrence de crises bancaires. Enfin, à partir de l'approche de Kauppi et Saikkonen (2008), les chapitres 3 et 4 ont pour objectif de contribuer à la littérature sur les cycles économiques. Le chapitre 3 propose une étude empirique permettant de vérifier la validité du lien entre spread de taux et probabilité future de changement de phase, en testant l'homogénéité de cette relation dans un panel de pays membres de l'OCDE. Enfin, le quatrième chapitre a pour but de permettre une meilleure compréhension de l'impact d'un choc exogène sur les cycles économiques en introduisant les fonctions de réponses des modèles dichotomiques.

Chapitre 1

Le premier chapitre s'intéresse à l'impact de l'investissement socialement responsable (ISR) sur la performance financière des fonds actions. Les encours sous gestion dans cette classe d'actifs ont beaucoup augmenté ces dernières années. Une large partie de la littérature s'est alors intéressée à l'impact que les critères extra-financiers pouvaient avoir sur la performance financière des fonds ouverts. La plupart de ces études sont basées sur la dichotomie qui est faite entre les fonds ouverts conventionnels et les fonds ouverts qui se présentent comme intégrant des critères d'investissement éthique. Plus récemment, Borgers (2015), et El Ghoul et Karoui (2017) ont proposé une approche dans laquelle les fonds sont comparés sans qu'aucune distinction ne soit faite entre fonds conventionnels et fonds se présentant comme éthique. Dans ces études, la performance des fonds est généralement estimée à partir du CAPM (Sharpe, 1964 ; Fama et French, 1993 ; Carhart, 1997). Les résultats de ces études convergent difficilement. Toutefois, un consensus semble émerger et dire que l'impact d'un critère éthique sur les performances financières serait soit négatif, soit non significatif.

Dans ce chapitre, Bertrand Candelon, Jean-Baptiste Hasse et moi-même nous interrogeons pour savoir si les décisions des fonds ouverts se présentant comme éthiques sont en accord avec les principes annoncés. Pour répondre à cette question, nous étudions un panel de plus 600 fonds européens et près de 900 fonds américains. Les rendements des fonds sont extraits de la base de données Morningstar, et couvrent la période 2013-2018 avec une fréquence mensuelle. Le biais du survivant est corrigé à partir de la méthode d'Elton et al (1996). Concernant les critères extra-financiers, nous utilisons deux bases de données différentes: Morningstar Sustainability Rating et MSCI ESG Fund Metrics. Les deux fournisseurs de données attribuent aux fonds des notes à partir de leurs investissements et de critères Environnementaux, Sociaux, et de Gouvernance (ESG). [9] Le fait d'avoir deux indicateurs quantitatifs calculés à partir de critères différents nous permet de nous assurer de la robustesse des résultats trouvés. En effet, il nous a paru nécessaire de vérifier la solidité de nos résultats suite à la critique de Berg, Koelbel et Rigobon (2019). Dans leur papier, ils montrent que de nombreuses divergences existent au sein des fournisseurs de données éthiques.

Notre démarche consiste en un premier temps à identifier les fonds qui se présentent comme éthiques. Pour cela, nous construisons un dictionnaire de mots éthiques en étendant la liste de Nofsinger et Varma (2014). Ensuite, nous recherchons les fonds qui contiennent ces mots dans leur présentation, dans leur DICI (Document d'Information Clé pour l'Investisseur), ou dans leur nom. Cela nous permet de différencier deux groupes : les fonds conventionnels, et les fonds qui se présentent comme éthiques. Nous comparons la distribution des notes ESG de ces deux groupes. Les fonds conventionnels et les fonds

[9] Morningstar Sustainability Rating construit ses notes en s'appuyant sur les données de Sustainalytics qui calcule des scores pour chaque action à partir de 163 indicateurs différents. Ces scores sont ensuite concaténer pour chaque portefeuille. Les notes de MSCI sont calculées de façon similaire, mais la source de données est interne et s'appuie sur 68 indicateurs.

éthiques devraient avoir théoriquement deux distributions disctinctes, ou significativement différentes. En effet, les notes des fonds éthiques devraient être en moyenne supérieures aux notes des fonds conventionnels. Toutefois, le chevauchement de ces deux distributions confirme que certains fonds qui se présentent comme éthiques ont une note plus faible que certains fonds conventionnels. Ce premier résultat suggère que les fonds qui se présentent comme éthiques ne le sont pas forcément. Afin d'approfondir ces résultats de façon quantitative, nous estimons la performance des fonds dans un panel avec le CAPM à 4 facteurs (Fama-French, 1993 ; Carhart, 1997) [10] en incluant une variable dichotomique à la Hansen (2000). La variable dichotomique prend la valeur de 1 lorsque le fonds est annoncé comme éthique, et 0 sinon. Ce modèle permet d'estimer si les fonds éthiques et conventionnels sont significativement différents. La note ESG est également inclue comme variable explicative dans notre régression.

Nos résultats indiquent qu'il n'y a pas de différence significative dans la performance de ces deux types de fonds. Le réel déterminant ESG d'un fonds n'est donc pas son affichage marketting, mais ses choix d'investissement. Un investisseur qui souhaite investir en intégrant des contraintes éthiques ne peut pas se limiter au DICI, à la description du fonds, ou au label de ce dernier. D'autre part, nos résultats indiquent que l'impact des critères éthiques sur la performance financière est négatif et significatif. En effet, le coefficient associé à la note éthique des fonds est négatif quelque soit le comportement affiché des fonds conventionnels et éthiques. Cela confirme qu'en moyenne, un fonds qui est plus exposé à des contraintes ESG aura une performance financière inférieure. Ces résultats montrent que le fait d'investir dans des entreprises ESG restreignent l'univers d'investissement, entraînant de ce fait une réduction de la performance financière.

Chapitre 2

Le deuxième chapitre étudie la relation entre le développement financier et l'occurrence des crises bancaires. Après qu'une large partie de la littérature se soit intéressée à l'impact

[10] Le CAPM à 4 facteurs contient le facteur de marché, le facteur taille (Small Minus Big - SMB), le facteur de croissance (High Minus Low - HML), et le facteur Momentum (MOM)

positif du développement financier sur la croissance, [11] de nombreuses études empiriques ont montré que le développement financier pouvait également être à l'origine de crises bancaires (Demirgüç-Kunt et Detragiache, 2005, Cihák, 2007, Davis et al., 2011; Schularick et Taylor, 2012; et Duca et Peltonen, 2013). Toutefois, ces études empiriques ne considèrent pas les différents aspects du développement financier. Čihák et al. (2012) décomposent le développement des institutions financières et des marchés financiers en termes de profondeur, d'accès et d'efficacité, aboutissant à l'indice final de développement financier. De plus, ces études recourent à des modèles probit/logit statiques. Dans ces modèles, la variable à expliquer correspond à une variable binaire égale à 1 lorsque le pays connaît une crise bancaire, et 0 sinon. Or, Kauppi et Saikkonen (2008) et Candelon, Dumitrescu et Hurlin (2014) ont montré que cette forme statique pouvait être à l'origine d'estimations trompeuses lorsque la variable étudiée était persistante. Les crises bancaires étant de natures persistantes, pouvant parfois durer plusieurs trimestres, nous nous proposons d'analyser les différents aspects du développement financier avec un panel dynamique. Notre étude empirique porte sur un panel d'une centaine de pays avec trois niveaux de revenu : pays développés, pays émergents, et pays à faibles revenus. Concernant l'indice qui reflète le développement financier des pays de notre panel, nous utilisons la base de données construite par Čihák et al. (2012), Sahay et al. (2015) et étendue par Svirydzenka (2016). Le développement financier est ainsi décomposé en six indices qui résument le développement des institutions financières et des marchés financiers selon leur profondeur, accès et efficacité. Enfin, pour les crises bancaires, nous utilisons la base de données de Laeven et Valencia (2013) étendue avec les données de Candelon et al. (2020).

Kauppi et Saikkonen (2008) proposent quatre spécifications: la première est un modèle probit ou logit statique ; la seconde renvoie à un modèle probit ou logit dynamique incluant la variable binaire retardée; la troisième est également dynamique, et intègre l'indice sous-jacent retardé du modèle comme variable explicative ; enfin, la quatrième prend à la fois en compte la variable binaire retardée et l'indice sous-jacent retardé du modèle. Parmi ces quatre spécifications développées par Kauppi et Saikkonen (2008), nous sélectionnons la deuxième qui minimise le critère AIC. Ensuite, nous suivons Candelon, Dumitrescu et

[11] Schumpeter (1934) , Robinson (1952), Goldsmith (1970) , McKinnon et Shaw (1973) , Shaw (1973) , Lucas Jr (1988) , Greenwood et Jovanovic (1990), Bencivenga et Smith (1991), Saint-Paul (1998), Rousseau et Wachtel (2000), Levine (2005).

Hurlin (2014), et nous implémentons la correction á la Carro (2007) permettant de corriger d'éventuels biais liés à nos effets fixes dans notre analyse en panel.

Les résultats montrent que l'accès aux institutions financières, la profondeur des institutions financières et la profondeur des marchés financiers accentuent la fréquence de crises bancaires pour les pays développés. A l'inverse, l'efficacité des institutions financières tend à réduire les occurrences des crises bancaires à venir. Pour les pays à faibles revenus, l'accès aux institutions financières et aux marchés financiers réduit la probabilité de survenance des crises bancaires tandis que la profondeur des institutions financières et des marchés financiers, ainsi que l'efficacité des institutions financières ont tendance à l'augmenter. On observe donc que l'impact du développement financier n'est pas homogène et varie en fonction des pays étudiés ainsi que des indicateurs testés. Ces résultats ont des conséquences pour les politiques macroprudentielles, suggérant que la réglementation financière devrait prend en compte les spécificités des marchés émergents par rapport aux pays avancés. Par exemple, des exigences de fonds propres plus élevées devraient être imposées aux banques des économies avancées afin de contrebalancer l'augmentation de l'accès et de la profondeur du financement, alors que cela ne devrait pas être le cas pour les institutions financières des marchés émergents.

Chapitre 3

Dans le troisième chapitre, Jean-Baptiste Hasse et moi-même étudions la relation entre *yield spread* et les phases du cycle économique. Le rôle des variations du yield spread comme prédicteur du cycle économique est assumé comme un *stylized fact* dans la littérature. De nombreuses études ont étudié cette relation sous différents angles. La plupart d'entre elles ont porté sur les cycles aux Etats-Unis, avec notamment les recherches de Estrella et Hardouvelis (1991), Wright (2006), ou encore Kauppi et Saikkonen (2008). Certaines d'entre elles ont également étudié des pays européens, comme par exemple Duarte, Venetis et Paya (2005), ou encore Moneta (2005). Récemment, Chinn et Kucko (2015) ont proposé une approche univariée pour estimer cette relation sur 9 pays. Dans ce chapitre, nous étendons ces derniers résultats en étudiant 13 pays de l'OCDE sur la période 1975-2020 avec une approche en panel. [12]

[12] Les pays du panel sont : Allemagne, Australie, Belgique, Canada, Etats-Unis, France, Italie, Japon, Nouvelle-Zélande, Pays-Bas, Royaume-Uni, Suède, et Suisse.

Ce *stylized fact* n'avait jusqu'ici pas été étudié en panel. Cette approche nous permet de tester cette relation sur la période 1975-2020, tout en la contrôlant par de nombreuses variables macroéconomiques établies dans la littérature : le taux de banque centrale (Wright, 2006), les rendements du marché action (Nyberg, 2010), le cours du pétrole (Engemann, et al., 2011; Kilian et Vigfusson, 2017), l'immobilier (Ng, 2012), un indice de sentiment (Christiansen, Eriksen et Molleret, 2014), le spread de crédit (Ponka, 2017), le spread de liquidité (Ng, 2012 ; Erdogan, Bennett, et Ozylidirim, 2015), l'incertitude (Karnizova et Li, 2014), et la volatilité (Adrian, Estrella et Shin, 2010). L'approche en panel permet de disposer des données suffisantes pour tester l'ensemble de ces variables tout en contrôlant la persistance de la relation entre les *yield spread* et les cycles économiques.

Dans ce chapitre, nous utilisons un modèle de panel dynamique développé par Candelon, Dumitrescu et Hurlin (2014). Nous préférons une fonction logistique à une gaussienne. En effet, le modèle Logit est plus approprié lorsque l'on cherche à étudier des événements extrêmes telles que des crises économiques. Les résultats confirment que le *yield spread* est robuste et significatif aux différentes variables de contrôle citées précédemment sur la période 1975-2020. Une sous-section est également consacrée au segment 2000-2020, ceci dans le but de montrer la persistance de la relation sur les 20 dernières années.

Enfin, dans une deuxième partie, nous développons une approche en clusters, afin d'établir les groupes de pays pour lesquels la relation entre le *yield spread*, la politique monétaire et le cycle économique est homogène. Cette approche économétrique consiste à estimer les effets fixes de chaque pays dans un premier temps, puis les coefficients des variables de contrôle dans un second temps, en testant l'ensemble des partitions possibles de pays. Cette estimation en deux étapes permet ainsi d'étendre les travaux de Zhang, Wang et Zhu (2019) aux panels dichotomiques. Nos résultats confirment une homogénéité partielle de notre panel : 2 groupes principaux de pays se distinguent. Dans le premier groupe, le *yield spread* apparaît comme significatif confirmant ainsi sa capacité à prévoir les cycles économiques contrairement au deuxième groupe. [13] Enfin, le lien entre la politique monétaire conventionnelle et non conventionnelle est discuté afin de comprendre pourquoi le coefficient du taux directeur des banques centrales n'est pas significatif pour l'ensemble

[13] La Belgique, le Canada, le Japon, les Pays-Bas, la Suède et les Etats-Unis sont dans le premier groupe, tandis que l'Australie, l'Allemagne, la Nouvelle-Zélande et le Royaume-Unis. Enfin, la France, l'Italie et la Suisse apparaissent dans des groupes différents.

des groupes.

Chapitre 4

Dans le quatrième chapitre, nous déterminons la fonction impulsion-réponse des modèles dichotomiques. La forme exacte adaptée aux quatre spécifications de Kauppi et Saikkonen (2008)[14] est établie et permet ainsi d'étudier les méchanismes de propagation d'un choc exogène sur une série temporelle binaire. Sims (1980) a proposé une méthode permettant d'analyser le comportement d'une série temporelle à la suite d'un choc exogène dans le cadre d'un modèle linéaire. Koop, Pesaran et Potter (1996) ont défini le cadre de la fonction de réponse généralisée afin qu'elle soit adaptée à la fois aux modèles linéaire et non-linéaires.[15] Dans ce chapitre, nous proposons d'utiliser le cadre défini par Koop, Pesaran et Potter, et de l'adapter aux quatre spécifications de Kauppi et Saikkonen (2008). Après avoir formalisé la fonction de réponse, nous utilisons une méthode de block-bootstrapp pour déterminer les intervalles de confiance.

La forme exacte de la fonction de réponse est ensuite appliquée à l'étude de l'effet d'un choc exogène sur le *yield spread* sur le cycle économique. Dans cette étude, la variable binaire correspond aux données du cycle économique aux Etats-Unis déterminés par le *National Bureau of Economic Research* (NBER). Le *yield spread* comme variable explicative est calculé par la différence du taux 10 ans et du taux 3 mois. Les données à fréquence trimestrielle couvrent la période 1953-2020.

Nous estimons dans un premier temps les quatre spécifications de Kauppi et Saikkonen (2008): la première dans laquelle seul le yield spread est utilisé comme variable explicative, la deuxième et la troisième dans lesquelles nous incluons soit la variable binaire retardée soit l'indice sous-jacent du modèle retardé, et enfin la quatrième dans laquelle la variable binaire retardée et l'indice sous-jacent retardé sont pris en compte. Ensuite, nous suivons

[14]Kauppi et Saikkonen (2008) définissent dans leur papier 4 formes du modèle dichotomique. La premier modèle est un modèle statique dans lequel la variable binaire est estimée par différentes variables exogènes. Les trois modèles suivants sont dynamiques: le deuxième modèle intègre la variable binaire retardée comme variable explicative afin d'intégrer la persistance des crises; le troisième modèle intègre la variable latente retardée sous une forme autorégressive permettant également d'intégrer cette persistance; enfin, le quatrième modèle intègre à la fois la variable binaire retardée et la variable latente retardée.

[15]Des travaux sur la fonction de réponse généralisée ont également été menés par Gallant Raussi et Tauchen (1993) et Potter (2000)

Koop, Pesaran et Potter (1996) et nous déterminons un choc pour chaque spécification à partir de la distribution historique des innovations. D'autre part, afin de déterminer les conséquences du choc sur notre système d'équations, nous calculons le niveau de l'indice sous-jacent du modèle à partir duquel la variable binaire prendra la valeur de 1 ou 0. [16] Le choc est ensuite appliqué à l'indice sous-jacent du modèle. La deuxième et la quatrième spécifications étant les plus pertinentes au vu des critères AIC et BIC, nous nous focalisons sur ces dernières.

La fonction de réponse de la deuxième spécification ne dépend que de la variable binaire retardée. Si le choc ne suffit pas à ce que l'indice-sous jacent du modèle franchisse le seuil, la variable binaire restera inchangée et le choc disparaîtra et ne se propagera pas sur les périodes suivantes. Pour la quatrième spécification, la fonction de réponse prend en compte à la fois la variable binaire retardée et l'indice sous-jacent du modèle. La structure autorégressive de l'indice aura tendance à accentuer le phénomène de persistance, pouvant être amplifié par un changement de la variable binaire. Ainsi, si l'indice franchit le seuil, le risque d'observer une récession sur plusieurs périodes augmentera considérablement. Dans notre étude empirique, nos résultats montrent que le choc considéré entraînera une récession aux Etats-Unis au trimestre suivant pour la deuxième spécification. Toutefois, la persistance de cette récession se limitera à un trimestre. Concernant la quatrième spécification, la prévision affiche que sans choc une récession aurait lieu pendant les trois trimestres. Le choc exogène considéré allongera la durée de ce choc de deux trimestres supplémentaires.

[16] Dans un Early Warning System (EWS), deux types d'erreur existent: les crises mal identifiées (type I) et les faux signaux (type II). Le seuil est calculé conditionnellement à ces deux erreurs. Nous utilisons le seuil *Accuracy Measure* présenté dans Candelon, Dumitrescu et Hurlin (2014). En effet, ce dernier prend en compte à la fois l'erreur de type I et l'erreur de type II.

Plan

Cette thèse est organisée comme suit. Le chapitre I permet de déterminer si les décisions d'investissement des fonds se présentant comme éthiques sont en accord avec les principes annoncés. Le chapitre II étudie la relation entre le développement financier et l'occurence des crises bancaires pour un panel de pays. Le chapitre III s'intéresse au rôle des variations du yield spread comme prédicteur du cycle économique pour un panel de pays et propose une méthodologie innovante de cluster. Enfin, le dernier chapitre (chapitre IV) étudie les mécanismes de propagation d'un choc exogène pour les modèles dichotomiques. Une application empirique est proposée pour illuster son application sur les cycles aux Etats-Unis.

Bibliography

Adrian, T., A. Estrella, and H. S. Shin (2010). “Monetary cycles, financial cycles and the business cycle”. In: *FRB of New York Staff Report* 421.

Andrews, D. W. (1993). “Tests for parameter instability and structural change with unknown change point”. In: *Econometrica: Journal of the Econometric Society*, pp. 821–856.

— (2003). “Tests for parameter instability and structural change with unknown change point: A corrigendum”. In: *Econometrica*, pp. 395–397.

Bai, J. and P. Perron (1998). “Estimating and testing linear models with multiple structural changes”. In: *Econometrica*, pp. 47–78.

Bencivenga, V. R. and B. D. Smith (1991). “Financial intermediation and endogenous growth”. In: *The review of economic studies* 58.2, pp. 195–209.

Berg, F., J. F. Koelbel, and R. Rigobon (2019). *Aggregate Confusion: The Divergence of ESG Ratings*. MIT Sloan Research Paper.

Berkson, J. (1944). “Application of the logistic function to bio-assay”. In: *Journal of the American statistical association* 39.227, pp. 357–365.

Bliss, C. I. (1934). “The method of probits.” In: *Science* 79.2034, pp. 38–39.

Borgers, A. et al. (2015). “Do social factors influence investment behavior and performance? Evidence from mutual fund holdings”. In: *Journal of Banking & Finance* 60, pp. 112–126.

Burgess, S. M. (1992). “Asymmetric employment cycles in Britain: evidence and an explanation”. In: *The Economic Journal* 102.411, pp. 279–290.

Candelon, B., A. Carare, et al. (2020). “Globalization and the New Normal”. In: *International Economics* 161.3, pp. 139–158.

Candelon, B., E.-I. Dumitrescu, and C. Hurlin (2014). “Currency crisis early warning systems: Why they should be dynamic”. In: *International Journal of Forecasting* 30.4, pp. 1016–1029.

Candelon, B., E.-I. Dumitrescu, C. Hurlin, and F. C. Palm (2013). “Multivariate dynamic probit models: an application to financial crises mutation”. In: *VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims*. Emerald Group Publishing Limited.

Cao, C. Q. and R. S. Tsay (1992). “Nonlinear time-series analysis of stock volatilities”. In: *Journal of Applied Econometrics* 7.S1, S165–S185.

Carhart, M. M. (1997). “On persistence in mutual fund performance”. In: *The Journal of finance* 52.1, pp. 57–82.

Carro, J. M. (2007). “Estimating dynamic panel data discrete choice models with fixed effects”. In: *Journal of Econometrics* 140.2, pp. 503–528.

Chan, K. S. and H. Tong (1986). “On estimating thresholds in autoregressive models”. In: *Journal of time series analysis* 7.3, pp. 179–190.

Chauvet, M. and S. Potter (2005). “Forecasting recessions using the yield curve”. In: *Journal of forecasting* 24.2, pp. 77–103.

Chinn, M. and K. Kucko (2015). “The predictive power of the yield curve across countries and time”. In: *International Finance* 18.2, pp. 129–156.

Christiansen, C., J. N. Eriksen, and S. V. Møller (2014). “Forecasting US recessions: The role of sentiment”. In: *Journal of Banking & Finance* 49, pp. 459–468.

Čihák, M. (2007). “Systemic loss: A measure of financial stability”. In: *Czech Journal of Economics and Finance* 57.1-2, pp. 5–26.

Čihák, M. et al. (2012). “Benchmarking financial systems around the world”. In: *World Bank Policy Research Working Paper* 6175.

Davis, E. P., D. Karim, and I. Liadze (2011). “Should multivariate early warning systems for banking crises pool across regions?” In: *Review of World Economics* 147.4, pp. 693–716.

Demirgüç-Kunt, A. and E. Detragiache (2005). “Cross-country empirical studies of systemic bank distress: a survey”. In: *National Institute Economic Review* 192.1, pp. 68–83.

Duarte, A., I. A. Venetis, and I. Paya (2005). "Predicting real growth and the probability of recession in the Euro area using the yield spread". In: *International Journal of Forecasting* 21.2, pp. 261–277.

Duca, M. L. and T. A. Peltonen (2013). "Assessing systemic risks and predicting systemic events". In: *Journal of Banking & Finance* 37.7, pp. 2183–2195.

El Ghoul, S. and A. Karoui (2017). "Does corporate social responsibility affect mutual fund performance and flows?" In: *Journal of Banking & Finance* 77, pp. 53–63.

Elton, E. J., M. J. Gruber, and C. R. Blake (1996). "Survivor bias and mutual fund performance". In: *The review of financial studies* 9.4, pp. 1097–1120.

Engemann, K. M., K. L. Kliesen, and M. T. Owyang (2011). "Do oil shocks drive business cycles? Some US and international evidence". In: *Macroeconomic Dynamics* 15.S3, pp. 498–517.

Erdogan, O., P. Bennett, and C. Ozyildirim (2015). "Recession prediction using yield curve and stock market liquidity deviation measures". In: *Review of Finance* 19.1, pp. 407–422.

Estrella, A. and G. A. Hardouvelis (1991). "The term structure as a predictor of real economic activity". In: *The journal of Finance* 46.2, pp. 555–576.

Falk, B. (1986). "Further evidence on the asymmetric behavior of economic time series over the business cycle". In: *Journal of Political Economy* 94.5, pp. 1096–1109.

Fama, E. F. and K. R. French (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of financial economics* 33.1, pp. 3–56.

Gallant, A. R., P. E. Rossi, and G. Tauchen (1993). "Nonlinear dynamic structures". In: *Econometrica: Journal of the Econometric Society*, pp. 871–907.

Goldsmith, R. W. (1970). "Financial structure and development." In: *The Economic journal* 80, pp. 365–367.

Greenwood, J. and B. Jovanovic (1990). "Financial development, growth, and the distribution of income". In: *Journal of political Economy* 98.5, Part 1, pp. 1076–1107.

Hamilton, J. D. (1989). "A new approach to the economic analysis of nonstationary time series and the business cycle". In: *Econometrica: Journal of the Econometric Society*, pp. 357–384.

Hansen, B. E. (1996). "Inference when a nuisance parameter is not identified under the null hypothesis". In: *Econometrica: Journal of the econometric society*, pp. 413–430.

Hansen, B. E. (1999). “Threshold effects in non-dynamic panels: Estimation, testing, and inference”. In: *Journal of econometrics* 93.2, pp. 345–368.

— (2000). “Sample splitting and threshold estimation”. In: *Econometrica* 68.3, pp. 575–603.

Karnizova, L. and J. C. Li (2014). “Economic policy uncertainty, financial markets and probability of US recessions”. In: *Economics Letters* 125.2, pp. 261–265.

Kauppi, H. and P. Saikkonen (2008). “Predicting US recessions with dynamic binary response models”. In: *The Review of Economics and Statistics* 90.4, pp. 777–791.

Kilian, L. and R. J. Vigfusson (2017). “The role of oil price shocks in causing US recessions”. In: *Journal of Money, Credit and Banking* 49.8, pp. 1747–1776.

Koop, G., M. H. Pesaran, and S. M. Potter (1996). “Impulse response analysis in nonlinear multivariate models”. In: *Journal of econometrics* 74.1, pp. 119–147.

Laeven, L. and F. Valencia (2013). “Systemic banking crises database.” In: *IMF Economic Review* 61.2, pp. 225–270.

LeBaron, B. (1994). “Chaos and nonlinear forecastability in economics and finance”. In: *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences* 348.1688, pp. 397–404.

Lucas, R. E. (1988). “On the Mechanics of Economic Development”. In: *Journal of Monetary Economics* 22.1, pp. 3–42.

McKinnon, R. and E. Shaw (1973). *Financial deepening in economic development.* Washington, Brookings Institution.

Mizrach, B. (1992). “Multivariate nearest-neighbour forecasts of ems exchange rates”. In: *Journal of Applied Econometrics* 7.1, pp. 151–163.

Moneta, F. (2005). “Does the yield spread predict recessions in the Euro area?” In: *International Finance* 8.2, pp. 263–301.

Neftci, S. N. (1984). “Are economic time series asymmetric over the business cycle?” In: *Journal of Political Economy* 92.2, pp. 307–328.

Ng, E. C. (2012). “Forecasting US recessions with various risk factors and dynamic probit models”. In: *Journal of Macroeconomics* 34.1, pp. 112–125.

Nofsinger, J. and A. Varma (2014). “Socially responsible funds and market crises”. In: *Journal of Banking & Finance* 48, pp. 180–193.

Nyberg, H. (2010). “Dynamic probit models and financial variables in recession forecasting”. In: *Journal of Forecasting* 29.1-2, pp. 215–230.

Pesaran, M. H. and S. M. Potter (1992). “Nonlinear dynamics and econometrics: An introduction”. In: *Journal of Applied Econometrics* 7.1, pp. 1–7.

Ponka, H. (2017). “The role of credit in predicting US recessions”. In: *Journal of Forecasting* 36.5, pp. 469–482.

Potter, S. M. (2000). “Nonlinear impulse response functions”. In: *Journal of Economic Dynamics and Control* 24.10, pp. 1425–1446.

Robinson, J. (1952). *The Generalization of the General Theory.* The Rate of Interest and Other Essays (MacMillan, London).

Rousseau, P. L. and P. Wachtel (2000). “Equity markets and growth: cross-country evidence on timing and outcomes, 1980–1995”. In: *Journal of Banking & Finance* 24.12, pp. 1933–1957.

Sahay, R. et al. (2015). “Rethinking financial deepening: Stability and growth in emerging markets”. In: *Revista de Economía Institucional* 17.33, pp. 73–107.

Saint-Paul, G. (1992). “Technological choice, financial markets and economic development”. In: *European Economic Review* 36.4, pp. 763–781.

Schularick, M. and A. M. Taylor (2012). “Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008”. In: *American Economic Review* 102.2, pp. 1029–61.

Schumpeter, J. (1934). “the theory of Economic Development, tr”. In: *By Redvers Opie, Harvard Economic Studies* 40, p. 20.

Sharpe, W. F. (1964). “Capital asset prices: A theory of market equilibrium under conditions of risk”. In: *The journal of finance* 19.3, pp. 425–442.

Shaw, E. S. (1973). *Financial deepening in economic development.* New York, NY (USA) Oxford Univ. Press.

Sims, C. A. (1980). “Macroeconomics and reality”. In: *Econometrica: journal of the Econometric Society*, pp. 1–48.

Svirydzenka, K. (2016). *Introducing a new broad-based index of financial development.* International Monetary Fund.

Terasvirta, T. and H. M. Anderson (1992). “Characterizing nonlinearities in business cycles using smooth transition autoregressive models”. In: *Journal of applied econometrics* 7.S1, S119–S136.

Tong, H. (1990). *Non-linear time series: a dynamical system approach*. Oxford University Press.

Wright, J. H. (2006). *The yield curve and predicting recessions*. FEDs Working Paper.

Zhang, Y., H. J. Wang, and Z. Zhu (2019). “Quantile-regression-based clustering for panel data”. In: *Journal of Econometrics* 213.1, pp. 54–67.

Chapter 1

Ethics and Information Asymmetry

Authors: Bertrand Candelon, Jean-Baptiste Hasse, Quentin Lajaunie

About this chapter

The authors thank Florian Berg, Franz Fuerst, Yannick Le Pen, Arno Riedl and Paul Smeets for helpful comments as well as the seminar participants at the 2nd Annual Conference of the Global Research Alliance for Sustainable Finance and Investment - GRASFI (University of Oxford, 2019), 9th International Conference of the Financial Engineering and Banking Society - FEBS (University of Economics, Prague, 2019), 68th Annual Meeting of the French Economic Association - AFSE (University of Orléans, 2019), the seminars of the Université Catholique de Louvain (2018) and the Institut Louis Bachelier (2018). This research was conducted as part of the research program entitled “Risk Management, Investment Strategies and Financial Stability” under the aegis of the Europlace Institute of Finance, a joint initiative with Insti7. The usual disclaimer applies.

Abstract

In this paper, we analyze the asymmetric information between investors and asset managers about financial and extra-financial mutual funds performances. Using a unique panel dataset consisting of European and US equity mutual funds, we empirically test the information transparency about the classification and performance of conventional and socially responsible mutual funds. Specifically, we first highlight that investors may have difficulty identifying the extra-financial characteristics of a given mutual fund due to the weak correspondence between asset managers' branding and portfolio holdings. Second, we empirically test the impact of extra-financial classifications on mutual funds performances. It is found that only the *de facto* ethical positioning matters for the funds' financial performance. Both results advocate for a unified regulation framework to reduce information asymmetry in the SRI market.

Keywords: Socially Responsible Investing (SRI); Environmental, Social and Governance (ESG) Criteria ; Ethical Mutual Funds; Performance Measurement.

1.1 Introduction

In this paper, we analyze the asymmetric information between investors and asset managers about financial and extra-financial mutual funds performances. We test the information transparency about the classification and performance of conventional and socially responsible mutual funds. Specifically, we first highlight that investors may have difficulty identifying the extra-financial characteristics of a given mutual fund due to the weak correspondence between asset managers' branding and portfolio holdings. Then, as asset managers' investment strategies determine their mutual funds' financial performance objectives related to their benchmark indices, the performance measurement is also impacted by the fund classification. Second, we therefore investigate the impact of different extra-financial classifications on financial performance.

Such an asymmetric information in the socially responsible investing (SRI) market would constitute a key issue for both investors and the mutual fund industry. According to the Social Investment Forum's (SIF) report (2016), the SRI market represents more than 20% of the mutual fund industry's total assets under professional management in the United States. In Europe, the same phenomenon is observed, as the number of socially responsible mutual funds grew by 12% between 2014 and 2016.[1] This significant growth in the SRI market has been driven by ethical criteria from investors regarding the environmental, social and governance (ESG) impacts of their investments (See Barber et al., 2018; Hartzmark and Sussman, 2019). This evolution originates from the preferences governing socially responsible investors behavior. Such preferences can be represented by a multi-attribute utility function in the sense that some investors appear to derive utility from being exposed to the socially responsible attribute (Statman, 2004; Bollen, 2007; Fama and French, 2007). Broadly, socially responsible mutual funds invest in firms that adhere to social, environmental, moral or religious beliefs. Although SRI has become customary in the current language, a precise definition of such mutual funds remains relatively subjective. In 2015, the French financial market regulator (Autorité des Marchés Financiers - AMF) defined SRI as *"a polymorphous and evolving concept that is sometimes difficult to understand"*.[2] This lack of transparency has led to the development of an industry aimed at providing SRI labels for mutual funds. In France, for example, the Novethic label was established in 2009 for European countries, and the FNG label was adopted in 2001 in German-speaking countries. Similarly, nonprofit organizations, such as the SIF, publish their own socially responsible asset manager lists, which are built from memberships in several SRI communities or adherence to ethical criteria of investing (e.g., Principles for Responsible Investment – PRI) and accreditation for labels. In light of the growing supply and subjective definitions of SRI, asset managers now also seek to obtain these labels in addition to adopting an ethically styled name as a branding strategy to attract investors. Indeed, asset managers design their mutual funds supply to meet the demand of investors. This demand depends on the investors' multi-attribute utility function, which can be expressed as financial and socially responsible attributes (Bollen, 2007). However, behavioral biases in the decision making of investors exist. Among others, the

[1] The authors' calculations based on several sources: the Global Sustainable Investment review (2016), available at `www.gsi-alliance.org`; the European SRI Study (2016), available at `http://www.eurosif.org/`; and the US SRI trend report (2016), available at `https://www.ussif.org/`.

[2] Quote extracted from "Rapport de l'AMF sur l'investissement responsable dans la gestion collective", November 26, 2015.

appearances of assets impact investment choices. For instance, the company name affects investors' decisions regarding its stocks (Green and Jame, 2013; Jacobs and Hillert, 2016). Similarly, mutual funds names bias investors' decisions (Cooper et al., 2001; Doellman et al., 2019). Furthermore, this behavioral bias also exists concerning ESG decision criteria (El Ghoul and Karoui, 2020). Asset managers can benefit from this behavioral bias using mutual funds' names as a branding tool. Cooper et al. (2005), and more recently Espenlaub et al. (2017), provide evidence of such opportunistic behavior (Schwarz, 2003). In the end, asymmetric information emerges in the SRI market from both investors' bounded rationality and asset managers' opportunistic behavior. [3]

This paper proposes to shed new light on socially responsible mutual funds by going beyond these appearances driven by asset managers' branding. It investigates whether an SRI branding (via self-presentation or label accreditation), broadly speaking, indeed signals that a mutual fund invests according to ESG criteria or whether it constitutes only a purely branding position. To this end, a 2-dimensional measure of the SRI characteristics of mutual funds is built. The first dimension concerns the branding of asset managers or the accreditation (or not) by a specialized audit agency of the social responsibility of the mutual funds. This information provides us with the willingness of the asset managers to identify their mutual funds as socially responsible (via self-identification or via a labeling partner). In this paper, "SRI branding" refers to asset managers ethically positioning the brand of the mutual fund. Furthermore, we qualify such investments "*de jure* socially responsible", as this ethical positioning does not necessarily involve ESG criteria. Following the literature, we identify this *de jure* ethical positioning by mutual funds names and labels. The second dimension covers the realizations of these commitments. This aspect is measured via the newly available Morningstar Sustainability Rating and MSCI ESG Fund Metrics databases, which present many advantages. In particular, these extra-financial ratings are continuous (defined between 0-100 and 0-10, respectively), normalized, and homogeneous for all European and US mutual funds. Interestingly, Morningstar's ratings are free of industry bias and propose a controversy score. Hence, Morningstar's ratings are considered the benchmark in the sequel of the paper, whereas the MSCI's ratings will be introduced as a robustness check.

An analysis of the correspondence between the two dimensions tells us whether SRI

[3]See Rhodes (2010) for a broader discussion about information asymmetry in the SRI market.

branding provide sufficient information about the SRI identification of mutual funds. It also highlights the dimension that matters the most for an investor who has a genuine ethical objective. This new approach also extends the empirical literature on the performance of socially responsible mutual funds. Since the emergence of SRI, empirical studies have sought to estimate whether there is a cost of being ethical and, if so, to evaluate it. The results are mixed, and the literature has hardly converged toward a consensus of assessing whether SRI cannot be achieved without a cost to financial performance. As this debate finds its roots in the categorization and the various methodological implementations employed, this paper proposes a new framework to investigate the *de jure* and *de facto* identity of mutual funds. Is there a difference between *what is said* and *what is done*? If that is the case, it would partially explain the divergences in the cost of ethics in the literature. From this new identification process, we empirically assess the impact of *de jure* and *de facto* SRI on financial performance. To do so, we build a new database, which is exploited using a novel econometric methodology to avoid (i) a matching procedure between differently categorized funds and (ii) a two-step approach and its potential statistical biases.

Specifically, whereas earlier studies propose simple comparisons of the risk-adjusted measures, more recent studies consider factor-augmented models à la Fama and French (1993) or Carhart (1997). This paper implements both approaches, free from any assumptions regarding the categorization or list. First, it offers a comparison of the returns among different categories of socially responsible funds without any matching procedure between differently categorized funds. It also follows the second route, considering the most recent papers (Bauer et al., 2005, Renneboog et al., 2008; El Ghoul and Karoui, 2017) and the 4-factor model à la Carhart (1997). Nonetheless, to avoid a two-step approach and its potential statistical biases, the 4-factor model is estimated via a non-linear panel data model to allow for the inclusion of the two dimensions of SRI.

This paper enriches the existing literature in two ways. First, the empirical analysis includes two dimensions of extra-financial information: (i) mutual fund's names or labels (i.e. *what is said* about SRI) (ii) mutual funds' holdings (i.e. *what is done* about SRI). Estimations begin with static ratings, whereas a robustness check is proposed that includes dynamic ratings. Furthermore, we verify that our results are robust to different time samples and that ESG ratings are stable over time. Second, our study relies on *SRI* and

ESG scores from the Morningstar Sustainability Rating the MSCI ESG Fund Metrics databases. Indeed, Morningstar is the provider used most frequently by investors, and its ratings have a significant influence on investors' behavior (Armstrong et al., 2017, Hartzmark and Sussman, 2019). Berg et al. (2019) show that ESG ratings from different providers are only weakly correlated: they differ in rating methodology and metrics. Thus, we conduct our study using both databases, focusing on large institutional mutual funds (conventionally defined as having over 100 million euros in assets under management) to avoid any selection bias. Hence, we empirically verify that our results are robust to different ESG ratings.

The empirical section covers both US and European domestic equity mutual funds. To this end, we build a new database of mutual fund returns that is robust to survival bias due to disappearing funds (right-censored) and the "newborn" bias (left-censored) generated by newly created funds. Hence, we obtain a novel database that includes approximately 600 European and 900 US domestic equity funds over the period $2013 - 2018$. The purpose of such a new database is to be able to exploit scores provided by Morningstar Sustainability Rating and those provided by MSCI ESG Fund Metrics for every mutual fund. In addition to this *de facto* SRI measure, we also introduce a dummy variable as an indicator that describes the identity of every mutual fund in the database. This *de jure* SRI measure is built from asset managers' branding (self-presentation and label accreditation).

Based on our results, we observe that in Europe and the US, SRI features cannot be exclusively summarized by asset managers' branding or labels. It appears that a group of *de jure* socially responsible mutual funds exhibits low *SRI scores*, even though they have committed to be managed according to ESG criteria. Moreover, some conventional mutual funds exhibit higher *SRI scores* than some *de jure* socially responsible funds. Such a result clearly questions the pertinence of brand management processes in the industry and thus calls for a harmonization of ESG measures and labels such that this objective becomes directly applicable for investors. In a further step, a performance analysis using a panel version of Carhart's 4-factor model indicates that SRI branding do not truly matter for the funds' financial performance. The exclusive ethical driver is the SRI and ESG scores, i.e., funds' holdings or *de facto* SRI.

These results have several policy implications. First, asset managers' SRI branding (*de jure* SRI) do not reflect their funds' ESG performance (*de facto* SRI). Then, only the use of ESG criteria as investment constraints (*de facto* SRI) has an impact on mutual funds' financial performance. An information asymmetry emerges in the SRI market: investors may have difficulty identifying the true nature of mutual funds (conventional or socially responsible) and evaluating their financial performance conditional on their ESG performance. Thus, regulators should adequately monitor the coherence between the asset managers' branding and their portfolios' holdings to reduce information asymmetry on the SRI market. An appropriate regulation is a *sine qua non* condition to improve market efficiency in this segment (i.e., to enable a better capital allocation between conventional and socially responsible mutual funds). This finding might also explain the divergent results regarding the cost of ethics reported in the literature, as some studies consider measures of *de jure* SRI to be *de facto SRI scores*, whereas others include only realized *SRI scores*.

The paper is organized as follows. Section 1.2 describes the SRI databases used in the paper: the Morningstar Sustainability Rating, the MSCI ESG Fund Metrics, the mutual funds' names and labels and their basic historical statistics. In addition, we introduce a novel homogeneous mutual fund returns database for the US and Europe, corrected for survivor and "newborn" biases. Section 1.3 evaluates the financial performance of mutual funds according to corresponding SRI objectives. Section 1.4 is devoted to the robustness checks. Section 1.5 concludes the paper.

1.2 Database

1.2.1 *de facto* SRI : using the Morningstar and MSCI databases

To conduct a proper empirical analysis of SRI, the choice of database is crucial. As ethical standards differ across investors, asset managers and labeling organizations, the categorization of conventional and socially responsible mutual funds is highly debatable. In a recent paper, Statman and Glushkov (2016) highlight this difficulty in describing the differences between the databases used in the literature (e.g., Lipper's list, the Social Investment Forum (SIF) list and the Standard & Poor's ($S\&P$) list) and their consequences for empirical studies. In this paper, we choose to use two newly available databases from Morningstar and MSCI that score mutual funds with respect to the ethical quality of their investment holdings. These new databases[4] present two advantages. First, more than 90% of existing mutual funds are rated, whether they are identified as socially responsible or not. In addition, SRI and ESG scores cover more than 90% of mutual funds' exposure. Second, both scores of each mutual fund are composites built from the aggregation of firm-level ratings and normalized. Although they share many common features, Morningstar and MSCI databases exhibit a few, but interesting, differences. First, Morningstar and MSCI provide comparable ESG scores, but in addition, Morningstar provides a *SRI score* based on an ESG score and a Controversy score. Second, Morningstar Sustainability Rating's scores are free of industry bias, which is not the case for the ESG score from MSCI ESG Fund Metrics.

We restrict our initial analysis to European and US domestic equity mutual funds and, more precisely, the large-cap funds. The database covers the period $2013 - 2018$ at a monthly frequency. Several steps are implemented to build a balanced and consistent database. We require each mutual fund to have the same geographical investment area and the same currency to avoid associated risks. To be more precise, for the European mutual funds, we study the funds whose investment zone covers all of Europe and trades in euros. Our final database thus contains 606 funds in Europe and 887 funds in the United States. Each fund has a monthly *SRI score* that incorporates the ESG (environmental, social, and governance) score and the Controversy score.[5] This database is thus balanced and homogeneous and lists all European and US mutual funds with an SRI rating provided by Morningstar. To complete our analysis and be prepared for the robustness check, we extract from MSCI[6] the ESG score for the whole mutual funds sample.

[4]The use of the Morningstar Sustainability Rating and MSCI ESG Fund Metrics databases is a novelty in the literature. To the best of our knowledge, only Hartzmark and Sussman (2019) use fund-level data from Morningstar, showing that investors widely refer to Morningstar Sustainability Rating. However, they are working with pre-categorization SRI ratings (called ”globes”), whereas we instead consider continuous ratings (underlying these ”globes”). Our choice is motivated by the desire of avoiding potential non-linear effects in the model. If MSCI ESG Research and MSCI ESG KLD STATS are widely used firm-level databases in the literature (e.g., El Ghoul and Karoui, 2017), the introduction and use of MSCI ESG Fund Metrics is an innovation. Compared to other data providers, Morningstar and MSCI are the only ones that provide fund-level and historical SRI and ESG scores.

[5]Appendix 1 offers a description of the rating methodology developed by Morningstar.

[6]Appendix 2 offers a description of the rating methodology developed by MSCI.

1.2.2 The Survivor bias

Since the early 1990s, a large body of literature has measured the importance of survivor bias for mutual fund performance (Grinblatt and Titman, 1989; Brown et al., 1992; Brown and Goetzmann, 1994). This bias occurs with the disappearance of many mutual funds from the market, simply closing or merging with other funds because of weak or poor performance. Neglecting such bias would lead to an overestimation of the funds' performance. To the best of our knowledge, the CRSP database is the only existing survivorship-corrected and updated mutual fund returns database, but it exclusively focuses on the US mutual funds market.[7] As the objective of the paper consists of studying and comparing both the American and European markets, it requires building a comparable database for Europe and the US, particularly with respect to the treatment of survivor bias.[8] First, we track every fund existing during our sample period, as in Brown and Goetzmann (1994), Carhart (1994; 1997) and Malkiel (1995). Then, following Elton et al. (1996), we use the risk-adjusted returns and perform a 4-factor CAPM (Carhart, 1997) using a single-index model. However, our approach differs from previous studies, as we complete missing returns not only at the end of the sample period but also for the missing returns of mutual funds at the beginning of the sample. Indeed, we take into account newly born funds as soon as they exist for at least two years. This decision is motivated by the fact that we will consider a balanced panel framework, and thus, we cannot afford to have missing returns at the end or at the beginning of the sample. In addition, completely excluding these "newborn" funds would have reinforced the issue of selection bias.

1.2.3 *de jure* SRI : a new list from mutual funds' marketing: names and labels

The former database provides us with normalized and continuous SRI and ESG scores from two alternative sources: Morningstar Sustainability Rating and MSCI ESG Fund Metrics. In addition, this rating is attributed ex post, i.e., once the investments are realized. In

[7]The CRSP database was initiated by Carhart (1997). It is now updated by the Center for Research in Security Prices – The University of Chicago Booth School of Business.

[8]See Hanke et al. (2018) on the causes of the non-comparability of different databases such as CRSP and Morningstar.

other words, asset managers can explicitly make commitments in favor of SRI investments or even obtain SRI label accreditation. This is in line with investors' behavior: as shown by Bauer and Smeets (2015) or Riedl and Smeets (2017), investing according to ESG criteria allows investors to be identified as socially responsible investors. This can take the form of advertising action and is a clear decision of the fund to refuse to invest in sectors that are related to unethical industry. Thus, another classification can be made between funds that have made such commitments, which are labelled *de jure* socially responsible funds, and others that are considered conventional funds. To discriminate between *de jure* SRI and conventional mutual funds, we build a dummy variable to determine whether asset managers claim that their portfolio follows ESG criteria. To this end, we focus on mutual funds' names, looking for advertised extra-financial objectives. Mutual funds' names are an excellent proxy for advertising. Cooper et al. (2005) investigate the impact of changes in mutual funds' names driven by hot investment styles. They empirically show that investors are irrationally influenced by cosmetic effects. Similarly, Espenlaub et al. (2017) conduct a natural experiment and empirically show that mutual funds' names have a significant impact on investors. Using a regulation change (SEC Rule 35d-1) in July 2001, the authors highlight that superficial changes in mutual funds' names led to a significant increase in capital flows.

We follow Nofsinger and Varma's (2014) methodology to classify conventional and socially responsible mutual funds. From a discrete selection process, they build a list of words related to SRI terminology: "social", "socially", "environment", "green", "sustainability", "sustainable", "ethics", "ethical", "faith", "religion", "Christian", "Islam", "Baptist" and "Lutheran". Then, using the dictionary defined above, the authors keyword-search mutual funds' names to identify socially responsible mutual funds. We go further than Nofsinger and Varma (2014), using several lexical databases to broaden the SRI terminology. Our purpose is to build a more complete dictionary that enables us to search for words (nouns, adjectives or verbs) associated with SRI. First, we store a preliminary list of words from the terminology identified by the USSIF (2018):[9] "community", "ethical", "green", "impact", "mission", "responsible", "socially", "sustainable", and "values". Then, we extend this initial list using the lexical database developed by Miller (1995) and Fellbaum (1998) and

[9]Terminology available at `https://www.ussif.org/sribasics`

hosted/updated by Princeton University.[10] To the 9 initial words from USSIF (2018)'s terminology, we add 18 additional words, which we present in Table 1.1. Our extended dictionary is then used to classify mutual funds via pattern search on mutual funds names. The rationale for using pattern search instead of keyword search as in Nofsinger and Varma (2014) is to track words (nouns, adjectives and verbs) based on the same stem as keywords from our dictionary. The indicator variable is then built from a matching procedure, and we check the results using Bloomberg's description of mutual funds. In case of doubt about a given European mutual fund, the classification is triple-checked using the mutual fund's key investor information document (KIID).

Our extended dictionary provides relevant results because it enables a better identification of *de jure* socially responsible mutual funds. Our dictionary is able to accurately discriminate between conventional and socially responsible mutual funds. The resulting classification is double-checked using Bloomberg's description of mutual funds (description texts are provided by asset managers) and mutual funds' prospectuses. Moreover, our dictionary outperforms Nofsinger and Varma (2014)'s dictionary, which fails to classify every SRI-labeled mutual fund as SRI.

[10] See Princeton University "About WordNet." WordNet. Princeton University. 2010.

Table 1.1: SRI Terminology

Nofsinger and Varma (2014)	Extended dictionary
Baptist	Baptist
Christian	blue
environment	carbon
ethical	Catholic
ethics	Christian
faith	climate
green	community
Islam	durable
Lutheran	environment
religion	ESG
Social	ethical
socially	faith
sustainable	governance
sustainability	green
	human rights
	impact
	Islam
	Lutheran
	mission
	moral
	peace
	philosophy
	religion
	responsible
	social
	solidarity
	subsidiarity
	sustainable
	sustainability
	values

1.2.4 Beyond the SRI classification

Table 1.2 reports the number of conventional and ethical funds as well as the descriptive statistics (their *de facto SRI score* mean and their standard deviation).

Table 1.2: Europe/US - Descriptive Statistics - Conventional vs *de jure* socially responsible mutual funds

Europe

	Conventional funds	*de jure* SRI funds	Total
Number	554	52	606
$Mean_{SRIscore}$	55.18	57.61	55.39
$\sigma_{SRIscore}$	1.996	1.953	2.103

The United States

	Conventional funds	*de jure* SRI funds	Total
Number	862	25	887
$Mean_{SRIscore}$	45.85	48.19	45.92
$\sigma_{SRIscore}$	1.725	1.500	1.761

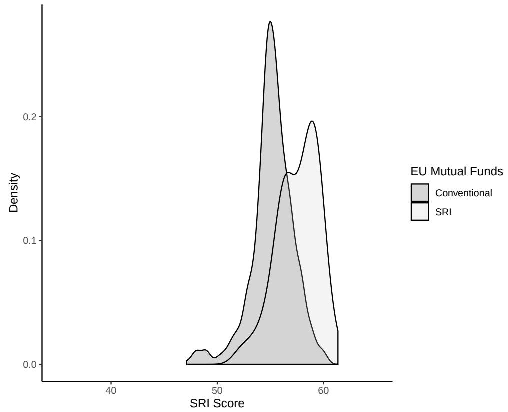
Notes: This table reports the number of funds included in our database and the corresponding SRI rating means and standard errors.

From the 606 European funds (resp. 887 US funds), we detect 52 funds (resp. 25 funds) presenting themselves as socially responsible funds, resulting in 554 conventional funds for Europe (resp. 862 conventional funds for the US). We also investigate when this classification is robust to SRI labels. In comparing our classification for European funds with that proposed by Novethic,[11] the largest European ethical label provider, we find that all 19 mutual funds possessing a label are classified as ethical, whereas none of the conventional funds present such a feature. This classification can thus be interpreted as either an ethical objective or an ethical label. In the case of the US, such a comparison is not possible because, to the best of our knowledge, it does not have ethical labels. It also appears that socially responsible funds represent a minority (approximately 9% in Europe and 3% in the United States), whereas conventional funds are numerous. However, it is striking to observe that on average, the corresponding SRI rates are almost the same (and

[11] http://www.novethic.fr/labellisation-de-linvestissement-responsable.html

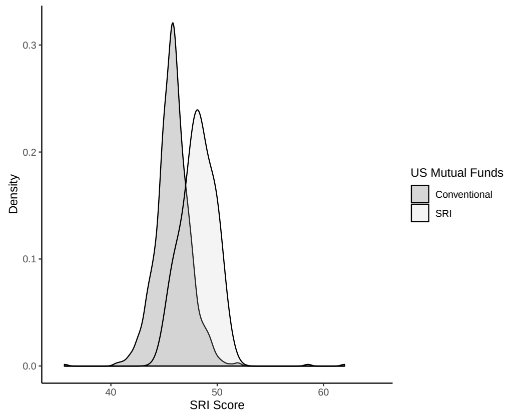
not significantly different) across ethical and conventional funds. This suggests that some SR funds have lower *SRI scores* than some conventional funds, whereas some conventional funds outperform socially responsible funds. This reveals a difference in ethical investments between what is announced and what is realized by such funds.

Figure 1.1: Density functions of mutual funds' SRI score: Europe



Notes: This figure plots the distribution of SRI scores of European mutual funds. Scores proxy the *de facto* SRI. The classification "Conventional" / "SRI" represents *de jure* SRI. Distributions of the ESG scores from Morningstar and MSCI exhibit the same features (figure available upon request).

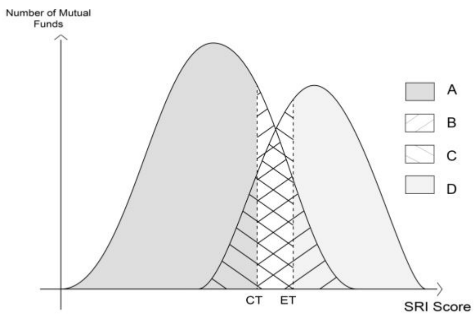
Figure 1.2: Density functions of mutual funds' SRI score: the United States



Notes: This figure plots the distribution of SRI scores of US mutual funds. Scores proxy the *de facto* SRI. The classification "Conventional" / "SRI" represents *de jure* SRI. Distributions of the ESG scores from Morningstar and MSCI exhibit the same features (figure available upon request).

Figure 1.1 and Figure 1.2 plot the distribution of SRI ratings for these two groups of funds in Europe and in the US. It is obvious that the peak of the distribution for socially responsible mutual funds is higher than that calculated for conventional funds. However, it also reveals the presence of a very large overlap between the distributions, thus confirming that some socially responsible funds have a poor *SRI score*, whereas some conventional funds have a relatively good *SRI score*. Such a stylized fact paves the way for a second dimension of ethics that addresses the realization of SRI investments that might differ from the SRI branding of the funds. To implement such a distinction in a preliminary analysis, it is necessary to find a threshold (CT) above which a conventional fund is managed ethically and a threshold (ET) below which a socially responsible mutual fund invests conventionally. To this end, we consider a simple rule that is similar to the conditional value-at-risk (CoVaR) measure. The threshold CT is defined as the *SRI score* given to the lowest 10% of socially responsible funds, and ET is defined as the *SRI score* given to the highest 10% of conventional funds. Figure 1.3 illustrates this definition.

Figure 1.3: Scheme of the SRI classification from the density function of the SRI score



Notes: This figure is an illustrative scheme about the 2-dimensional measure of SRI. The left-hand distribution corresponds to the conventional mutual funds (category A ∪ B), and the right-hand distribution corresponds to the *de jure* socially responsible mutual funds (category C ∪ D). Specifically, we define two thresholds (CT and ET) to discriminate subcategories (A, B, C and D). Then, mutual funds on the left of the threshold CT (subcategory A) corresponds to conventional mutual funds with low score, whereas mutual funds on the right of the threshold CT (subcategory B) corresponds to conventional mutual funds with high score. Similarly, *de jure* socially responsible mutual funds on the left of the threshold ET (subcategory C) corresponds to *de jure* socially responsible mutual funds with low score, whereas mutual funds on the right of the threshold ET (subcategory D) corresponds to *de jure* socially responsible mutual funds with high score.

1.2.5 Preliminary Analysis

For Europe, we find that $CT = 55.204$ and $ET = 57.598$, and for the US, we find that $CT = 46.056$ and $ET = 47.720$. Conditional on these threshold values, Tables 1.3 and 1.4 summarize our categorization of the funds according to their SRI objectives and realizations. For simplicity, we label them from A to D, as reported in Table 1.2.

Table 1.3: Categorization of funds according to *de jure* / *de facto* SRI

Europe

	Conventional Real.	Ethical Real.
Conventional Obj.	A n=276 (49.82%)	B n=278 (50.18%)
Ethical Obj.	C n=23 (44.23%)	D n=29 (55.77%)

The United States

	Conventional Real.	Ethical Real.
Conventional Obj.	A n=510 (59.16%)	B n=352 (40.84%)
Ethical Obj.	C n=9 (36.00%)	D n=16 (64.00%)

Notes: This table reports the categorization of mutual funds. The double entry table classifies funds relatively to their objectives (*de jure*) and to their realized investments (*de facto*).

We find that 50.18% (resp. 40.84%) of the European (resp. US) conventional funds still present very high *SRI scores*, indicating that they respect ethical principles. On the contrary, and perhaps more interestingly, 44.23% (resp. 36.00%) of the European (resp. US) ethical funds have a low ethical grade. This result indicates that 23 (resp. 9) ethical funds do not respect their commitments in terms of SRI investments. What is said does not seem to match what is done in terms of SRI. When looking specifically at the European mutual funds having a label, they are all classified in the C and D categories. Labels are coherent with announcements. Still, 4 of them belong to group C, highlighting that labels are only weak leading indicators of the effective respect given to ethical objectives. This industry should thus depart from considering ethical as a marketing positioning and instead integrate the effective realization of socially responsible investments. In addition, when considering the management fees of the mutual funds, we do not observe significant differences across categories. Instead, mutual funds with a label present significantly higher management fees, [12] suggesting that they charge investors for the ethical label.

[12] SRI labeled mutual funds' fees are on average 39.78% higher than the other *de jure* socially responsible mutual funds.

1.3 Evaluating the performance of socially responsible mutual funds

1.3.1 Literature review of socially responsible fund performance

The literature on the financial impacts of SRI focuses on the dichotomy between conventional and SRI-labelled mutual funds. The early studies by Hamilton et al. (1993), Goldreyer and Diltz (1999) and Statman (2000) compare risk-adjusted returns of SRI indices and mutual funds in the US in the 1990s and show that the impacts of ethical investment are not significantly different from zero. Luther and Matatko (1992; 1994) and Mallin et al. (1995) run the same experiments for the UK trusts for $1984 - 1990$ and $1986 - 1994$, respectively. They find similar results, supporting the idea that ethical screening could introduce bias both about the size of firms and the reference benchmark choice. From this conclusion, Goldreyer et al. (1999) move from the mean-variance framework to the CAPM model of Sharpe (1964) and Lintner (1965), as Kreander et al. (2005) and Renneboog et al. (2008a; 2008b) use the three-factor model of Fama and French (1992; 1993) to evaluate the performance of European and US mutual funds. Their results reach the same conclusions, namely, financial performance tends to be lower when investment is ethical, but their results are hardly significant. The introduction of the four-factor model of Carhart (1997) by Bauer et al. (2005), who study the financial performance of US, UK and German socially responsible mutual funds, does not contrast the significance of the performance differences observed in previous studies. Statman and Glushkov (2009) and Hong and Kacperczyk (2009) choose to consider “sin stocks” in investment portfolios. A positive and significant excess performances of such funds would support the idea that ethics has a price. Nofsinger and Varma (2014) and Petitjean (2019) investigate financial performance of conventional and ethical investments during crisis times. Their purpose is to compare tail risks in both investment strategies but their results diverge. These seemingly different results obtained in the empirical literature highlight a methodological issue regarding the dual approach of conventional versus ethical investment. Consequently, the most recent empirical literature focuses on a lack of strict methodology for comparing performance. Indeed, early studies on the impact of ethics (Hamilton et al., 1993; Bauer et al., 2005) are based on a direct comparison between conventional and socially responsible mutual funds but do not assess the relevance of a comparison between such mutual funds. Since then, this methodology has been enriched following two major approaches. On the one hand, Statman (2000, 2006) and Schröder (2007) use financial indices rather than mutual funds to compare the financial performances of ethical and conventional benchmarks. On the other hand, a preliminary matching procedure among conventional, ethical (Mallin et al., 1995; Gregory et al., 1997; Kreander et al., 2005) and even "sin" (Hong and Kacperczyk, 2009; Humphrey and Tan, 2014; Borgers et al., 2015) mutual funds is introduced to improve the relevance of quantitative performance comparisons. In the most recent papers, the use of cross scores based on ESG criteria among data covering both conventional and ethical assets allows researchers to avoid statistical biases inherent to a binary classification of mutual funds. El Ghoul and Karoui (2017) build an asset-weighted CSR score for mutual funds based on their exposures by using firm-level data from MSCI ESG KLD STATS. This preliminary data treatment allows the authors to compare US domestic mutual fund performance on the same basis over the $2003-2011$ period. Borgers et al. (2015) use the same approach with US mutual fund holdings from 2004 to 2012 to measure the impact of social factors on financial performance.

1.3.2 A preliminary analysis of mutual fund performance

Given the funds' classification along the two dimensions of ethics, it is possible to run a first analysis of their performance, as in Hamilton et al. (1993), Goldreyer and Diltz (1999) and Statman (2000). Following this literature, we first study two clusters related to *de jure* SRI: conventional mutual funds (AB) and socially responsible mutual funds (CD). We then analyze the two other clusters related to *de facto* SRI considering realized *SRI Score*: low-ranked mutual funds (AC) and top-ranked mutual funds (BD). Table 1.4 reports the descriptive statistics (mean return, standard error and Sharpe ratio) for each of these fund categories. It appears that, in line with the existing empirical studies in the literature, conventional funds (AB) outperform socially responsible funds (CD) in both the US and Europe. Thus, *de jure* SRI seems to behave as added value for investors because socially responsible mutual funds exhibit higher returns and lower risks than conventional funds. However, the financial performance analysis conditional on realized investments (*de facto* SRI) leads to a different interpretation: low-ranked mutual funds (AC) exhibit better financial performance than high-ranked mutual funds (BD). Thus,*de* *facto* SRI has a financial cost: in line with theory, extra-financial constraints lead to less efficient portfolios.

Table 1.4: Descriptive statistics - Mutual funds performance - 2013-2018

Europe

	AB	CD	AC	BD
Fund Return (μ)	7.21%	7.92%	7.51%	7.04%
Fund Risk (σ)	11.72%	11.10%	11.72%	11.61%
Sharpe ($\frac{\mu}{\sigma}$)	0.615	0.713	0.641	0.606

The United States

	AB	CD	AC	BD
Fund Return (μ)	13.50%	13.59%	13.97%	12.85%
Fund Risk (σ)	10.51%	10.37%	10.67%	10.27%
Sharpe ($\frac{\mu}{\sigma}$)	1.285	1.310	1.309	1.251

Notes: This table reports the annualized average returns of different types of funds. It also reports corresponding standard deviations and Sharpe ratios.

This preliminary analysis (Table 1.4) illustrates the divergence in the literature between SRI and financial performance. In the next section, we deepen the analysis using a more sophisticated econometric framework.

1.3.3 Methodology

To offer a more extensive comparison of socially responsible mutual funds, we elaborate on the CAPM approach. Recent papers have typically estimated factor models, such as the traditional Fama-French (1993) model, which integrates 3 or more factors. A recently developed 4-factor risk-adjusted performance model has also been proposed by Carhart (1997). This model considers a market return index (r^m), the monthly premium of the book-to-market factor (r^{HML}), the monthly premium of the size factor (r^{SMB}) and momentum in stock markets (r^{MOM}). The model can thus be rewritten as

$$r_{i,t} - r_t^f = \alpha + \beta_{r^m}.(r_t^m - r_t^f) + \beta_{SMB}.r_t^{SMB} + \beta_{HML}.r_t^{HML} + \beta_{MOM}.r_t^{MOM} + \eta_i + \epsilon_{i,t}, \quad (1.1)$$

where r_i is the fund's i return, r^f is the monthly risk-free rate, and α is the net-of-fees annual risk-adjusted performance of fund i. To avoid a two-step approach, which can introduce statistical bias, we rely on recent studies (Petersen, 2009; Ando and Bai, 2015) that propose evaluating the performance of funds using a large-dimensional panel, i.e., considering in a single step both the time and the cross-sectional dimension. To this aim, a fixed effect factor η_i is added in order to take into account the potential unobserved heterogeneity. The model is estimated independently for each fund i such that $\epsilon_{i,t}$ has i.i.d. white noise. The model can be estimated for a period of time and for a set of funds i (cross-sectional dimension) or for a particular fund i for a period of time $t = 1, ..., T$ (time series dimension). El Ghoul and Karoui (2017) apply the latter strategy. In the first step, they estimate for each fund individually (1) to obtain an individual estimate of α the conditional return of the funds. The βs that represent the sensitivity to market factors remain a common factor for all funds. They introduce the SRI characteristics of the fund in the second step, in which they regress the estimated conditional return on the particular features of the fund.

Then, to distinguish between conventional and socially responsible funds, a non-linear panel is considered. More precisely, we split the panel for these two types of funds and obtain the following model:

$$\begin{aligned} r_{i,t} - r_t^f = \mathbb{1}_{CONV}(\alpha_c + \beta_{c,r^m}.(r_t^m - r_t^f) + \beta_{c,SMB}.r_t^{SMB} + \beta_{c,HML}.r_t^{HML} + \beta_{c,MOM}.r_t^{MOM}) \\ + \mathbb{1}_{SRI}(\alpha_s + \beta_{s,r^m}.(r_t^m - r_t^f) + \beta_{s,SMB}.r_t^{SMB} + \beta_{s,HML}.r_t^{HML} + \beta_{s,MOM}.r_t^{MOM}) + \eta_i + \epsilon_{i,t}, \end{aligned} \tag{1.2}$$

where $\mathbb{1}_{CONV}(.)$ is an index variable that takes a value of 1 if the fund is conventional and 0 otherwise and $\mathbb{1}_{SRI}(.)$ is an index that takes a value of 1 if the fund is ethical and 0 otherwise. The subscript C refers to estimates associated with conventional funds, and the subscript s corresponds to socially responsible fund coefficients. Model (2) is estimated via GLM, and a Driscoll-Kraay (1998) correction is implemented to avoid bias due to cross-sectional dependence. It is thus possible to test whether a category of funds offers extra returns and whether it is more sensitive to a peculiar factor. In a sense, this approach is in line with papers that test for homogeneous breaks in slopes, such as those by Pesaran and Yamagata (2008) or Blomquist and Westerlund (2013). Here, the breaks are exogenous and are driven by economic motivations: socially responsible and conventional funds.

Furthermore, we follow Hansen (2000), and instead of considering model (2), we estimate a model integrating both the whole sample and the ethical funds subsample. The non-linear panel Carhart model takes the following form:

$$r_{i,t} - r_t^f = \alpha + \beta_{r^m}.(r_t^m - r_t^f) + \beta_{SMB}.r_t^{SMB} + \beta_{HML}.r_t^{HML} + \beta_{MOM}.r_t^{MOM}$$
$$+\mathbb{1}_{SRI}(\tilde{\alpha}_s + \tilde{\beta}_{s,r^m}.(r_t^m - r_t^f) + \tilde{\beta}_{s,SMB}.r_t^{SMB} + \tilde{\beta}_{s,HML}.r_t^{HML} + \tilde{\beta}_{s,MOM}.r_t^{MOM}) + \eta_i + \epsilon_{i,t}. \tag{1.3}$$

This representation offers more precise estimates and straightforward interpretations. If a coefficient associated with a socially responsible mutual fund (denoted with an underscore s) is significant, then it would indicate a particular behavior of socially responsible funds. In the opposite case, it would suggest that they behave similarly to conventional funds. The model thus separates ethical from conventional funds. However, as we stress in the previous section, some funds do not respect their commitments. Furthermore, some socially responsible funds present a low *SRI score*, and some conventional funds present a high *SRI score*. We thus consider in model (4) the *SRI score* obtained from the final non-linear panel-augmented Carhart model:

$$r_{i,t} - r_t^f = \alpha^* + \beta_{r^m}.(r_t^m - r_t^f) + \beta_{SMB}.r_t^{SMB} + \beta_{HML}.r_t^{HML} + \beta_{MOM}.r_t^{MOM}$$
$$+ \mathbb{1}_{SRI}(\tilde{\alpha}_s + \tilde{\beta}_{s,r^m}.(r_t^m - r_t^f) + \tilde{\beta}_{s,SMB}.r_t^{SMB} + \tilde{\beta}_{s,HML}.r_t^{HML} + \tilde{\beta}_{s,MOM}.r_t^{MOM})$$
$$+ \alpha_{SRI}.SRI_i + \eta_i + \epsilon_{i,t}. \tag{1.4}$$

Let us note that in such a specification, the fixed term effect is omitted, as it would be highly correlated with the *SRI score* if it is fixed over the period or presents a low variability. In this section, *SRI score* is fixed over the given time period 2018. In the robustness sub-section 1.4, time-varying *SRI score* is considered. In model (4) the estimated return of the mutual fund $\hat{\alpha}^*$ is now calculated as $\hat{\alpha} - \hat{\alpha}'_{SRI}$.

1.3.4 Empirical application

Through a preliminary analysis and to identify a benchmark, we estimate the basic linear Carhart model (1) for $2013 - 2018$ without considering any SRI dimensions. The market benchmark (r^m) is the MSCI USA Index and MSCI Europe Index for the US and European mutual funds, respectively, the risk-free rate (r^f) is the US and EU short-term interest rates, respectively, and returns (r) are net of fees. Table 1.5 reports the model (1) estimates.

It appears that almost all explanatory variables except momentum (and HML for Europe) have a significant effect on the risk-adjusted performance of the funds. The $Adjusted - R^2$ value is also quite high (0.76 for Europe and 0.84 for the US). The alpha coefficient is not significant. Such a result is consistent with the efficient market hypothesis. Specifically, the results show that the market factor has almost a proportional impact on the funds' returns. SMB also affects returns positively but with less elasticity. By contrast, the value premium is negative but with a relatively small coefficient.[13]

Table 1.5: Estimation of the panel version of Carhart's model (2013-2018)

	Europe		US	
	Estimates	*s.e.*	Estimates	*s.e.*
$\hat{\alpha}$	-0.0018	0.0135	-0.0017	0.0011
$\hat{\beta}_{r^m}$	0.9803***	0.0185	1.0068***	0.0084
$\hat{\beta}_{SMB}$	0.1996***	0.0465	0.0711***	0.0158
$\hat{\beta}_{HML}$	-0.0268	0.0215	-0.0470**	0.0210
$\hat{\beta}_{MOM}$	0.0031	0.0239	0.0207	0.0217
$Adj.R^2$	0.7599		0.8466	

	Europe	US
Fixed-Effects:		
Fund	Yes	Yes
Time	Yes	Yes
Obs	40,602	59,429

Notes: This table reports estimates of the augmented non-linear panel version of Carhart's model (Eq. 1.1) based on the GLM method. The Driscoll and Kraay (1998) correction is applied such that standard errors are robust to heteroscedasticity and autocorrelation. The notations ***, ** and * indicate that the null hypothesis of a zero coefficient is statistically rejected at 99%, 95% and 90%.

[13] In his famous survey, Schwert (2003) confirms this fact, concluding that small-firm anomalies have almost disappeared in the most recent period.

The significance of annual risk-adjusted performance ($\hat{\alpha}$) is not significantly different from zero, corroborating the results of Fama and French (1993) and Carhart (1997), indicating that on average, mutual funds do not exhibit extra returns. Table 1.6 gathers the results of the estimation of the 4-factor model augmented by a *de jure* dummy indicating whether the fund has committed to investing ethically. It appears that this dummy is not significantly different from 0, supporting the literature's findings (Bauer et al., 2005; Renneboog et al., 2008b). Indeed, SRI branding do not affect the performance of the fund. The same estimation is performed using an SRI label dummy instead of the name-based variable. It leads to the same conclusion (see Appendix 3).

Table 1.6: Estimation of the panel version of Carhart's model (2013-2018) with dummy *de jure*

	Europe		US	
	Estimates	*s.e.*	Estimates	*s.e.*
$\hat{\alpha}$	-0.0018	0.0069	-0.0017	0.0010
$\hat{\beta}_{rm}$	0.9803***	0.0185	1.0068***	0.0084
$\hat{\beta}_{SMB}$	0.1996***	0.0465	0.0711***	0.0158
$\hat{\beta}_{HML}$	-0.0268	0.0215	-0.0470**	0.0210
$\hat{\beta}_{MOM}$	0.0031	0.0239	0.0207	0.0217
$\hat{D}ummyDeJure$	0.0011	0.0073	0.0041	0.0020
$Adj.R^2$	0.7545		0.8464	
Fixed-Effects:				
Fund	Yes		Yes	
Time	Yes		Yes	
Obs	40,602		59,429	

Notes: This table reports estimates of the augmented non-linear panel version of Carhart's model (Eq. 1.1) based on the GLM method with an extra *de jure* dummy. The Driscoll and Kraay (1998) correction is applied such that standard errors are robust to heteroscedasticity and autocorrelation. The notations ***, ** and * indicate that the null hypothesis of a zero coefficient is statistically rejected at 99%, 95% and 90%.

In the next step, the non-linear augmented Carhart model represented by equation (1.4) can now be estimated. Remember that $\mathbb{1}(.)_{SRI}$ is an indicator function that takes a value of 1 (resp. 0) if the fund is ethical (resp. conventional); i.e., there SRI branding (resp. there is no clear objective in favor of SRI). This corresponds to the *de jure* dummy variable. The second variable, *SRI Score*, corresponds to the grade given by Morningstar. This variable is a proxy for the *de facto* dimension of the SRI because it is independent of any branding. The model is estimated by the generalized linear method (GLM), and the results are reported in Table 1.7.

Table 1.7: Estimation of the augmented non-linear panel version of Carhart's model (2013-2018)

	Europe		US	
	Estimates	*s.e.*	Estimates	*s.e.*
		Full Sample		
$\hat{\alpha^*}$	0.0088**	0.0035	0.0094***	0.0028
$\hat{\beta}_{r^m}$	0.9638***	0.0178	1.0218***	0.0141
$\hat{\beta}_{SMB}$	0.1136***	0.0312	0.0896***	0.0259
$\hat{\beta}_{HML}$	-0.0953***	0.0315	-0.0361	0.0271
$\hat{\beta}_{MOM}$	-0.0192	0.0145	-0.0006	0.0140
		de jure Socially Responsible Mutual Funds		
$\hat{\hat{\alpha}}_s$	-0.0013***	0.0003	-0.0004*	0.0003
$\hat{\hat{\beta}}_{s,r^m}$	0.0181	0.0175	-0.0154	0.0143
$\hat{\hat{\beta}}_{s,SMB}$	0.0940***	0.0336	-0.0190	0.0261
$\hat{\hat{\beta}}_{s,HML}$	0.0750**	0.0345	-0.0112	0.0283
$\hat{\hat{\beta}}_{s,MOM}$	0.0244	0.0188	0.0219*	0.0116
		de facto SRI Score		
$\hat{\alpha}_{SRI}$	-0.0118**	0.0000	-0.0170***	0.0000
Adj.R^2	0.7528		0.8418	
Fixed-Effects:				
Fund	Yes		Yes	
Time	Yes		Yes	
Obs	40,602		59,429	

Notes: This table reports estimates of the augmented non-linear panel version of Carhart's model (Eq. 1.4) based on the GLM method. The Driscoll and Kraay (1998) correction is applied such that standard errors are robust to heteroscedasticity and autocorrelation. The notations ***, ** and * indicate that the null hypothesis of a zero coefficient is statistically rejected at 99%, 95% and 90%.

The estimates obtained in the first part of the model (i.e., for the full model) are similar in both sign and magnitude to those obtained using the previous linear model (Table 1.5) for Europe and the United States. The only slight difference is the increase in the value of α^* when *SRI Score* is introduced. Such a result can be explained by the negative value of the *de facto SRI score*, suggesting that extra-financial constraints constitute a penalty for funds' performance (Bollen, 2007, Fama and French, 2007). In addition, when calculating the overall α[14] in model (4), it turns out to be very close to 0.

The second part of Table 1.7 addresses the impact of SRI branding (proxied here by a regime characterized by the *de jure* dummy). The results between the US and

[14]The overall *alpha* ($\hat{\alpha}$) is the sum of the full-sample *alpha* ($\hat{\alpha}^*$) and the *alpha* specific to socially responsible funds ($\hat{\alpha}_{SRI}$) such that $\hat{\alpha} = \hat{\alpha}^* + \hat{\alpha}_{SRI}.\bar{SRI}$

Europe highlight several common features. First, the $\hat{\tilde{\beta}}_{s,r^m}$ is not significant, meaning that SRI branding have no significant impact on the market risk exposure of these funds. Second, neither the β_{MOM} factor nor the $\hat{\tilde{\beta}}_{s,MOM}$ for *de jure* socially responsible funds is significant, indicating the absence of persistence in the funds' returns. SRI thus does not impact the persistence of funds' returns. Finally, it appears that the *SRI score* has a negative and significant effect on the funds' returns, supporting the idea that *de facto* socially responsible mutual funds have a return penalty (Bollen, 2007; Fama and French, 2007).

However, we observe differences in funds' behavior. For the US, none of the *de jure* factors explain the funds' returns at a 95% confidence level, confirming that SRI branding have no impact on mutual funds' performance. In contrast, in Europe, $\tilde{\beta}_{s,SMB}$ and $\tilde{\beta}_{s,HML}$ are significant and positive. Thus, *de jure* SRI could have an indirect impact on financial performance if investing in small businesses or value firms is considered ethical. This is consistent with the fact that in Europe, asset managers tend to combine ethics with investing in small firms. This difference between de jure SMB and HML between the EU and the US comes from the fact that in the US, the names of mutual funds must reflect the "real" strategy of funds, which is not the case in Europe. Such an observation can also explain why the number of *de jure* socially responsible mutual funds is relatively low in Europe. Finally, a difference can be observed when comparing the $\hat{\tilde{\alpha}}_s$ (*de jure*) and the $\hat{\alpha}_{SRI}$ (*de facto*). We observe that *de jure* socially responsible funds present significantly lower average returns in Europe, whereas their performance in the US is identical to that of conventional funds.

Regardless of the region, it appears that the magnitude of the *de facto SRI score* is much higher than that of the *de jure* dummy variable. The *de facto* socially responsible mutual funds exhibit a return penalty in the EU and the US, as supported by theory (Bollen, 2007, Fama and French, 2007), and are thus much more important than SRI branding.

Investors who are genuinely interested in ethical investing should not base their choice on labels or SRI branding from the funds. Instead, they should focus exclusively on the "ex post" SRI scores. Finally, substantial concerns are raised about the mutual fund labels issued by specialized audit agencies, as they do not appear to be good leading indicators of real respect for ethical rules. Such a conclusion corroborates the findings regarding labels presented in the last section.

1.4 Robustness Checks

The following section presents a series of robustness checks. We begin by considering the stability of our results (i.e., by considering a smaller sample size). In a second experiment, we consider a time-varying SRI score. Then, the SRI scores are split into ESG and Controversy grades. Next, the robustness of our results is analyzed while restricting mutual funds to Euro-area funds. Finally, we categorize the funds according to size.

In a first exercise, a smaller sample score is considered for $2017-2018$, i.e., with 24 observations. Given the number of explanatory variables considered and the non-linear nature of our model (including 11 explanatory variables), this is the smallest sample to be considered without being subjected to severe finite-sample bias. Table 1.8 reports the estimation results.

These results are qualitatively equivalent to those obtained for the whole sample, $2013-2018$, and the main previous findings still hold. It can be noticed that socially responsible mutual funds do not perform significantly worse than conventional funds do in Europe and the United States. It also appears that *SRI scores* negatively affect the adjusted-risk returns of all mutual funds.

In a second experiment, we consider time-varying SRI ratings. Even though a quick analysis of the Morningstar database would reveal that the the SRI ratings do not vary much over time, we estimate the non-linear panel-augmented Carhart model with a time-varying *SRI score*, which thus can be expressed as:

$$\begin{aligned} r_{i,t} - r_t^f &= \alpha^* + \beta_{r^m}.(r_t^m - r_t^f) + \beta_{SMB}.r_t^{SMB} + \beta_{HML}.r_t^{HML} + \beta_{MOM}.r_t^{MOM} \\ &+ \mathbb{1}_{SRI}(\tilde{\alpha}_s + \tilde{\beta}_{s,r^m}.(r_t^m - r_t^f) + \tilde{\beta}_{s,SMB}.r_t^{SMB} + \tilde{\beta}_{s,HML}.r_t^{HML} + \tilde{\beta}_{s,MOM}.r_t^{MOM}) \\ &+ \alpha_{SRI}.SRI_{i,t} + \eta_i + \epsilon_{i,t}. \end{aligned} \quad (1.5)$$

Because of data availability, only seven monthly historical SRI ratings are available for the US. Table 1.9 reports the results obtained.

Table 1.8: Static (constant SRI ratings) estimation of the augmented non-linear panel version of Carhart's model (2017-2018)

	Europe		US	
	Estimates	*s.e.*	Estimates	*s.e.*
	Full Sample			
$\hat{\alpha}^*$	0.0129	0.0127	0.0266	0.0275
$\hat{\beta}_{r^m}$	0.8951***	0.0316	0.9625***	0.0121
$\hat{\beta}_{SMB}$	-0.0549	0.0607	-0.0243	0.0244
$\hat{\beta}_{HML}$	-0.1095**	0.0165	-0.0696	0.0759
$\hat{\beta}_{MOM}$	0.1086***	0.0368	-0.0152	0.0608
	de jure Socially Responsible Mutual Funds			
$\hat{\hat{\alpha}}_s$	-0.0007	0.0007	-0.0014	0.0011
$\hat{\hat{\beta}}_{s,r^m}$	0.0367	0.0240	0.0066	0.0088
$\hat{\hat{\beta}}_{s,SMB}$	0.1379***	0.0277	-0.0123	0.0119
$\hat{\hat{\beta}}_{s,HML}$	0.0727**	0.0287	0.0943	0.0720
$\hat{\hat{\beta}}_{s,MOM}$	0.0463*	0.0229	0.1012	0.0501
	de facto SRI Score			
$\hat{\alpha}_{SRI}$	-0.0213***	0.0001	-0.0524**	0.0001
Adj.R^2	0.7566		0.8381	
Fixed-Effects:				
Fund	Yes		Yes	
Time	Yes		Yes	
Obs	7,878		11,531	

Notes: This table reports estimates of the augmented non-linear panel version of Carhart's model (Eq. 1.4) derived from the GLM method. The Driscoll and Kraay (1998) correction is applied such that standard errors are robust to heteroscedasticity and autocorrelation. The notations ***, ** and * indicate that the null hypothesis of a zero coefficient is statistically rejected at 99%, 95% and 90%.

Table 1.9: Dynamic (time-varying SRI ratings) estimation of the augmented non-linear panel version of Carhart's model (2017-2018)

	Europe		US	
	Estimates	*s.e.*	Estimates	*s.e.*
		Full Sample		
$\hat{\alpha^*}$	0.0118	0.0145	0.0071	0.0065
$\hat{\beta}_{rm}$	0.9083***	0.0326	0.9277***	0.0331
$\hat{\beta}_{SMB}$	-0.0496	0.0620	0.0100	0.0394
$\hat{\beta}_{HML}$	-0.0989***	0.0172	-0.0358	0.0580
$\hat{\beta}_{MOM}$	0.1085***	0.0381	-0.0271	0.0357
		de jure Socially Responsible Mutual Funds		
$\hat{\hat{\alpha}}_s$	-0.0006	0.0007	-0.0003	0.0008
$\hat{\hat{\beta}}_{s,rm}$	0.0174	0.0221	0.0127	0.0340
$\hat{\hat{\beta}}_{s,SMB}$	0.1348***	0.0293	-0.0140	0.0415
$\hat{\hat{\beta}}_{s,HML}$	0.0659**	0.0281	0.0876	0.0607
$\hat{\hat{\beta}}_{s,MOM}$	0.0496*	0.0231	0.0986	0.0385
		de facto SRI Score		
$\hat{\alpha}_{SRI}$	-0.0196***	0.0001	-0.0154***	0.0001
$Adj.R^2$	0.7515		0.7980	
Fixed-Effects:				
Fund	Yes		Yes	
Time	Yes		Yes	
Obs	7,878		11,531	

Notes: This table reports estimates of the augmented non-linear panel version of Carhart's model (Eq. 1.5) derived from the GLM method. The Driscoll and Kraay (1998) correction is applied such that standard errors are robust to heteroscedasticity and autocorrelation. The notaions ***, ** and * indicate that the null hypothesis of a zero coefficient is statistically rejected at 99%, 95% and 90%.

The results obtained using time-varying *SRI scores* are qualitatively similar to those obtained with static *SRI scores* and are reported in Table 1.7. First, these results signal that considering static or time-varying scores does not matter for the result, as *SRI scores* are not very volatile. Such a result supports the long-term adherence to ethical objectives. Second, it turns out again that SRI branding and adherence to ethical labels thus do not harm funds' performance. In contrast, we observe that the estimated coefficient of the *SRI Score* (which holds for all mutual funds) is negative and significantly different from zero.

In a fourth experiment, we divide the SRI into its two distinct components: the ESG score, which considers environmental, social and governance aspects, and the Controversy score, which evaluates risks associated with a controversial announcement of an investment. Both measures correspond to effective measures and not to announcements. The results of the estimation are reported in Table 1.10.

Table 1.10: Estimation of the augmented non-linear panel version of Carhart's model (2013-2018) with ESG and Controversy scores

	Europe		US	
	Estimates	*s.e.*	Estimates	*s.e.*
	Full Sample			
$\hat{\alpha^*}$	0.0121***	0.0004	0.0083***	0.0020
$\hat{\beta}_{r^m}$	0.9638***	0.0178	1.0218***	0.0141
$\hat{\beta}_{SMB}$	0.1136***	0.0312	0.0896***	0.0267
$\hat{\beta}_{HML}$	-0.0953***	0.0315	-0.0361	0.0297
$\hat{\beta}_{MOM}$	-0.0192	0.0145	-0.0007	0.0119
	de jure Socially Responsible Mutual Funds			
$\hat{\hat{\alpha}}_s$	-0.0012***	0.0003	-0.0002	0.0002
$\hat{\hat{\beta}}_{s,r^m}$	0.0181	0.0175	-0.0154	0.0144
$\hat{\hat{\beta}}_{s,SMB}$	0.0940***	0.0336	-0.0190	0.0261
$\hat{\hat{\beta}}_{s,HML}$	0.0750**	0.0345	-0.0112	0.0283
$\hat{\hat{\beta}}_{s,MOM}$	0.0244	0.0188	0.0220*	0.0116
	de facto SRI Score			
$\hat{\alpha}_{ESGScore}$	-0.0139***	0.0001	0.0000	0.0000
$\hat{\alpha}_{ControversyScore}$	-0.0215***	0.0001	-0.0130***	0.0000
$Adj.R^2$	0.7527		0.8416	
Fixed-Effects:				
Fund	Yes		Yes	
Time	Yes		Yes	
Obs	40,602		59,429	

Notes: This table reports estimates of the augmented non-linear panel version of Carhart's model (Eq. 1.4) based on the GLM method. The Driscoll and Kraay (1998) correction is applied such that standard errors are robust to heteroscedasticity and autocorrelation. The notations ***, ** and * indicate that the null hypothesis of a zero coefficient is statistically rejected at 99%, 95% and 90%.

It appears that both the ESG and Controversy scores have negative and significant impacts on mutual fund adjusted risk performance. This confirms that both aspects of ethics have costs in terms of performance. Interestingly, the magnitude of the controversy score is twice as large as that of the ESG score, suggesting that it is of greater importance. Such a finding can be explained by the construction of the Controversy index, which relies on a 5-class categorization[15] before normalization on a $0-100$ scale.

In a fifth and final robustness check, we include the MSCI ratings in our CAPM regression. In focusing on the ESG scores provided by these two data providers, the aim is to check whether our results are robust to different extra-financial data sources. The two

[15]The score ranges from 0, which means no controversy, to 5, which indicates high controversy.

ESG scores are comparable because they measure the portfolio exposures to companies involved in environmental, social and governance challenges. Both ESG scores are fund-level measures built aggregating firm-level ratings. Morningstar and MSCI thus provide continuous and normalized ratings.[16] Although their ESG scores are very similar, they nevertheless present a difference: Morningstar ratings are free of industrial bias, whereas MSCI ratings are not. Thus, our experiment constitutes a robustness check to different extra-financial ratings and to different scoring methodologies. To avoid any selection bias due to the adjustment of the MSCI and Morningstar databases, we restrict our initial sample to institutional mutual funds (assets under management larger than USD/EUR 100 million) because both data providers have excellent coverage on this market segment (sharing approximately 92% of common mutual funds). The results of this experiment are reported in Table 1.11. First, the results confirm that our conclusions are robust to different extra-financial rating sources. Indeed, the impact of both Morningstar and MSCI ESG scores is negative and significant for the US and Europe. Interestingly, estimated coefficients obtained using the MSCI ESG score are smaller than those obtained with the Morningstar ESG score. It is highly probable that such a difference arises from the industry bias absent in Morningstar but present in MSCI.

[16]We slightly modify the normalization of MSCI ESG score to scale the rating between 0 and 100 instead of 0 and 10 to make the interpretation of the coefficients easier. See Appendix 1 and 2.

Table 1.11: Estimation of the augmented non-linear panel version of Carhart's model (2013-2018)

	Europe				US			
	Morningstar		MSCI		Morningstar		MSCI	
	Estimates	*s.e.*	Estimates	*s.e.*	Estimates	*s.e.*	Estimates	*s.e.*
				Full Sample				
$\hat{\alpha}^*$	0.0240***	0.0043	0.0096***	0.0019	0.0549**	0.0239	0.0079**	0.0036
$\hat{\beta}_{rm}$	0.9859***	0.0052	0.9859***	0.0052	1.0127***	0.0043	1.0127***	0.0043
$\hat{\beta}_{SMB}$	0.1327***	0.0219	0.1327***	0.0219	0.0799***	0.011	0.0799***	0.0110
$\hat{\beta}_{HML}$	-0.1067***	0.0218	-0.1067***	0.0218	-0.0541***	0.0138	-0.0541***	0.0138
$\hat{\beta}_{MOM}$	-0.0381***	0.0087	-0.0381***	0.0087	0.0196	0.0142	0.0196	0.0142
				de jure Socially Responsible Mutual Funds				
$\hat{\hat{\alpha}}_s$	-0.0022***	0.0008	-0.0043***	0.0012	-0.0010	0.0008	-0.0047***	0.0015
$\hat{\hat{\beta}}_{s,rm}$	0.0108***	0.0037	0.0108***	0.0037	-0.0043*	0.0024	-0.0043*	0.0024
$\hat{\hat{\beta}}_{s,SMB}$	0.0599***	0.0145	0.0599***	0.0145	-0.0095**	0.0047	-0.0095**	0.0047
$\hat{\hat{\beta}}_{s,HML}$	0.0287***	0.0084	0.0287***	0.0084	-0.0073	0.0101	-0.0073	0.0101
$\hat{\hat{\beta}}_{s,MOM}$	0.0554***	0.0053	0.0554***	0.0053	0.0073*	0.0042	0.0073*	0.0042
				de facto SRI Score				
$\hat{\alpha}_{SRI}$	-0.0340***	0.0062	-0.0104***	0.0020	-0.0993**	0.0442	-0.0089*	0.0053
Adj.R^2	0.7569		0.7569		0.8478		0.8478	
Fixed-Effects:								
Fund	Yes		Yes		Yes		Yes	
Time	Yes		Yes		Yes		Yes	
Obs	14,606		14,606		42,478		42,478	

Notes: This table reports estimates of the augmented non-linear panel version of Carhart's model (Eq. 1.4) based on the GLM method. The Driscoll and Kraay (1998) correction is applied such that standard errors are robust to heteroscedasticity and autocorrelation. The notations ***, ** and * indicate that the null hypothesis of a zero coefficient is statistically rejected at 99%, 95% and 90%.

1.5 Conclusion

This paper examines whether there is a difference between *what is said* and *what is done* in the SRI industry. To this end, it analyzes the socially responsible dimension of mutual funds along two dimensions. The first dimension addresses mutual funds' self-presentations and labels from specialized audit agencies (*de jure* or *what is said* about SRI), whereas the second dimension addresses the funds' holdings (*de facto* or *what is done* about SRI). This last aspect is measured via the newly available Morningstar Sustainability Rating and MSCI ESG Fund Metrics databases, which present many advantages. In particular, SRI and ESG scores are continuous, normalized and homogeneous for all European and US mutual funds. The performance of these funds is analyzed using a panel-augmented Carhart (1997) model, in which conventional and socially responsible mutual funds are split into two different clusters. To achieve comparable analysis of returns, a new database is built for mutual funds' returns in Europe and in the US. The database is corrected for potential survivorship and entrance biases.

This paper enriches the literature on SRI and leads to several institutional and professional implications. First, it reveals the weak correspondence between *what is said* and *what is done* in the socially responsible mutual fund industry; branding from asset managers and from specialized audit agencies appear to serve only as adverstising. Second, it reveals a dichotomy between the effect on performance of *what is said* and *what is done* about SRI: the *de jure* SRI has a small and marginally significant impact on a fund's financial performance, whereas *de facto* SRI has a significant negative effect on the performance. This last result is in line with the existing theoretical literature on the consequences of extra-financial constraints on mutual fund performance.[17] See Bollen (2007) for a discussion of the substitutability property of investors' utilities and, more generally, Fama and French (2007) about different taste consequences. For a behavioral approach, see also Levitt and List (2007) and Doskeland and Pedersen (2016) about investors' financial and moral utilities. In addition, this dichotomy explains the puzzle often

[17] Within a Markowitz (1952) mean-variance framework, adding an extra-financial constraint can only penalize the optimization program of the investor. Recently, Gasser et al. (2017) revisited the cost of extra-financial constraints on portfolio performance. The authors computed the three-dimensional capital allocation plane and showed that investors who maximize their extra-financial utility face a statistically significant decrease in expected returns.

encountered in the empirical literature about the financial impact of ethics. In summary, information asymmetry emerges in the SRI market from investors' difficulty in identifying the true nature of mutual funds and evaluating their performance conditionally on their SRI classification.

Therefore, this paper stresses the existing urgent need to regulate the SRI market. One approach would consist of creating a public and specialized audit agency. This agency would only evaluate the *de facto* aspect of the socially responsible mutual funds to set the path and improve the quality of other SRI labels, which only consider *de jure* SRI aspects. Asymmetric information between funds' managers and investors would thus be reduced. It would also lead to harmonization of the evaluation of the SRI objectives of the funds. Favoring the transparency of the SRI market as well as the evaluation of extra-financial returns would generate confidence in this segment of funds. It is a *sine qua non* condition to improve market efficiency on this segment (i.e., to enable a better capital allocation between conventional and socially responsible mutual funds).

References

Ando, T. and J. Bai (2015). “Asset pricing with a general multifactor structure”. In: *Journal of Financial Econometrics* 13.3, pp. 556–604.

Armstrong, W. J., E. Genc, and M. Verbeek (2019). “Going for gold: An analysis of Morningstar analyst ratings”. In: *Management Science* 65.5, pp. 2310–2327.

Barber, B. M., A. Morse, and A. Yasuda (2019). *Impact investing.* Tech. rep. National Bureau of Economic Research.

Bauer, R., K. Koedijk, and R. Otten (2005). “International evidence on ethical mutual fund performance and investment style”. In: *Journal of Banking & Finance* 29.7, pp. 1751–1767.

Bauer, R. and P. Smeets (2015). “Social identification and investment decisions”. In: *Journal of Economic Behavior & Organization* 117, pp. 121–134.

Berg, F., J. F. Koelbel, and R. Rigobon (2019). *Aggregate Confusion: The Divergence of ESG Ratings.* MIT Sloan Research Paper.

Blomquist, J. and J. Westerlund (2013). “Testing slope homogeneity in large panels with serial correlation”. In: *Economics Letters* 121.3, pp. 374–378.

Bollen, N. P. (2007). “Mutual fund attributes and investor behavior”. In: *Journal of Financial and Quantitative Analysis* 42.3, pp. 683–708.

Borgers, A. et al. (2015). “Do social factors influence investment behavior and performance? Evidence from mutual fund holdings”. In: *Journal of Banking & Finance* 60, pp. 112–126.

Brown, S. J. and W. N. Goetzmann (1994). “Attrition and mutual fund performance”. In: *Journal of Finance*. Vol. 49. 3, pp. 1055–1056.

Brown, S. J., W. Goetzmann, et al. (1992). “Survivorship bias in performance studies”. In: *The Review of Financial Studies* 5.4, pp. 553–580.

Carhart, M. M. (1997). “On persistence in mutual fund performance”. In: *The Journal of finance* 52.1, pp. 57–82.

Cooper, M. J., O. Dimitrov, and P. R. Rau (2001). “A rose. com by any other name”. In: *The journal of Finance* 56.6, pp. 2371–2388.

Cooper, M. J., H. Gulen, and P. R. Rau (2005). "Changing names with style: Mutual fund name changes and their effects on fund flows". In: *The Journal of Finance* 60.6, pp. 2825–2858.

Doellman, T. W. et al. (2019). "Alphabeticity bias in 401 (k) investing". In: *Financial Review* 54.4, pp. 643–677.

Døskeland, T. and L. J. T. Pedersen (2016). "Investing with brain or heart? A field experiment on responsible investment". In: *Management Science* 62.6, pp. 1632–1644.

Driscoll, J. C. and A. C. Kraay (1998). "Consistent covariance matrix estimation with spatially dependent panel data". In: *Review of economics and statistics* 80.4, pp. 549–560.

El Ghoul, S. and A. Karoui (2017). "Does corporate social responsibility affect mutual fund performance and flows?" In: *Journal of Banking & Finance* 77, pp. 53–63.

Elton, E. J., M. J. Gruber, and C. R. Blake (1996). "Survivor bias and mutual fund performance". In: *The review of financial studies* 9.4, pp. 1097–1120.

Espenlaub, S., I. ul Haq, and A. Khurshed (2017). "It's all in the name: Mutual fund name changes after SEC Rule 35d-1". In: *Journal of Banking & Finance* 84, pp. 123–134.

Fama, E. F. and K. R. French (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of financial economics* 33.1, pp. 3–56.

— (2007). "Disagreement, tastes, and asset prices". In: *Journal of financial economics* 83.3, pp. 667–689.

Fellbaum, C. (1998). "A semantic network of English verbs". In: *WordNet: An electronic lexical database* 3, pp. 153–178.

Goldreyer, E. F. and J. D. Diltz (1999). "The performance of socially responsible mutual funds: Incorporating sociopolitical information in portfolio selection". In: *Managerial Finance.*

Green, T. C. and R. Jame (2013). "Company name fluency, investor recognition, and firm value". In: *Journal of Financial Economics* 109.3, pp. 813–834.

Gregory, A., J. Matatko, and R. Luther (1997). "Ethical unit trust financial performance: small company effects and fund size effects". In: *Journal of Business Finance & Accounting* 24.5, pp. 705–725.

Grinblatt, M. and S. Titman (1989). "Mutual fund performance: An analysis of quarterly portfolio holdings". In: *Journal of business*, pp. 393–416.

Hamilton, S., H. Jo, and M. Statman (1993). “Doing well while doing good? The investment performance of socially responsible mutual funds”. In: *Financial Analysts Journal* 49.6, pp. 62–66.

Hanke, B. et al. (2018). “Survivorship bias and comparability of UK open-ended fund databases”. In: *Economics Letters* 172, pp. 110–114.

Hartzmark, S. M. and A. B. Sussman (2019). “Do investors value sustainability? A natural experiment examining ranking and fund flows”. In: *The Journal of Finance* 74.6, pp. 2789–2837.

Hong, H. and M. Kacperczyk (2009). “The price of sin: The effects of social norms on markets”. In: *Journal of Financial Economics* 93.1, pp. 15–36.

Humphrey, J. E. and D. T. Tan (2014). “Does it really hurt to be responsible?” In: *Journal of business ethics* 122.3, pp. 375–386.

Jacobs, H. and A. Hillert (2016). “Alphabetic bias, investor recognition, and trading behavior”. In: *Review of Finance* 20.2, pp. 693–723.

Kreander, N. et al. (2005). “Evaluating the performance of ethical and non-ethical funds: a matched pair analysis”. In: *Journal of Business Finance & Accounting* 32.7-8, pp. 1465–1493.

Levitt, S. D. and J. A. List (2007). “What do laboratory experiments measuring social preferences reveal about the real world?” In: *Journal of Economic perspectives* 21.2, pp. 153–174.

Lintner, J. (1969). “The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets: A reply”. In: *The review of economics and statistics*, pp. 222–224.

Luther, R. G. and J. Matatko (1994). “The performance of ethical unit trusts: choosing an appropriate benchmark”. In: *The British Accounting Review* 26.1, pp. 77–89.

Luther, R. G., J. Matatko, and D. C. Corner (1992). “The Investment Performance of UK”Ethical”Unit Trusts”. In: *Accounting, Auditing & Accountability Journal* 5.4, p. 57.

Malkiel, B. G. (1995). “Returns from investing in equity mutual funds 1971 to 1991”. In: *The Journal of finance* 50.2, pp. 549–572.

Mallin, C. A., B. Saadouni, and R. J. Briston (1995). “The financial performance of ethical investment funds”. In: *Journal of Business Finance & Accounting* 22.4, pp. 483–496.

Markowitz, H. (1952). “Portfolio Selection”. In: *Journal of Finance* 7.1, pp. 77–91.

Miller, G. A. (1995). "WordNet: a lexical database for English". In: *Communications of the ACM* 38.11, pp. 39–41.

Nofsinger, J. and A. Varma (2014). "Socially responsible funds and market crises". In: *Journal of Banking & Finance* 48, pp. 180–193.

Pesaran, M. H. and T. Yamagata (2008). "Testing slope homogeneity in large panels". In: *Journal of econometrics* 142.1, pp. 50–93.

Petersen, M. A. (2009). "Estimating standard errors in finance panel data sets: Comparing approaches". In: *The Review of Financial Studies* 22.1, pp. 435–480.

Renneboog, L., J. Ter Horst, and C. Zhang (2008a). "Socially responsible investments: Institutional aspects, performance, and investor behavior". In: *Journal of banking & finance* 32.9, pp. 1723–1742.

— (2008b). "The price of ethics and stakeholder governance: The performance of socially responsible mutual funds". In: *Journal of corporate finance* 14.3, pp. 302–322.

Rhodes, M. J. (2010). "Information asymmetry and socially responsible investment". In: *Journal of business ethics* 95.1, pp. 145–150.

Riedl, A. and P. Smeets (2017). "Why do investors hold socially responsible mutual funds?" In: *The Journal of Finance* 72.6, pp. 2505–2550.

Schröder, M. (2007). "Is there a difference? The performance characteristics of SRI equity indices". In: *Journal of Business Finance & Accounting* 34.1-2, pp. 331–348.

Schwert, G. W. (2003). "Anomalies and market efficiency". In: *Handbook of the Economics of Finance* 1, pp. 939–974.

Sharpe, W. F. (1964). "Capital asset prices: A theory of market equilibrium under conditions of risk". In: *The journal of finance* 19.3, pp. 425–442.

Statman, M. (2000). "Socially responsible mutual funds (corrected)". In: *Financial Analysts Journal* 56.3, pp. 30–39.

— (2004). "What do investors want?" In: *The Journal of Portfolio Management* 30.5, pp. 153–161.

— (2006). "Socially responsible indexes". In: *The Journal of Portfolio Management* 32.3, pp. 100–109.

Statman, M. and D. Glushkov (2009). "The wages of social responsibility". In: *Financial Analysts Journal* 65.4, pp. 33–46.

Statman, M. and D. Glushkov (2016). “Classifying and measuring the performance of socially responsible mutual funds”. In: *The Journal of Portfolio Management* 42.2, pp. 140–151.

Appendices

Appendix 1: Description of Morningstar Sustainability Rating

Portfolio sustainability scores are obtained from Morningstar. This score reflects the quality of mutual funds relatively to environmental, social and governance (ESG) firm-level scores and is calculated as follows:

$$Portfolio.Sustainability.Score = Portfolio.ESG.Score - Portfolio.Controversy.Deduction \tag{1.6}$$

The calculation of the Portfolio ESG score is based on data provided by "Sustainalytics" [18], a leader in ESG asset rating. To attribute an ESG score to a company (or an asset), Sustainalytics compares a company to other companies of the same industry and uses different indicators on a $0-100$ scale. Morningstar then normalizes scores as follows:

$$Zc = \frac{ESG_i + \mu_{industry}}{\sigma_{industry}}, \tag{1.7}$$

$$ESG.Norm_i = 50 + 10Zc. \tag{1.8}$$

Morningstar determines the Portfolio ESG score from the weighted average of an asset's ESG score. For a fund to be graded, Morningstar requires the fund to have at least 50% of assets with an ESG score obtained from Sustainalytics (the percentage of assets scored is rescaled to 100%).

$$Portfolio.ESG.Score = \sum_{i=1}^{n} w_i ESG.Norm_i, \tag{1.9}$$

where:

$\sum_{i=1}^{n} w_i = 1$ and for each i, $w_i = \frac{x_i}{\sum_{i=1}^{n} x_i}$ if $\sum_{i=1}^{n} x_i > 50\%$.

Portfolio controversy deduction is also obtained by Sustainalytics. Sustainalytics assesses companies involved in ESG-related incidents on a $0-100$ scale. This negatively contributes to the Portfolio sustainability score. For a fund to be graded, Morningstar requires that 50% of the fund's assets have a controversy score assigned by Sustainalytics

[18] http://www.sustainalytics.com

(the percentage of assets scored is rescaled to 100%). Portfolio controversy deduction is calculated as follows:

$$Portfolio.Controversy.Score = \sum_{i=1}^{n} w_i.S.Contr_i. \tag{1.10}$$

Appendix 2: Description of MSCI ESG Fund Metrics

In March 2016, MSCI launched the MSCI ESG Fund Metrics tool to provide fund-level data to investors. Among ESG for the total portfolios, ESG Quality Score measures mutual fund exposure of companies that address environmental, social and governance (ESG) challenges. The ESG Quality Score has a high coverage score (around 90% in our dataset and at least 65% in the whole MSCI database), the score is a composite index built aggregating the firm-level ratings extracted from MSCI ESG Research. The ESG Quality Score has similar features with MSCI's ESG score: the score is continuous defined and normalized. To allow the fair comparison between Morningstar and MSCI ESG scores, we slightly modify the normalization step to equalize the rating support (Morningstar and MSCI ESG scores are defined between 0-100 and 0-10 respectively). More formally, the MSCI ESG score is denoted ESG_{score} and is defined as follows :

$$Zc = \frac{ESG_{QualityScore} + \mu_{score}}{\sigma_{score}}, \tag{1.11}$$

$$ESG_{score} = 50 + 10Zc. \tag{1.12}$$

Appendix 3: Panel version of Carhart's model with dummy *de jure* (labels)

This model with a dummy for funds that have been granted a label is only performed for European mutual funds, as there is no existing unified label in the US.

Table 1.12: Estimation of the panel version of Carhart's model (2013-2018) with labels

	Europe	
	Estimates	*s.e.*
$\hat{\alpha}$	-0.0027	0.0067
$\hat{\beta}_{rm}$	0.9765***	0.0192
$\hat{\beta}_{SMB}$	0.1754***	0.0492
$\hat{\beta}_{HML}$	-0.0170	0.0234
$\hat{\beta}_{MOM}$	0.0045	0.0250
$\hat{D}ummyDeJure$	0.0012	0.0068
$Adj.R^2$	0.7668	
Fixed-Effects:		
Fund	Yes	
Time	Yes	
Obs	40,602	

Notes: This table reports estimates of the augmented non-linear panel version of Carhart's model (Eq. 1) based on the GLM method with an extra *de jure* dummy for labelised funds. Driscoll and Kraay (1998) correction is applied such that standard errors are robust to heteroscedasticity and autocorrelation. ***, ** and * indicate that the null hypothesis of a zero coefficient is statistically rejected at 99%, 95% and 90%.

Chapter 2

Taming Financial Development to reduce Crises

Authors: Sami Ben Naceur, Bertrand Candelon, Quentin Lajaunie

About this chapter

The authors thank Juliana Araujo, Norbert Funke, Nigel Nagarajan, Kate Phylaktis (the editor-in-charge of the "Emerging Market Review"), Tim Rose, one anonymous referee as well as the participants at the 1st Annual Conference of the JRC Community of Practice in Financial Research (CoPFiR) for stimulating comments and questions. They also benefit for excellent research assistance on the data collection from Sena Oztosun. This research has been conducted with the research program "Risk Management and Investment Strategies" under the aegis of the Europlace Institute of Finance, Insti7, and of IMF-ICD's Visiting Scholar program. The usual disclaimer applies: the views expressed here are those of the authors, and the paper does not represent the views of the IMF, its Executive Board or its management

Abstract

This paper assesses whether and how financial development triggers the occurrence of banking crises. It builds on a database that includes financial development as well as financial access, depth and efficiency for almost 100 countries. Through estimation of a dynamic logit panel model, it appears that financial development, from an institutional dimension and to a lesser extent from a market dimension, triggers financial stability within a one- to two-year horizon. Additionally, whereas financial access is destabilizing for advanced countries, it is stabilizing for emerging and low incomes ones. Both results have important implications for macroprudential policies and financial regulations.

Keywords: Financial Development ; Banking crises ; Regulation.

2.1 Introduction

Early writers have found that financial development has been associated with higher growth, lower inequality/poverty and reduced economic volatility (Levine 2005). More recent literature has highlighted the vanishing effect of financial development on growth (Cecchetti and Kharoubi, 2012) and the existence of nonlinearities in the financial growth-nexus, showing that financial development starts lowering growth when a threshold of credit-to-private/GDP is crossed (Arcand et al., 2015).

This vanishing effect stems from financial deepening, rather than from better inclusion or higher efficiency (Sahay et al. 2015). Similarly, a large number of empirical studies have examined how financial development may generate future banking crises (cf. Demirgüç–Kunt and Detragiache, 2005 or Cihák, 2007). The researchers all find that credit growth is the most important factor at the origin of banking crises. Still, these works are incomplete in two dimensions. First, contrary to the literature focusing on growth, these works include financial development through different aggregates or proxies, but they do not consider its different dimensions (depth, access and efficiency). Regulators are thus left with the option to limit or to facilitate financial development, enabling the precise targeting of which dimension (access, depth or efficiency) should be favored or limited. The practical implementation of optimal rules on financial development is thus impossible. Second, all these studies consider static binary models (probit/logit), whereas Kauppi and Saikkonen (2008) and more recently Candelon et al. (2014) have shown that this specification can be misleading if the banking crisis or the dependent variables exhibit persistence.

This papers aims to fill these two gaps. It considers a panel of 98 countries, for which the systemic banking crisis database of Leaven and Valencia (2013) has been extended to 2016. Following Svirydzenka (2016), the financial development variable is decomposed into 6 subindices, which measure depth, access and efficiency for both institutions and the market sector. Finally, the relationship between banking crises and the financial development indicators is analyzed using a dynamic panel logit model. It turns out that for the whole panel, financial development indeed increases the probability of occurrence of a crisis. It can be via either financial institution development or to a lesser extent financial market development. Considering now three groups of countries clustered according to their degree of development, we observe that in advanced economies, depth (FID) and access (FIA) cause banking crises, whereas for least income developing countries (LIDC) and emerging markets (EM), only the financial institution's depth constitutes a leading indicator for future crises. In the latter case, access to financial services enhances financial stability, whereas it should be limited in developed countries.

These findings convey important messages to regulators. The results first confirm the potential destabilizing effect of financial development leading to systemic banking crises. The findings hence support the implementation of regulatory measures, such as capital requirements and access control to loans and deposits for financial institutions, in order to stabilize the system. Second, the results show that regulation should not be unique but that it should take into account the degree of development of the country. Whereas access to financial institutions is destabilizing for advanced countries (increasing, for example, the amount of nonperforming loans), it is stabilizing for the other countries (of middle and low incomes) via the promotion of financial inclusion and the reduction of inequalities. Regulators should thus impose strict access control for financial intermediaries in advanced countries. In contrast, regulators should enhance its access, supporting for example fintech industry and its financial innovations (mobile application payments, etc.) in low-income countries.

The paper is composed as follows. In Section 2.2, a literature review is presented. Section 2.3 deals with the methodology. Section 2.4 describes the database. Section 2.5

exhibits the empirical results for the whole sample of countries as well as for 3 clusters of countries. Section 2.6 provides several robustness checks, and Section 2.7 exposes the consequences for regulators and concludes.

2.2 Literature Review

The literature on financial development started with considerations of the relationship between financial deepening and economic growth.[1] Nevertheless, it overlooked the issue of whether a larger financial system is associated with a higher occurrence of crises. Several theories have been proposed to explain why finance can lead to banking crises. In the 1980s and 1990s, many developed and developing countries witnessed a wave of systemic banking crises affecting growth. Bernanke and Blinder (1988) and Kiyotaki and Moore (1997) elaborate theoretical justifications to prove that the quantity of credit in the economy is positively associated with the probability of a banking crisis. Minsky (1986) explain that an extended period of financial stability encourages excess borrowing which may led to a banking crisis. Keeley (1990) and Dell'Ariccia and Marquez (2004) argue that financial liberalization comes with more competition and a lower standard of lending, which can lead to an increase in lending to lower quality borrowers and a higher probability of banking crises.

From an empirical point of view, the literature on the finance-banking crises nexus is linked to the development of early warning systems (EWS) for banking crises. To summarize, EWS has come up with two main approaches: the signal approach and the binary regression approach. The signal approach identifies individual variables that best signal a threat to financial stability (Kaminski and Reinhart, 1999nocitekaminsky1999twin; Borio and Drehmann, 2009; Drehmann and Juselius, 2014). The second approach, the binomial or multinomial logit or probit, relates a binary banking crisis dummy to multiple explanatory variables to predict banking crises (Demirgüç–Kunt and Detragiache, 1998, 2000, 2002; Davis et al., 2011; Schularick and Taylor 2012 F; and Duca and Peltonen 2013). Davis and Karim (2008a, 2008b) suggest that the logit approach outperforms the

[1] See inter alii Schumpeter , Robinson (1952), Goldsmith (1970), McKinnon and Shaw (1973), Shaw (1973), Lucas (1988), Greenwood and Jovanovic (1990), Bencivenga (1991), Saint-Paul (1992) Rousseau and Wachtel (2000) for theoretical considerations and Levine (2005) for a review of empirical studies.

signaling approach because the former exhibits lower type I (missed crises) and type II (false alarms) errors compared to the latter. The literature has proposed various early warning indicators such as a high inflation rate, a large current account deficit, house price inflation, an increase in the real interest rate, and excessive domestic credit (see Kauko, 2014 for a review). We will focus on the credit variables (measures of financial deepening) as banking crisis determinants, since our interest is to examine the impact of financial development on banking crises.

Two kinds of credit variables have been often considered as banking crisis early warning indicators, namely, the credit-to-GDP ratio and the credit growth rate. Demirgüç–Kunt and Detragiache (1998, 2005) find that the level of the credit-to-GDP ratio and the growth rate of credit have a robust positive effect on bank crisis occurrences. Sahay et al. (2015) find that a faster pace of financial deepening increases financial instability when the financial system is weakly regulated and supervised. In contrast, Davis and Karim (2008), in replicating the Demirgüç–Kunt and Detragiache 2005 analysis on a larger sample and a longer period, find that the credit-to-GDP ratio is not a good predictor of banking crises. This result was confirmed by Hahm et al. (2013) and Von Hagen and Ho (2007). Rose and Spiegel (2011) also concluded that the ratio of credit relative to GDP was of no use as a predictor of the Global Financial Crisis of 2007-2008.

In a recent paper, Mathonnat and Minea (2018) revisited these conflicting findings using a large sample of banking crises and five financial development variables, finding that the level of the credit-to-GDP ratio jointly introduced with its growth and volatility does not affect significantly the occurrence of banking crises.

The lagged value of the other credit variable, the growth rate of credit, has been tested as a bank crisis predictor. Jordà et al (2011), Schularick and Taylor (2012), Demerguc-Kunt and Detragiache (2005), Kaminsky et al (1998, 1999) and Bordo and Meissner (2012), among others, show that credit growth lagged two years and up is a good predictor of crises. Bunda and Ca'Zorzi (2010) and Barrell et al. (2011) conclude that credit growth lagged one year is of no use as a crisis predictor. Büyükkarabacak and Valev (2010) decompose private credit into household credit and enterprise credit. They find that a rapid expansion of household credit is a significant predictor of banking crises. Mathonnat and Minea (2018) used instead monetary aggregate growth and found that the growth of M3/GDP impacts positively and significantly the probability of a banking crisis. Drehmann et al.

(2011) substituted the growth rate of credit by the trend deviation of the credit-to-GDP ratio and found evidence that this variable is the best of ten different potential variables to predict banking crises, confirming Borio and Lowe's (2002) initial conclusion suggesting that a credit-to-GDP trend deviation reaches its peak three years before the occurrence of a banking crisis.

2.3 Methodology: The Dynamic Panel Logit model

The methodology used in this paper belongs to the second approach of EWS, relating a binary banking crisis dummy to multiple explanatory variables to predict banking crises. It builds on a dynamic panel logit version proposed by Kauppi and Saikkonen (2008) and Candelon et al. (2014).

Let us denote by $\{y_{i,t}\}_{t=1}^{T}$ the banking crisis binary variable for country i, $i \in \{1, 2, ...N\}$[2], which takes the value 1 during crisis periods and 0 otherwise. Similarly, $\{x_{i,t}\}_{t=1}^{T}$ represents the matrix composed of the k explanatory variables, which are the indicators of financial development and the macroeconomic control variables in our case.

The dynamic panel logit model has the following form:

$$\Pr(y_{it} = 1) = F(\pi_{i,t}) = F(\alpha y_{i,t-1} + x_{i,t-1}\beta + \delta\pi_{i,t-1} + \eta_i), \text{ for } t = 1, 2, ...T, \text{ and } i = 1, 2, ..N, \tag{2.1}$$

where N is the number of countries; $\Pr_{t-1}(y_t = 1)$ is the conditional probability of observing a banking crisis given the information set we have at our disposal at time $t-1$; and π_t is the index at time t. F is the logistic c.d.f., which is preferred to the Gaussian one, as it is more appropriate for the study of extreme events such as crises. η_i is a country fixed effect for the control of unobserved heterogeneity and potential bias. The coefficient β informs us about the one-step-ahead causal relationship between the explanatory variables (financial development proxies and/or the macroeconomic variables) and the banking crises. If the sign is positive (resp. negative) it indicates that the probability of occurrence of a crisis in a horizon of one year will increase (resp. decrease). The dynamic of the crisis is captured by the coefficients α and π. α is associated with the lagged binary banking crisis variable, whereas π is linked to the lagged index. Both terms capture the persistence of the crisis and constitute the innovation proposed by Kauppi and Saikkonen (2008). If

[2]For ease of notation, the country index i is omitted hereafter.

one of them is significantly different from 0, then it implies that the traditional static logit models are biased and that their interpretations may be misleading. Candelon et al. (2014) show that the different alternatives of this general model can be estimated under the same exact maximum likelihood (EML) framework.

To be more precise, the log-likelihood function has the following general form:

$$\mathrm{LogL}(\theta, \eta_i) = \sum_{i=1}^{N} \mathrm{LogL}_i(\theta, \eta_i) = \sum_{i=1}^{N} \sum_{t=1}^{T} [y_{it} \log(F_{it}) + (1 - y_{it}) \log(1 - F_{it})], \tag{2.2}$$

where θ represents the vector of parameters.

The EML estimators have the desired large-sample properties. As shown in Candelon et al. (2014), 4 different models can be considered, each of which correspond to particular restrictions of the general log-likelihood function.

The first model is the static logit model. In such a case $\alpha = \pi = 0$. Only the exogenous macroeconomic variables affect the future occurrence of a banking crisis. The second and third models are dynamic and include either the lagged value of the binary dependent variable y_{t-1} or the lagged index π_{t-1}. Finally, the most complex dynamic model combines the two previous cases and includes both the lagged dependent variable y_{t-1} and the lagged index π_{t-1}. The best model is chosen as the one minimizing the Akaike information criterion (AIC).

Finally, since we do not make any assumptions about the distribution of $\{\eta_i\}_{i=1}^{N}$, they are treated as parameters to be estimated, and our approach is a fixed effects one. In addition, we assume no cross-sectional dependence. In such a case, we follow Candelon et al. (2014) and implement a correction á la Carro (2007).[3]

2.4 Data

For a long time, financial development has been measured by proxies such as the credit-to-GDP ratio or stock market capitalization (see Rajan and Zingales, 1998 or more recently Arcand et al., 2015). Still, financial development (FD) has evolved and is now multidimensional. In many countries, financial institutions (FI) have grown. Traditional players such as investment banks, insurance companies, mutual funds, pension funds, and venture capital firms are now in competition with many other types of nonbank financial

[3]The correction is explained in details in Candelon et al. (2014).

institutions, which are now playing substantive roles. Additionally, it is possible to enter the markets bypassing these traditional institutions. Internet trading platforms allow you to invest directly your savings. This finance market (FM) is relatively important in the US and developed countries, whereas it remains so far limited in low- and middle-income countries. An adequate index of financial development should encompass both of these dimensions. Recent studies (Cihák et al., 2012, Sahay et al., 2015) propose to disaggregate Financial development into different dimensions: depth (D), corresponding to the size and liquidity of the markets; access (A), measuring the ability of individuals and companies to access financial services; and efficiency (E), indicating the level of activity in capital markets and the ability of institutions to provide financial services at low cost and with sustainable revenues. We are thus left with 9 measures for financial development: the global one FD, composed of two sub-indices FI and FM, which are each finally decomposed into 3 individual indicators FID, FIA, FIE, FMD, FMA, and FME.

Sahay et al, (2015) explain in detail how indices are built. The first step consists in building the 6 sub-indices FID, FIA, FIE, FMD, FMA, and FME. Each sub-index depends on a set of specific variables. For example, Financial Institutions Depth (FID) is a composite indicator including Private-sector credit, Pension fund assets, Mutual fund assets, and Insurance premiums, life and non-life. Appendix 1 provides with a list of variables considered for each sub-indices. The weights of each variable in the composite index is obtained via a principal component analysis (PCA). In a second step, once the sub-indices built, indices FI and FM are also constructed via a PCA approach based on the sub-indices. In a final step FD index is built from FI and FM. Svirydzenka (2016) follows this methodology to set up a global database for 183 countries covering the period $1980 - 2016$ at an annual frequency. Besides, Svirydzenka (2016) categorizes countries according to their income level into advanced, emerging and low-income developing countries. Appendix 2 table 2.9 provides with the average value of the financial development indicators according to this categories.

For the banking crisis dummy, we use the Leaven and Valencia (2013) database, which encounters systemic financial market failures. As the sample size only covers the period until 2011, we complete it for the period $2012 - 2016$ with the database of Candelon et al. (2018).[4] The banking crisis database thus contains 100 countries from 1980 to 2016 on a

[4]They expand the sample to more countries and more years, up to 2016, using data from Harvard Busi-

yearly basis.[5]

Following Demirgüç–Kunt and Detragiache (1998), the macroeconomic control variables retained are the output growth rate, the interest rate spread built as the difference between the $10-year$ treasury rate and a $3-month$ monetary rate. A fast growing economy increases banks' asset value and decrease the amount of non performing loans. Both factors reduce the probability of occurrence of a banking crisis. The term spread (long minus short rate) corresponds to the slope of the yield curve, which is a good leading indicator for economic activity.[6] Both are yearly and extracted from the International Financial Statistics (IFS) database of the International Monetary Fund. In summary, our sample comprises 98 countries for the period $1980-2016$.

2.5 Results

Our empirical strategy consists in estimating model (2.1) considering the different proxies for financial development. We begin with the most aggregated one (FD), then consider FI and FD, to finish with the most detailed ones (FID, FIA, FIE, FMD, FMA, and FME).

For each model, we have estimated all the four specifications of the EWS (static, with a lag binary variable, with the index, with the lag binary as well as the index).[7] We observe that the second specification including one lag binary variable presents in each case the lowest AIC and thus is always selected. This observation then confirms that considering a static logit model is not adequate and that the persistence should be introduced by the lagged binary variable, indicating that causality is nonlinear in essence.[8]

ness School (HBS) $http://www.hbs.edu/faculty/initiatives/behavioral-finance-and-financial-stability/Pages/global.aspx$.

[5]Appendix 3 provides with a list of the countries as well as the dates of banking crises. It also reports the country group they belong to. Figure 2.1 report banking crisis frequency on a year-by-year basis.

[6]Several other control variables have been tested, as the index of banking competition -the concentration index in the financial sector (Lerner index)- inflation or the government surplus as a share of GDP. Besides being available only for a very restrictive number of countries, they turn out to be not significantly different from 0. For sake of space, they are not reported, but available from authors upon request.

[7]Results of these estimations are not reported to save space but are available upon request from the authors.

[8]See Candelon at al. (2013) for a discussion on causality in binary dynamic models. Even if it is proposed in the case of multivariate model it can be easily translated to univariate ones.

Table 2.1 reports the estimates of the EWS, where the dependent variable is the banking crisis index explained by different indicators of financial development and macro-variables. In this first exercise, all the countries are considered for the period 1980–2016.

The selected macroeconomic control variables (output growth, inflation and interest rate term spread) are always significant and exhibit the expected signs. Indeed, an increase in the output growth rate stimulates banks' returns and reduces significantly the probability of occurrence of a banking crisis at a horizon of one year. Similarly, an inversion of the yield curve, i.e., the negative interest rate term spread signals a higher risk of banking crisis. Finally, inflation turns out to deteriorate banks' balance sheet and increases banking crises probability.

When estimating the panel for the 98 countries over the period 1980–2016, financial development appears to increase financial instability, increasing the probability of occurrence of a crisis in a one-year period (see column (1)). It can be via either the financial institution (column (2)) or the financial market development (column (3)).

Concerning financial market development (column (4) and (6)), we observe an opposite effect between the depth and the access dimension. Whereas the financial institution's depth is destabilizing, the access dimension actually reduces the future occurrence of a crisis. When a financial institution's efficiency has a negative sign, a banking crisis is not significantly affected. Similar results are observed for the indicators of financial market development: a positive sign for the depth and a negative sign for the access dimension. Nevertheless, almost all coefficients are not statistically different from 0, and thus, the impact of an improvement in financial institution or market efficiency is quite small.

Table 2.1: Estimation Results

Model	(1)	(2)	(3)	(4)	(5)	(6)
constant	-3.122***	-3.166***	-2.927***	-3.235***	-2.970***	-3.197***
	(0.686)	(0.703)	(0.682)	(0.735)	(0.681)	(0.740)
lag binary	3.880***	3.904***	3.856***	3.840***	3.843***	3.815***
	(0.170)	(0.170)	(0.170)	(0.171)	(0.170)	(0.172)
			Financial development variables			
FD_{-1}	2.722***					
	(0.900)					
FI_{-1}		1.755*				
		(1.065)				
FM_{-1}			2.236***			
			(0.662)			
FID_{-1}				5.690***		5.144***
				(1.470)		(1.861)
FIA_{-1}				-2.700***		-2.784***
				(1.309)		(1.366)
FIE_{-1}				-1.370		-0.239
				(0.889)		(0.884)
FMD_{-1}					2.260***	1.482
					(0.954)	(1.045)
FMA_{-1}					-1.450	-1.911*
					(1.010)	(1.106)
FME_{-1}					0.704	0.836
					(0.557)	(0.565)
			Macro-control variables			
Spread	-0.062*	-0.0501	-0.066**	-0.069**	-0.069**	-0.075***
	(0.025)	(0.034)	(0.035)	(0.037)	(0.036)	(0.038)
Output growth	-0.101***	-0.099***	-0.123***	-0.100***	-0.106***	-0.105***
	(0.019)	(0.184)	(0.018)	(0.018)	(0.019)	(0.019)
			Relevant Statistics			
AIC	1288.3	1295.1	1285.8	1284.5	1284.7	1281.4
$Pseudo - R^2$	0.521	0.518	0.522	0.524	0.524	0.529
#Observations	3626	3626	3626	3626	3626	3626

Notes: This table reports the estimates obtained from the dynamic logit models (1) to (6) for the panel of 98 countries from $1980 - 2016$. Standard errors are reported within brackets below the estimates. $***$, $**$ and $*$ report significance at 99%, 95% and 90%.

Tables 2.2 –2.4 report the estimates of the EWS with a lag binary dependent variable, where the dependent variable is the systemic banking crisis index explained by indicators of financial development and macro-variables, considering clusters of countries according to their financial market development.[9] Following Svirydzenka (2016), three clusters have thus been created: one for the least income developing countries (LIDC), one for the emerging markets (EM), and the last one for the advanced markets (AM).

[9]See Svirydzenka, 2016.

Table 2.2: Estimation Results - Advanced Economies

Model	(1)	(2)	(3)	(4)	(5)	(6)
constant	-7.471***	-10.408***	-5.533***	-8.381***	-5.358***	-6.222**
	(1.399)	(2.173)	(1.062)	(2.217)	(1.061)	(2.471)
lag binary	4.315***	4.428***	4.290***	4.257***	4.268***	4.156***
	(0.364)	(0.368)	(0.360)	(0.374)	(0.360)	(0.377)
Financial development variables						
FD_{-1}	6.279***					
	(1.444)					
FI_{-1}		9.849***				
		(2.503)				
FM_{-1}			3.952***			
			(0.941)			
FID_{-1}				6.660***		2.363
				(2.146)		(3.127)
FIA_{-1}				5.892***		6.226**
				(2.286)		(2.586)
FIE_{-1}				-9.114**		-11.972***
				(3.710)		(4.065)
FMD_{-1}					2.958**	3.063*
					(1.473)	(1.717)
FMA_{-1}					-0.437	-1.162
					(1.527)	(1.752)
FME_{-1}					0.790	0.545
					(0.987)	(1.051)
Macro-control variables						
Spread	-0.109	-0.099	-0.101	-0.148*	-0.118*	-0.184**
	(0.068)	(0.069)	(0.066)	(0.082)	(0.070)	(0.082)
Output growth	-0.149**	-0.131*	-0.156**	-0.109	-0.173**	-0.149**
	(0.070)	(0.070)	(0.070)	(0.073)	(0.072)	(0.075)
Relevant Statistics						
AIC	331.35	335.03	333.4	327.13	336.00	325.30
$Pseudo-R^2$	0.595	0.590	0.592	0.607	0.594	0.619
#Observations	851	851	851	851	851	851

Notes: This table reports the estimates obtained from the dynamic logit models (1) to (6) for the panel of 23 countries from $1980-2016$. Standard errors are reported within brackets below the estimates. $***$, $**$ and $*$ report significance at 99%, 95% and 90%.

Table 2.3: Estimation Results - Emerging Markets

Model	(1)	(2)	(3)	(4)	(5)	(6)
constant	-2.386***	-2.256***	-2.696***	-2.367***	-2.763***	-2.791***
	(0.690)	(0.764)	(0.622)	(0.884)	(0.653)	(0.923)
lag binary	3.597***	3.590***	3.621***	3.579***	3.548***	3.527***
	(0.257)	(0.258)	(0.257)	(0.265)	(0.260)	(0.268)
Financial development variables						
FD_{-1}	-2.223					
	(1.454)					
FI_{-1}		-1.861				
		(1.369)				
FM_{-1}			-1.693			
			(1.261)			
FID_{-1}				8.422***		11.339***
				(2.969)		(3.302)
FIA_{-1}				-10.139***		-9.247***
				(2.621)		(2.666)
FIE_{-1}				2.334*		2.338*
				(1.239)		(1.229)
FMD_{-1}					-1.020	-1.549
					(1.925)	(2.109)
FMA_{-1}					-2.818	-3.402*
					(1.727)	(1.964)
FME_{-1}					0.866	0.899
					(0.755)	(0.784)
Macro-control variables						
Spread	-0.027	-0.036	-0.023	-0.048	-0.023	-0.030
	(0.050)	(0.049)	(0.050)	(0.053)	(0.053)	(0.056)
Output growth	-0.077***	-0.079***	-0.076***	-0.082***	-0.079***	-0.081***
	(0.023)	(0.023)	(0.023)	(0.024)	(0.023)	(0.024)
Relevant Statistics						
AIC	563.73	564.22	564.23	551.48	561.92	549.7
$Pseudo-R^2$	0.514	0.514	0.514	0.531	0.520	0.539
#Observations	1702	1702	1702	1702	1702	1702

Notes: This table reports the estimates obtained from the dynamic logit models (1) to (6) for the panel of 46 countries from $1980-2016$. Standard errors are reported within brackets below the estimates. $***$, $**$ and $*$ report significance at 99%, 95% and 90%.

Table 2.4: Estimation Results - Low-Income Countries

Model	(1)	(2)	(3)	(4)	(5)	(6)
constant	-2.349***	-2.443***	-2.644***	-1.479	-2.613***	-1.439
	(0.905)	(0.903)	(0.698)	(1.035)	(0.700)	(1.064)
lag binary	3.581***	3.583***	3.606***	3.347***	3.578***	3.331***
	(0.314)	(0.315)	(0.314)	(0.319)	(0.315)	(0.322)
			Financial development variables			
FD_{-1}	-5.216					
	(7.581)					
FI_{-1}		-2.292				
		(4.232)				
FM_{-1}			-7.839			
			(12.625)			
FID_{-1}				7.953		8.289
				(6.545)		(7.112)
FIA_{-1}				-66.472**		-64.489**
				(26.328)		(26.065)
FIE_{-1}				2.008		1.873*
				(1.625)		(1.641)
FMD_{-1}					-0.804	0.047
					(5.377)	(6.762)
FMA_{-1}					-42.402	-24.078
					(123.610)	(147.837)
FME_{-1}					-19.683	-19.087
					(19.790)	(22.410)
			Macro-control variables			
Model	(1)	(2)	(3)	(4)	(5)	(6)
Output growth	-0.112***	-0.112***	-0.111***	-0.101***	-0.107***	-0.099***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
			Relevant Statistics			
AIC	385.43	385.61	385.51	379.28	388.36	384.16
$Pseudo-R^2$	0.463	0.463	0.463	0.480	0.465	0.482
#Observations	1073	1073	1073	1073	1073	1073

Notes: This table reports the estimates obtained from the dynamic logit models (1) to (6) for the panel of 29 countries from 1980 − 2016. Because of data availability, the term spread is not included for this group of countries. Standard errors are reported within brackets below the estimates. $***$, $**$ and $*$ report significance at 99%, 95% and 90%.

We observe that for all of them, financial institution development has a destabilizing impact. Even if the aggregate variables FD, FI and FM are only significant in the case of the advanced countries' cluster, subindices are significant for the three clusters. From analysis of models (4) to (6), it appears that financial institution development is more important for predicting banking crises than financial markets are. This fact highlights the key role played by financial intermediates in the occurrence of banking turmoil.

Still, the analysis of Tables 2.2 to 2.4 reveal important differences between the three groups of countries. In advanced economies, depth (FID) and access (FIA) are actually destabilizing and increase the probability of a future occurrence of a crisis. This result suggests that regulators in these countries should not only control the amount invested/borrowed or the efficiency of financial institutions but also have a particular monitoring of the access to financial institutions.

In contrast, for LIDC and EM, only the financial institution depth constitutes a leading indicator for future crises. The access to financial institutions increases financial stability and reduces the probability of financial crises a year ahead. Such a result is clearly linked to the problem of financial inclusion. In LIDC/EM countries, inequality in the access to indirect finance constitutes a constraint for the real economy and thus increases the probability of occurrence of a banking crisis. A difference between the groups appears also with respect to the macro-fundamental variables. We observe that while term spread and output growth are important for the occurrence of future banking crisis in advanced economies, only the last term matters in EM. Such a finding could be due to the low level of maturity in the least advanced countries. For LIDC, we do not include the spread term because of data availability, which is not very important as most of these countries are small open economies that are thus interest rate takers with unmature financial capital markets.

2.6 Robustness checks

To assess our results, several robustness checks are proposed:

In the first three robustness checks, we consider a nonlinear model for which financial development interacts with the regimes of the nonperforming loans (NPL) and the capital account openness (KAO, measured here by the method of Chinn and Ito, 2006). These variables z_j, with $j = 1, ..., 3$, are included in the model via an index $\mathbb{1}_z$, which takes the value of 1 for a particular year t and a particular country i if $z_{i,t} > median(z_i)$ and 0 otherwise. Model (1) takes the following form:

$$\Pr(y_{it} = 1) = F(\alpha.y_{i,t-1} + x_{i,t-1}\beta + \delta\pi_{i,t-1} + \gamma.x_{i,t-1}.\mathbb{1}(z_{j,i,t-1}) + \eta_i). \qquad (2.3)$$

A γ associated with z_{NPL} that is positive and significantly different from 0 indicates that a high amount of net performing loans would amplify banking instability brought by financial development. Similarly, a positive coefficient for the interactive term z_{KAO} suggests that the more open the capital account of a country is, the more destabilizing financial development is. We also consider the exchange rate regime (Err) measured via the method of Levy-Yeyati and Sturzenegger (2005), (2016). A dummy is then simultaneously introduced for fixed and flexible exchange rate regimes.[10]

[10] The sum of the dummy variables for fixed and flexible exchange rates does not amount to one as it exists in some cases of undefined or intermediate exchange rate regimes.

Table 2.5 reports the results obtained for the previous robustness checks. It turns out that the introduction of interactive terms (KAO and NPL) does not affect the relationship between financial development and the banking crisis, as none of the coefficients associated with the interaction term are significantly different from zero at 99%. This result thus signifies that our previous findings hold whatever the degree of capital openness or the amount of nonperforming loans. Similar results are obtained for fixed exchange rates (third panel of Table 2.5). In contrast, it appears that the flexible exchange rate regime affects the previous results. Indeed, in this exchange rate regime, access to financial institutions becomes destabilizing, whereas financial deepening is stabilizing. This finding can be explained by capital movement on the foreign exchange markets. In this case, only financial institutions' efficiency appears to decrease the occurrence of a banking crisis.

The model is reestimated for the period before 2008 in order to check if the great crisis has structurally modified the relationship between finance and the crises. The results of the estimations are reported in Table 2.6 and do not show a major quantitative difference from those reported in Table 2.1. This observation thus signifies that the impact of financial development on future banking crises is not driven by the great crisis and is quite stable over time.

So far, models have been estimated for a one-lag horizon. In other words, the previous results show how financial development is improving or deteriorating the probability of occurrence of a banking crisis in the coming year. This last robustness check explores this relationship for a horizon of two years and considers hence the following model: $Pr(y_{it} = 1) = F(\eta_i + \alpha y_{i,t-2} + x_{i,t-2}\beta + \delta\pi_{i,t-2})$. The results of the estimations are reported in Table 2.7. Again, the results obtained are similar to those reported in Table 2.1. These findings as well as the previous one (for the pre-crisis period) support the idea that the link obtained between financial development and the probability of a banking crisis is structural: this holds whatever the sample and the horizon considered. This outcome clearly calls for structural regulation policies, which should be independent of the business cycle or a specific temporary event.

Table 2.5: Estimation Results - Kaopen - NPL - Exchange Rate

	Kaopen				NPL				Exchange Rate							
Model	(1)	(2)	(3)	(6)	(1)	(2)	(3)	(6)	(1)		(2)		(3)		(6)	
constant	-3.137***	-3.188***	-2.933***	-3.235***	-27.041	-25.981	-23.552	-26.870	-3.176***		-3.303***		-2.899***		-3.699***	
	(0.687)	(0.706)	(0.681)	(0.756)	(6487.027)	(6655.876)	(6498.003)	(6433.435)	(0.696)		(0.712)		(0.693)		(0.759)	
lag binary	3.859***	3.896***	3.832***	3.783***	3.99***	3.645***	3.990***	3.980***	3.898***		3.925***		3.876***		3.843***	
	(0.179)	(0.179)	(0.179)	(0.184)	(0.486)	0.438	(0.481)	(0.530)	(0.180)		(0.179)		(0.179)		(0.187)	
	Financial development variables															
FD_{-1}	3.040***				21.155***				4.399***							
	(1.063)				(4.469)				(1.124)							
FI_{-1}		1.849				10.690**					3.369***					
		(1.213)				(4.375)					(1.240)					
FM_{-1}			2.517***				14.436***						3.472***			
			(0.792)				(3.060)						(0.919)			
FID_{-1}				4.305**				-1.630							8.852***	
				(2.002)				(6.079)							(2.355)	
FIA_{-1}				-1.331				1.408							-7.423***	
				(1.537)				(4.468)							(2.021)	
FIE_{-1}				-0.329				1.999							0.944	
				(0.978)				(4.590)							(1.008)	
FMD_{-1}				0.931				12.491***							3.333**	
				(1.145)				(4.293)							(1.684)	
FMA_{-1}				-1.147				-1.863							-1.507	
				(1.227)				(3.918)							(1.468)	
FME_{-1}				0.947				4.717**							0.531	
				(0.686)				(1.934)							(0.774)	
	Macro-control variables															
Spread	-0.059*	-0.049	-0.065*	-0.061	-0.114	-0.144	-0.113	-0.161	-0.070*		-0.059*		-0.073**		-0.088**	
	(0.035)	(0.034)	(0.035)	(0.037)	(0.097)	(0.092)	(0.102)	(0.107)	(0.037)		(0.035)		(0.037)		(0.044)	
Output growth	-0.092***	-0.091***	-0.094***	-0.098***	-0.323***	-0.259***	-0.329***	-0.355***	-0.099***		-0.096***		-0.102***		-0.102***	
	(0.019)	(0.019)	(0.019)	(0.019)	(0.094)	(0.088)	(0.095)	(0.107)	(0.019)		(0.019)		(0.019)		(0.020)	
	Dummy-control variables															
TypeofRate									Fix	Float	Fix	Float	Fix	Float	Fix	Float
$FD_1.z_{-1}$	-0.617				-0.376				-1.018	-0.115						
	(0.750)				(0.627)				(0.751)	(0.557)						
$FI_1.z_{-1}$		-0.077				-0.184					-0.837	0.230				
		(0.623)				(0.559)					(0.605)	(0.450)				
$FM_1.z_{-1}$			-0.803				-0.948						-1.083	-0.214		
			(0.829)				(0.675)						(0.888)	(0.648)		
$FID_1.z_{-1}$				-1.456				6.591*							-1.064	-3.842*
				(2.501)				(3.504)							(2.389)	(2.032)
$FIA_1.z_{-1}$				-3.049*				2.542							-1.437	6.235***
				(1.736)				(2.560)							(2.275)	(1.610)
$FIE_1.z_{-1}$				1.163				1.687							0.106	-2.211***
				(0.731)				(2.857)							(0.784)	(0.839)
$FMD_1.z_{-1}$				4.246**				-6.346*							-0.405	-0.796
				(2.153)				(3.656)							(2.303)	(1.812)
$FMA_1.z_{-1}$				-2.544				-3.403*							1.342	0.170
				(1.870)				(1.975)							(2.233)	(1.356)
$FME_1.z_{-1}$				0.306				-1.794							-0.109	-0.091
				(1.166)				(1.619)							(1.342)	(0.911)
	Relevant Statistics															
AIC	1175.2	1181.7	1172.8	1172.8	330.32	350.1	329.26	333.61	1202.6		1213.1		1201.2		1186.8	
$Pseudo-R^2$	0.512	0.508	0.513	0.523	0.715	0.687	0.717	0.739	0.522		0.517		0.523		0.544	
#Observations	3256	3256	3256	3256	976	976	976	976	3162		3162		3162		3162	

Notes: This table reports the estimates obtained from the dynamic logit models (1), (2), (3), and (6) for the panel of: 88 countries from $1980-2016$ - Kaopen, 61 countries from $2001-2016$ - NPL, 93 countries from $1980-2013$ - Exchange Rate. Standard errors are reported within brackets below the estimates. $***$, $**$ and $*$ report significance at 99%, 95% and 90%.

Table 2.6: Estimation Results 1980 − 2008

Model	(1)	(2)	(3)	(4)	(5)	(6)
constant	-2.938***	-3.186***	-2.652***	-3.433***	-2.747***	-3.471***
	(0.674)	(0.695)	(0.667)	(0.726)	(0.669)	(0.733)
lag binary	3.437***	3.415***	3.420***	3.408***	3.437***	3.416***
	(0.195)	(0.195)	(0.194)	(0.197)	(0.195)	(0.198)
			Financial development variables			
FD_{-1}	4.131***					
	(1.086)					
FI_{-1}		4.132***				
		(1.341)				
FM_{-1}			2.730***			
			(0.761)			
FID_{-1}				9.694***		9.234***
				(2.077)		(2.342)
FIA_{-1}				-4.337**		-3.652
				(2.209)		(2.248)
FIE_{-1}				0.348		0.244
				(0.929)		(0.924)
FMD_{-1}					3.423***	2.217*
					(1.142)	(1.250)
FMA_{-1}					-1.274	-2.538*
					(1.122)	(1.304)
FME_{-1}					0.047	-0.063
					(0.643)	(0.666)
			Macro-control variables			
Spread	-0.100**	-0.085**	-0.102**	-0.110**	-0.107**	-0.112**
	(0.044)	(0.043)	(0.044)	(0.053)	(0.046)	(0.054)
Output growth	-0.090***	-0.084***	-0.092***	-0.089***	-0.095***	-0.093***
	(0.019)	(0.018)	(0.019)	(0.019)	(0.019)	(0.019)
			Relevant Statistics			
AIC	1140.7	1145.6	1142.6	1129.4	1141.6	1129.9
$Pseudo - R^2$	0.454	0.451	0.453	0.463	0.456	0.466
#Observations	2842	2842	2842	2842	2842	2842

Notes: This table reports the estimates obtained from the dynamic logit models (1) to (6) for the panel of 98 countries from 1980 − 2008. Standard errors are reported within brackets below the estimates. $***$, $**$ and $*$ report significance at 99%, 95% and 90%.

Table 2.7: Estimation Results with two lags 1980 - 2016

Model	(1)	(2)	(3)	(4)	(5)	(6)
constant	-2.517***	-2.608***	-2.293***	-2.741***	-2.352***	-2.683***
	(0.573)	(0.585)	(0.571)	(0.613)	(0.571)	(0.618)
lag binary	2.165***	2.214***	2.134***	2.107***	2.113***	2.069***
	(0.146)	(0.145)	(0.147)	(0.171)	(0.147)	(0.149)
			Financial development variables			
FD_{-2}	3.137***					
	(0.719)					
FI_{-2}		2.261***				
		(0.855)				
FM_{-2}			2.502***			
			(0.531)			
FID_{-2}				6.903***		5.953***
				(2.128)		(1.524)
FIA_{-2}				-3.184***		-3.254***
				(1.079)		(1.122)
FIE_{-2}				-0.053		-0.132
				(0.713)		(0.708)
FMD_{-2}					2.708***	1.849**
					(0.763)	(0.834)
FMA_{-2}					-1.285	-1.771**
					(0.801)	(0.890)
FME_{-2}					0.403	0.484
					(0.450)	(0.456)
			Macro-control variables			
Spread	-0.068**	-0.055*	-0.072**	-0.080**	-0.078**	-0.088**
	(0.030)	(0.029)	(0.031)	(0.037)	(0.032)	(0.036)
Output growth	-0.096***	-0.093***	-0.099***	-0.093***	-0.101***	-0.098***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
			Relevant Statistics			
AIC	1707.6	1720.8	1703.8	1693.4	1700.2	1687.3
$Pseudo - R^2$	0.322	0.316	0.323	0.330	0.327	0.335
#Observations	3626	3626	3626	3626	3626	3626

Notes: This table reports the estimates obtained from the dynamic logit models (1) to (6) for the panel of 98 countries from $1980 - 2016$. Standard errors are reported within brackets below the estimates. $***$, $**$ and $*$ report significance at 99%, 95% and 90%.

2.7 Conclusion and Policy Implications

This paper assesses whether and how financial development triggers the occurrence of banking crises. This question clearly fits the literature evaluating whether more finance is always good for growth and financial stability.

The innovation of the paper is twofold: First, it considers a database, decomposing financial development into its main components, i.e., access, depth and efficiency, and covering most of the world's economies (98 countries). Second, this study relies on a dynamic logit panel model, which includes past crisis observations in order to obtain unbiased estimators as well as a fixed effect to address unobserved heterogeneity.

It appears that financial development, from an institutional and to a lesser extent a market dimension, increases the probability of occurrence of a crisis within a one- to two-year horizon. Still, the explosion of the fintech industry (mobile payment, cryptocurrency, and offshore banking, to quote but a few) is such a matter of concern for regulators and supervisors, who should adjust their macroprudential rules accordingly. Going deeper, the paper indicates that the destabilizing dimension of financial development is different in advanced and emerging/low income countries. For advanced countries, we observe that financial access and depth are destabilizing, whereas efficiency reduced the future occurrence of a banking crisis. In contrast, for emerging countries, financial access is stabilizing, and depth/efficiency is not. We thus observe that the impact of financial development on stability is not homogeneous; rather, it varies with its component and the country under consideration.

Such results tend to support the papers which categorize private credit growth (as a percentage of gdp) as a leading indicator for banking crises (e.g. Demirgüç–Kunt and Detragiache, 1998). They also explain why conclusions vary with the countries considered and the credit proxy used. It hence suggests to analyse separately countries at different development stages. It also highlights that studies should go beyond the simple analysis of private credit and consider other dimensions for financial development.

These findings have also important consequences for macroprudential policies. First, financial stability assessments (such as the Financial Sector Assessment Program, FSAP, jointly conducted by the IMF and the World Bank) should include a shock associated with the degree of financial development (as well as with each of these components: access, depth and efficiency). By doing so, the financial sector's vulnerabilities would be better assessed, particularly in front of the surge of financial innovations. Second, financial regulation (in particular, the Basel agreements for the banking sector) should take into account the emerging markets' specificities compared to advanced countries. For example, higher capital requirements should be imposed on banks in advanced economies to smooth the increase in financial access and depth, whereas this should not be the case for emerging markets' financial institutions. Similarly, regulators should encourage higher efficiencies in the financial institutions. It is thus obvious that including these findings would definitively change regulation in the current times of high credit for advanced countries.

References

Arcand, J. L., E. Berkes, and U. Panizza (2015). “Too much finance?” In: *Journal of Economic Growth* 20.2, pp. 105–148.

Barrell, R. et al. (2011). “How idiosyncratic are banking crises in OECD countries?” In: *National Institute Economic Review* 216.1, R53–R58.

Bencivenga, V. R. and B. D. Smith (1991). “Financial intermediation and endogenous growth”. In: *The review of economic studies* 58.2, pp. 195–209.

Bernanke, S., A. S. Blinder, et al. (1988). “Credit, money and aggregate demand”. In: *American Economic Review* 78.2, pp. 435–439.

Bordo, M. D. and C. M. Meissner (2012). “Does inequality lead to a financial crisis?” In: *Journal of International Money and Finance* 31.8, pp. 2147–2161.

Borio, C. E. and M. Drehmann (2009). *Towards an operational framework for financial stability:'fuzzy'measurement and its consequences.* BIS Working Paper.

Borio, C. E. and P. W. Lowe (2002). *Asset prices, financial and monetary stability: exploring the nexus.* BIS working paper.

Bunda, I. and M. Ca'Zorzi (2010). “Signals from housing and lending booms”. In: *Emerging Markets Review* 11.1, pp. 1–20.

Büyükkarabacak, B. and N. T. Valev (2010). “The role of household and business credit in banking crises”. In: *Journal of Banking & Finance* 34.6, pp. 1247–1256.

Candelon, B., A. Carare, et al. (2020). “Globalization and the New Normal”. In: *International Economics* 161.3, pp. 139–158.

Candelon, B., E.-I. Dumitrescu, and C. Hurlin (2014). “Currency crisis early warning systems: Why they should be dynamic”. In: *International Journal of Forecasting* 30.4, pp. 1016–1029.

Candelon, B., E.-I. Dumitrescu, C. Hurlin, and F. C. Palm (2013). *Multivariate Dynamic Probit Models: An Application to Financial Crises Mutation', VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims (Advances in Econometrics, Volume 32).*

Carro, J. M. (2007). “Estimating dynamic panel data discrete choice models with fixed effects”. In: *Journal of Econometrics* 140.2, pp. 503–528.

Cecchetti, S. G. and E. Kharroubi (2012). *Reassessing the impact of finance on growth.* BIS working paper.

Chinn, M. D. and H. Ito (2006). “What matters for financial development? Capital controls, institutions, and interactions”. In: *Journal of development economics* 81.1, pp. 163–192.

Čihák, M. (2007). “Systemic loss: A measure of financial stability”. In: *Czech Journal of Economics and Finance* 57.1-2, pp. 5–26.

Čihák, M. et al. (2012). “Benchmarking financial systems around the world”. In: *World Bank Policy Research Working Paper* 6175.

Davis, E. P. and D. Karim (2008a). “Comparing early warning systems for banking crises”. In: *Journal of Financial stability* 4.2, pp. 89–120.

— (2008b). “Could early warning systems have helped to predict the sub-prime crisis?” In: *National Institute Economic Review* 206.1, pp. 35–47.

Davis, E. P., D. Karim, and I. Liadze (2011). “Should multivariate early warning systems for banking crises pool across regions?” In: *Review of World Economics* 147.4, pp. 693–716.

Dell’Ariccia, G. and R. Marquez (2004). “Information and bank credit allocation”. In: *Journal of Financial Economics* 72.1, pp. 185–214.

Demirgüç-Kunt, A. and E. Detragiache (1998). “The determinants of banking crises in developing and developed countries”. In: *Staff Papers* 45.1, pp. 81–109.

— (2000). “Monitoring banking sector fragility: a multivariate logit approach”. In: *The World Bank Economic Review* 14.2, pp. 287–307.

— (2002). “Does deposit insurance increase banking system stability? An empirical investigation”. In: *Journal of monetary economics* 49.7, pp. 1373–1406.

— (2005). “Cross-country empirical studies of systemic bank distress: a survey”. In: *National Institute Economic Review* 192.1, pp. 68–83.

Drehmann, M., C. E. Borio, and K. Tsatsaronis (2011). *Anchoring countercyclical capital buffers: the role of credit aggregates.* BIS working paper.

Drehmann, M. and M. Juselius (2014). “Evaluating early warning indicators of banking crises: Satisfying policy requirements”. In: *International Journal of Forecasting* 30.3, pp. 759–780.

Duca, M. L. and T. A. Peltonen (2013). “Assessing systemic risks and predicting systemic events”. In: *Journal of Banking & Finance* 37.7, pp. 2183–2195.

Goldsmith, R. W. (1970). “Financial structure and development.” In: *The Economic journal* 80, pp. 365–367.

Greenwood, J. and B. Jovanovic (1990). “Financial development, growth, and the distribution of income”. In: *Journal of political Economy* 98.5, Part 1, pp. 1076–1107.

Hahm, J.-h., H. S. Shin, and K. Shin (2013). “Noncore bank liabilities and financial vulnerability”. In: *Journal of Money, Credit and Banking* 45.s1, pp. 3–36.

Jordà, Ò., M. Schularick, and A. M. Taylor (2011). “Financial crises, credit booms, and external imbalances: 140 years of lessons”. In: *IMF Economic Review* 59.2, pp. 340–378.

Kaminsky, G. L. and C. M. Reinhart (1999). “The twin crises: the causes of banking and balance-of-payments problems”. In: *American economic review* 89.3, pp. 473–500.

Kaminsky, G., S. Lizondo, and C. M. Reinhart (1998). “Leading indicators of currency crises”. In: *Staff Papers* 45.1, pp. 1–48.

Kauko, K. (2014). “How to foresee banking crises? A survey of the empirical literature”. In: *Economic Systems* 38.3, pp. 289–308.

Kauppi, H. and P. Saikkonen (2008). “Predicting US recessions with dynamic binary response models”. In: *The Review of Economics and Statistics* 90.4, pp. 777–791.

Keeley, M. C. (1990). “Deposit insurance, risk, and market power in banking”. In: *The American economic review*, pp. 1183–1200.

Kiyotaki, N. and J. Moore (1997). “Credit cycles”. In: *Journal of political economy* 105.2, pp. 211–248.

Laeven, L. and F. Valencia (2013). “Systemic banking crises database.” In: *IMF Economic Review* 61.2, pp. 225–270.

Levine, R. (2005). “Finance and growth: theory and evidence”. In: *Handbook of economic growth* 1, pp. 865–934.

Levy-Yeyati, E. L. and F. Sturzenegger (2016). *Classifying exchange rate regimes: 15 years later*. HKS Working Paper.

Levy-Yeyati, E. and F. Sturzenegger (2005). “Classifying exchange rate regimes: Deeds vs. words”. In: *European economic review* 49.6, pp. 1603–1635.

Lucas, R. E. (1988). “On the Mechanics of Economic Development”. In: *Journal of Monetary Economics* 22.1, pp. 3–42.

Mathonnat, C. and A. Minea (2018). “Financial development and the occurrence of banking crises”. In: *Journal of Banking & Finance* 96, pp. 344–354.

McKinnon, R. and E. Shaw (1973). *Financial deepening in economic development.* Washington, Brookings Institution.

Minsky, H. P. et al. (1986). *Stabilizing an unstable economy: The lessons for industry, finance and government.*

Rajan, R. G. and L. Zingales (1996). *Financial dependence and growth.* Tech. rep. National bureau of economic research.

Robinson, J. (1952). *The Generalization of the General Theory.* The Rate of Interest and Other Essays (MacMillan, London).

Rose, A. K. and M. M. Spiegel (2011). “Cross-country causes and consequences of the crisis: An update”. In: *European Economic Review* 55.3, pp. 309–324.

Rousseau, P. L. and P. Wachtel (2000). “Equity markets and growth: cross-country evidence on timing and outcomes, 1980–1995”. In: *Journal of Banking & Finance* 24.12, pp. 1933–1957.

Sahay, R. et al. (2015). “Rethinking financial deepening: Stability and growth in emerging markets”. In: *Revista de Economía Institucional* 17.33, pp. 73–107.

Saint-Paul, G. (1992). “Technological choice, financial markets and economic development”. In: *European Economic Review* 36.4, pp. 763–781.

Schularick, M. and A. M. Taylor (2012). “Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008”. In: *American Economic Review* 102.2, pp. 1029–61.

Schumpeter, J. (1934). “the theory of Economic Development, tr”. In: *By Redvers Opie, Harvard Economic Studies* 40, p. 20.

Shaw, E. S. (1973). *Financial deepening in economic development.* New York, NY (USA) Oxford Univ. Press.

Svirydzenka, K. (2016). *Introducing a new broad-based index of financial development.* International Monetary Fund.

Von Hagen, J. and T.-k. Ho (2007). “Money market pressure and the determinants of banking crises”. In: *Journal of Money, Credit and Banking* 39.5, pp. 1037–1066.

Appendices

Appendix 1 - Financial Development indicator construction

Table 2.8: Variables used for building the Financial Development Index

Category	**Indicator**	**Index**
	Financial Institutions	FI
Depth	- Private-sector credit to GDP - Pension fund assets to GDP - Mutual fund assets to GDP - Insurance premiums, life and non-life to GDP	FID
Access	- Bank branches per 100,000 adults - ATMs per 100,000 adults	FIA
Efficiency	- Net interest margin - Lending-deposits spread - Non-interest income to total income - Overhead costs to total assets - Return on assets - Return on equity	FIE
	Financial Markets	FM
Depth	- Stock market capitalization to GDP - Stocks traded to GDP - International debt securities of government to GDP - Total debt securities of financial corporations to GDP - Total debt securities of nonfinancial corporations to GDP	FMD
Access	- Percent of market capitalization outside of top 10 largest companies - Total number of issuers of debt (domestic and external, nonfinancial and financial corporations)	FMA
Efficiency	- Stock market turnover ratio (stocks traded to capitalization)	FMD

Sources: Sahay et. al (2015) and Svirydzenka (2016)

Appendix 2 - Data Description - Indices

Table 2.9: Average indices from 1980 to 2016

	AM	EM	LIDC	Global
FD	0.63	0.29	0.12	0.31
FI	0.74	0.38	0.21	0.40
FM	0.50	0.20	0.02	0.21
FID	0.63	0.23	0.10	0.28
FIA	0.69	0.26	0.06	0.30
FIE	0.70	0.60	0.52	0.58
FMD	0.50	0.16	0.05	0.20
FMA	0.48	0.21	0.00	0.21
FME	0.51	0.22	0.01	0.22

Notes: This table reports the averages of each index from $1980 - 2016$. These averages are calculated for Advanced Markets (AM), Emerging Markets (EM) Low-Income Developing Countries (LIDC) and for all countries (Global).

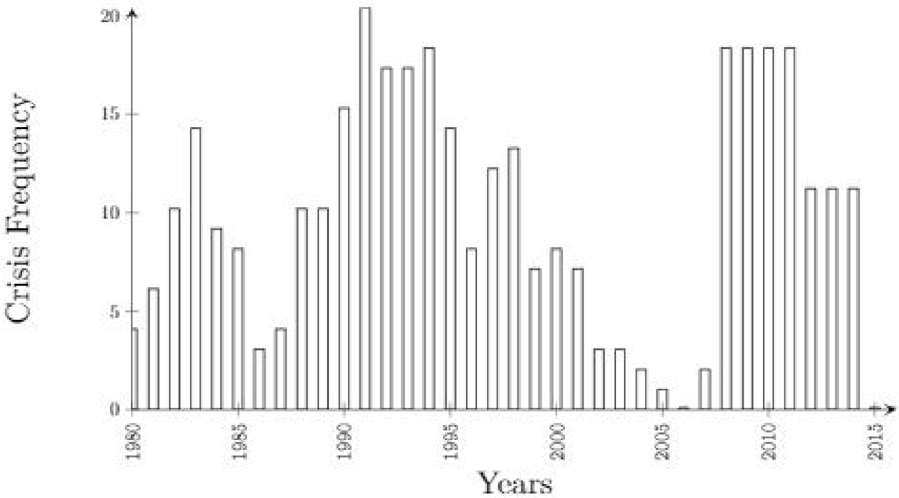
Appendix 3 - Data Description

Table 2.10: Country Data

Country	IMF Region	Country Group	Banking Crisis
Algeria	Middle East and Central Asia	EM	1990, 1991, 1992, 1993, 1994
Angola	Africa	EM	
Argentina	Western Hemisphere	EM	1980, 1981, 1982, 1989, 1990, 1991, 1995, 2001, 2002, 2003
Australia	Asia and Pacific	AM	
Austria	Europe	AM	2008, 2009, 2010, 2011
Bahamas, The	Western Hemisphere	EM	
Bangladesh	Asia and Pacific	LIDC	1987
Barbados	Western Hemisphere	EM	
Belgium	Europe	AM	2008, 2009, 2010, 2011, 2012, 2013, 2014
Belize	Western Hemisphere	EM	
Benin	Africa	LIDC	1988, 1989, 1990, 1991, 1992
Bolivia	Western Hemisphere	LIDC	1986, 1994
Botswana	Africa	EM	
Brazil	Western Hemisphere	EM	1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998
Burkina Faso	Africa	LIDC	1990, 1991, 1992, 1993, 1994
Burundi	Africa	LIDC	1994, 1995, 1996, 1997, 1998
Cameroon	Africa	LIDC	1987, 1988, 1989, 1990, 1991, 1995, 1996, 1997
Canada	Western Hemisphere	AM	
Chad	Africa	LIDC	1983, 1992, 1993, 1994, 1995, 1996
Chile	Western Hemisphere	EM	1981, 1982, 1983, 1984, 1985
Colombia	Western Hemisphere	EM	1982, 1998, 1999, 2000
Congo, Dem. Rep.	Africa	LIDC	1983, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998
Congo, Rep.	Africa	LIDC	1992, 1993, 1994
Costa Rica	Western Hemisphere	EM	1987, 1988, 1989, 1990, 1991, 1994, 1995
Cote d'Ivoire	Africa	LIDC	1988, 1989, 1990, 1991, 1992
Denmark	Europe	AM	2008, 2009, 2010, 2011, 2012, 2013, 2014
Dominican Republic	Western Hemisphere	EM	2003, 2004
Ecuador	Western Hemisphere	EM EM	1982, 1983, 1984, 1985, 1986, 1998, 1999, 2000, 2001, 2002
Egypt, Arab Rep.	Middle East and Central Asia	EM	1980
El Salvador	Western Hemisphere	EM	1989, 1990
Fiji	Asia and Pacific	EM	
Finland	Europe	AM	1991, 1992, 1993, 1994, 1995
France	Europe	AM	2008, 2009, 2010, 2011, 2012, 2013, 2014
Gabon	Africa	EM	
Germany	Europe	AM	2008, 2009, 2010, 2011
Ghana	Africa	LIDC	1982, 1983
Greece	Europe	AM	2008, 2009, 2010, 2011, 2012, 2013, 2014
Guatemala	Western Hemisphere	EM	
Guyana	Western Hemisphere	LIDC	1993
Honduras	Western Hemisphere	LIDC	
Hungary	Europe	EM	1991, 1992, 1993, 1994, 1995, 2008, 2009, 2010, 2011, 2012, 2013, 2014
Iceland	Europe	AM	2008, 2009, 2010, 2011, 2012, 2013, 2014
India	Asia and Pacific	EM	1993
Indonesia	Asia and Pacific	EM	1997, 1998, 1999, 2000, 2001
Ireland	Europe	AM	2008, 2009, 2010, 2011
Israel	Europe	EM	
Italy	Europe	AM	2008, 2009, 2010, 2011, 2012, 2013, 2014
Japan	Asia and Pacific	AM	1997, 1998, 1999, 2000, 2001
Kenya	Africa	LIDC	1985, 1992, 1993, 1994

Country	IMF Region	Country Group	Banking Crisis
Korea, Rep.	Asia and Pacific	EM	1997, 1998
Kuwait	Middle East and Central Asia	EM	1982, 1983, 1984, 1985
Lesotho	Africa	LIDC	
Luxembourg	Europe	AM	2008, 2009, 2010, 2011
Madagascar	Africa	LIDC	1988
Malawi	Africa	LIDC	
Malaysia	Asia and Pacific	EM	1997, 1998, 1999
Mali	Africa	LIDC	1987, 1988, 1989, 1990, 1991
Mauritius	Africa	EM	
Mexico	Western Hemisphere	EM	1981, 1982, 1983, 1984, 1985, 1994, 1995, 1996
Morocco	Middle East and Central Asia	EM	1980, 1981, 1982, 1983, 1984
Nepal	Asia and Pacific	LIDC	1988
Netherlands	Europe	AM	2008, 2009, 2010, 2011, 2012, 2013, 2014
New Zealand	Asia and Pacific	AM	
Nicaragua	Western Hemisphere	LIDC	1990, 1991, 1992, 1993, 2000, 2001
Niger	Africa	LIDC	1983, 1984, 1985
Norway	Europe	AM	1991, 1992, 1993
Oman	Middle East and Central Asia	EM	
Pakistan	Asia and Pacific	EM	
Panama	Western Hemisphere	EM	1988, 1989
Papua New Guinea	Asia and Pacific	LIDC	
Paraguay	Western Hemisphere	EM	1995
Peru	Western Hemisphere	EM	1983
Philippines	Asia and Pacific	EM	1983, 1984, 1985, 1986, 1997, 1998, 1999, 2000, 2001
Poland	Europe	EM	1992, 1993, 1994
Portugal	Europe	AM	2008, 2009, 2010, 2011, 2012, 2013, 2014
Romania	Europe	EM	1990, 1991, 1992
Rwanda	Africa	LIDC	
Senegal	Africa	LIDC	1988, 1989, 1990, 1991
Seychelles	Africa	EM	
Sierra Leone	Africa	LIDC	1990, 1991, 1992, 1993, 1994
Singapore	Asia and Pacific	EM	
South Africa	Africa	EM	
Spain	Europe	AM	1980, 1981, 2008, 2009, 2010, 2011, 2012, 2013, 2014
Sri Lanka	Asia and Pacific	EM	1989, 1990, 1991
St. Vincent and the Grenadines	Western Hemisphere	LIDC	
Sudan	Africa	LIDC	
Sweden	Europe	AM	1991, 1992, 1993, 1994, 1995, 2008, 2009, 2010, 2011
Switzerland	Europe	AM	2008, 2009, 2010, 2011
Thailand	Asia and Pacific	EM	1983, 1997, 1998, 1999, 2000
Togo	Africa	LIDC	1993, 1994
Trinidad and Tobago	Western Hemisphere	EM	
Tunisia	Middle East and Central Asia	EM	1991
Turkey	Europe	EM	1982, 1983, 1984, 2000, 2001
United Kingdom	Europe	AM	2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014
United States	Western Hemisphere	AM	1988, 2007, 2008, 2009, 2010, 2011
Uruguay	Western Hemisphere	EM	1981, 1982, 1983, 1984, 1985, 2002, 2003, 2004, 2005
Venezuela, RB	Western Hemisphere	EM	1994, 1995, 1996, 1997, 1998
Zambia	Africa	LIDC	1995, 1996, 1997, 1998

Figure 2.1: Banking Crises's frequency



Notes: This figure represent the yearly frequency of banking crises' occurrence in our sample of 98 countries.

Chapter 3

Does the Yield Curve Signal Recessions? New Evidence from an International Panel Data Analysis.

Authors: Jean-Baptiste Hasse, Quentin Lajaunie

About this chapter

The authors thank Bertrand Candelon, Yin-Wong Cheung, Elena Ivona Dumitrescu, Laurent Ferrara and Yannick Le Pen for helpful comments. This research was performed as part of a research program titled "Risk Management, Investment Strategies and Financial Stability" under the aegis of the Europlace Institute of Finance, a joint initiative with insti7. The usual disclaimer applies.

Abstract

In this paper, we reexamine the predictive power of the yield spread across countries and over time. Using a dynamic panel/dichotomous model framework and a unique dataset covering 13 OECD countries over a period of 45 years, we empirically show that the yield spread signals recessions. This result is robust to different econometric specifications, controlling for recession risk factors and time sampling. Using a new cluster analysis methodology, we present empirical evidence of a partial homogeneity of the predictive power of the yield spread. Our results provide a valuable framework for monitoring economic cycles.

Keywords: Yield Spread; Recession; Panel Binary Model; Cluster Analysis.

3.1 Introduction

In this paper, we reexamine the predictive power of the yield spread in a unique dataset covering 13 OECD countries over a period of 45 years. Using a dichotomous panel model, we estimate the relationship between yield spreads and future recessions, controlling for monetary policy stance and seven other recession risk factors selected from the recent literature. Our results, robust to different econometric specifications and time sampling, confirm that yield spreads signal recessions. In a further analysis, we investigate the homogeneity of the panel-based estimates via a new cluster analysis procedure. A few clusters of countries emerge, and ten countries out of thirteen are concentrated in two clusters only. Such clustering analysis indicates a partial homogeneity of the predictive power of the yield spread across countries. Furthermore, the predictive power of the yield curve appears to be unrelated to central bank policy rates, while cluster distribution could be linked to monetary policy frameworks (inflation targeting or alternative policy frameworks). Both empirical results are of major interest to policymakers who need to anticipate future economic conditions. Indeed, the relationship between the term spread and future recessions is a well-known stylized fact in economics; the yield curve has been monitored to detect recession signs for several decades (Wheelock and Wohar, 2009). However, two issues have recently been raised following the global financial crisis, namely, the homogeneity and the stability of the predictive power of the yield spread across countries and over time. On the one hand, the predictive power of the term spread appears to have declined since the early 1990s: monetary policy changes, the long-term interest rate conundrum and the zero lower bound have been identified as potential roots of this new regime (Chauvet and Potter, 2005). On the other hand, little is known about the predictive power of the yield spread outside the US. Indeed, most of the literature focuses on the US, while only a few authors extend empirical studies to several countries (Bernard and Gerlach, 1998; Ahrens, 2002; Moneta, 2005; Chinn and Kucko, 2015).

Concerns about stability and homogeneity of the predictive power of the yield curve have led to questions regarding the economic roots of the relationship between interest rates' spread and business cycles. The term spread's evolution has long been linked to business cycles and thus has been used as a valuable tool for monitoring such cycles. The cyclical behavior of the yield spread was first documented by Kessel (1971), who investigated the common variation of the term structure of interest rates and business cycles. Specifically, he showed that the yield spread tended to decline immediately before a recession. Similarly, Fama (1986) noted that the shape of the yield curve changed relative to expansion or recession periods. The author argued that this relationship could be consistent with the liquidity preference hypothesis and could be explained in an intertemporal CAPM framework. Considering the Fisher's expectation hypothesis for the term structure of interest rates,[1] Harvey (1988) provided analytical evidence that the yield spread was related to future consumption growth. Using the consumption CAPM (CCAPM) framework, the author empirically tested if expected Treasury bill returns contained information about expected consumption growth. The results indicated that the yield spread had more explanatory power than lagged consumption growth and lagged stock returns. This result supported the idea that investors' expectations of future expansion or recession could impact the shape of the yield curve. Indeed, CCAPM implies that a cyclical consumption growth should induce a cyclical movement in expected returns. Hence, in this framework the predictive power of the yield spread could originate from agents' anticipation of future recessions. More recently, Estrella (2005) built an analytical rational expectations model to investigate the theoretical roots of the usefulness of the yield curve as a predictor of

[1] See Dimand and Betancourt (2012) for a historical perspective.

output growth. The analytical results indicated that the yield curve in predictive relationships was a function of parameters of the monetary policy rule. Thus, the predictive power of the yield curve could not be said to be structural. Specifically, the predictor depended on the form of the monetary policy reaction function, which in turn might depend on explicit policy objectives.

Apart from the above theoretical studies, the vast majority of the literature focuses on assessing the empirical relationship between the term spread and the probability of future recessions. Seminal studies of Estrella and Hardouvelis (1991) and Kauppi and Saikkonen (2008) have enhanced the use of dichotomous models (probit and logit models) in a univariate framework. This empirical literature is mainly focused on the US, even if a few papers use data for other countries. Consequently, univariate and bivariate analyses are the most frequently used frameworks. Our contribution is a reexamination of the predictive power of the yield curve in an international panel data analysis. Specifically, this paper addresses two econometric challenges: (i) moving from a univariate/bivariate binary model analysis to a dynamic binary balanced panel framework, and (ii) proposing an innovative clustering method adapted to our new framework. Finally, our contribution to the literature is threefold: (i) introducing a unique database including a set of eight country-level recession risk factors, and covering 13 industrialized countries over 45 years at a monthly frequency, (ii) confirming the predictive power of the yield curve in an international panel data framework, and (iii) investigating the potential homogeneity of this predictive power across countries.

The rest of this paper is organized as follows. In Section 3.2, we present a review of the literature on the predictive power of the yield curve across countries and over time. In Section 3.3, we introduce a modeling framework to move from the univariate binary model analysis of Kauppi and Saikkonen (2008) to a dynamic panel/dichotomous model framework. Moreover, this framework is accompanied by a proposed methodological extension to the recent literature on clustering analysis (e.g., Zhang, Wang and Zhu, 2019). In Section 3.4, we discuss the empirical results of our international panel data and clustering analysis. Finally, we summarize the usefulness of our new framework for policymakers and economic forecasters and highlight policy implications with respect to the role of monetary policy in the predictive power of the yield spread.

3.2 Literature

3.2.1 Using the yield curve to forecast recessions

In the 1990s, early empirical studies investigated the relationship between the yield curve and future recessions using binary response models in a univariate framework. Among those studies, the seminal paper of Estrella and Hardouvelis (1991) introduced the use of probit models to forecast recessions. Focusing on the US, the authors showed that the yield spread had a greater predictive power than did the index of leading indicators and survey forecasts. Subsequently, Bernard and Gerlach (1998) and Estrella and Mishkin (1998) extended their findings by analyzing the yield curves in several countries. Similarly, Moneta (2005) focused on European countries to compare the predictive power of several yield spreads of different maturities. The results indicated that the "10 y minus 3 m" yield spread appeared to be the most useful indicator for predicting recession in the medium and long term (two quarters and eight quarters, respectively). In this strand of research, a consensus has emerged that binary response models based on the shape of the yield curve are useful tools for predicting recessions in the US. Specifically, probit models (Dotsey, 1998; Estrella, Rodrigues and Schich, 2003; Wright, 2006; Rosenberg and Maurer, 2008) and logit models (Sensier et al., 2004; Moneta, 2005) have all been used in the static and univariate setting.[2] The introduction of a dynamic approach by Kauppi and Saikkonen (2008) confirmed the usefulness of the yield spread in forecasting recessions (see also Duarte, Venetis and Paya, 2005; Nyberg, 2010; Ng, 2012, and other studies). The empirical results of the researchers' econometric specification provided additional evidence of usefulness of the yield spread for forecasting recessions in the US.

While a proper econometric specification is important, the choice of control variables is also crucial. Exploring this, Wright (2006) shows that using both the level of central banks' rates and term spreads can result in better predictive performance than can using term spreads alone. Indeed, this variable can be used to disentangle the origins of yield

[2] Wright (2006) and King, Levin and Perli (2007) are two exceptions, as they use a bivariate approach.

spread variations that could be related to variations of short- and/or long-term yields.[3] Furthermore, the author argues that term premiums should not be neglected, as they could have an impact on the shape of the yield curve; however, Rosenberg and Maurer (2008) provide evidence to the contrary. Specifically, the latter authors empirically show that taking term premiums into account does not lead to better recession forecasts. According to them, the expectation component of the yield spread indeed signals recession, but the term premium is uninformative. Other macroeconomic variables have predictive power and can improve recession forecasting accuracy. First, Nyberg (2010) provides some empirical evidence of the usefulness of stock market returns and the foreign term spread. These additional variables can capture the monetary policy stance but cannot be indicators of other potential risk factors for recessions. A subsequent study by Ng (2012) extends previous analyses by incorporating a more complete set of recession risk factors (financial market expectations of a gloomy economic outlook, credit or liquidity risks in the general economy, the risks of negative wealth effects resulting from the bursting of asset price bubbles, and signs of deteriorating macroeconomic fundamentals).[4] In a more recent study, Park, Simar and Zelenyuk (2020) replicate the results of Kauppi and Saikkonen (2008) obtained with a parametric linear dynamic probit model. Extending the data up to 2017, the researchers validate the results using both parametric and nonparametric validation approaches. To improve probit models' fitting in the US, other variables have also been suggested, such as sentiment (Christiansen, Eriksen and Moller, 2014), credit (Ponka, 2017), liquidity (Ng, 2012), money supply variables (Hwang, 2019), volatility index (VIX) (Adrian, Estrella and Shin, 2010) and economic policy uncertainty (Karnizova and Li, 2014).

[3]Specifically, the author argues that considering the yield spread only leads to the conclusion that an increasing (respectively, decreasing) short-term yield has an effect on recession probability similar to that of a decreasing (respectively, increasing) long-term yield.

[4]The authors use macroeconomic and financial indicators to proxy the four risk factors: the yield spread, the TED spread (the interest rate differential between 3-month LIBOR and 3-month T-bills), the equity price index, the housing price index, and a macro-leading index.

3.2.2 Predictive power of the yield curve across countries and over time

The economists' consensus based on a large body of empirical evidence is that binary response models perform well in forecasting future probabilities of recessions. However, the future predictive power of the yield curve remains fragile, as it is not structural but related to monetary policy (Estrella, 2005). Specifically, the predictive power of the yield spread is a function of parameters of the monetary policy rule. Such parameters are based on policy objectives (money supply targeting, inflation targeting or price level targeting) that could change over time and across monetary areas. Thus, the relevance of generalizing the use of yield spreads in forecasting recessions relies on both stability and homogeneity of their relationship.

On the one hand, the literature has been enriched with several empirical papers investigating the effects of structural breaks in monetary policy. Including a Markov-switching coefficient variation in the probit model, Dueker (1997) and Ahrens (2002) both reject the linearity hypothesis. Both authors observe significant regimes, and, in particular, the results of Ahrens (2002) indicate that the two estimated regimes are associated with expansions and recessionary periods, respectively. However, the results of Estrella, Rodrigues and Schich (2003) and Wright (2006) indicate that binary models of expansion-recession provide more stable estimates than does a continuous model of GDP growth. In a dynamic probit model framework, Chauvet and Potter (2002; 2005) and Bellego and Ferrara (2009) show that time-varying probit models improve in-sample fitting. However, following the modeling framework of Kauppi and Saikkonen (2008), Ng (2012) and more recently Hwang (2019) provide the opposite evidence (see Rudebusch and Williams, 2009 for a discussion of the puzzle of the enduring predictive power of the yield spread).
On the other hand, while the early literature focused on the US, more recent empirical studies have investigated the predictive power of the yield curve in an international framework. The objective is to assess the homogeneity of the relationship between the yield spread and probability of recessions across countries. For instance, Bernard and Gerlach (1998) and Ahrens (2002) perform the first international analyses. Their respective data samples cover eight countries among Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, the UK and the US. The authors' cross-country analyses highlight

that the yield curve predicts future recessions in all countries. However, the authors point out that its information content differs from one country to another. For instance, the predictive power of the yield curve is greater in Canada, Germany and the US than in Japan, Italy or the Netherlands. To a lesser extent, Estralla, Rodrigues and Schich (2003) and Sensier et al. (2004) focus on the US and several European countries and observe that the overall patterns in countries in the chosen set are similar. Duarte, Venetis and Paya (2005) and Moneta (2005) focus on the Euro area and obtain results that confirm the usefulness of the yield spread in predicting the likelihood of a future recession in the Euro area (see also Estrella and Mishkin (1997)). Moneta's results also emphasize that the German and French yield spreads are the most significant signals of recession in the Euro area and that country-level data provides better forecasting performance than does aggregated data. Chinn and Kucko (2015) have recently built the largest database covering nine countries (Canada, France, Germany, Italy, Japan, the Netherlands, Sweden, the UK and the US) over the period of 1970-2013. The results of the researchers' review of the predictive power of the yield curve indicate that probit models are a relatively good fit for the United States, Germany and Canada over the entire dataset, while the remaining models largely failed to anticipate the recessions of the 2000s. The models for Japan and Italy did not predict recessions well. While the literature is dominated by univariate time series analyses, a few papers deviate from this approach. For instance, Wright (2006) and King, Levin, and Perli (2007) use a bivariate time series approach to show that adding federal funds rates and credit spreads, respectively, as macroeconomic variables improves probit models' estimations in the US. Interestingly, the attempt of Ozturk and Pereira (2013) to reexamine the predictive power of the yield curve using a panel approach is, as far as we know, the first empirical study trying to compare the empirical results of probit models in an international framework. However, the researchers' results are weakened by several drawbacks pertaining to econometric and data issues.[5]

[5]Unfortunately, the empirical study of Ozturk and Pereira (2013) is based on unbalanced panel data; the researchers use a static binary model only, and their results are subject to statistical biases. Their modeling approach could have been dramatically improved using econometric specifications introduced in Kauppi and Saikkonen (2008) (i.e., using a dynamic dichotomous model with a lagged binary variable, a lagged index variable or both). Additionally, their estimations could have been adjusted for cross-sectional dependence using the correction of Driscoll and Kraay (1998).

3.3 Using a yield spread to predict recessions

3.3.1 Model

A dichotomous model enables us to regress a binary dependent variable on continuous independent variables. In this paper, we aim to estimate the probability at time $t-1$ that the economy will fall into recession at time t. The state of the economy is represented by a discrete variable y_t taking the value of 1 if the economy is in recession at time t, and equal to 0 otherwise. This dependent variable is estimated from various explanatory variables. The latter represent the state in which the economy will be at time t and make it possible to calculate the value π_t of an index. A binary financial analysis can be written as

$$y_t = \begin{cases} 1 & \text{if } \pi_t > 0 \\ 0 & \text{if } \pi_t \leq 0 \end{cases} \tag{3.1}$$

Estrella and Hardouvelis (1991), Estrella and Miskin (1996, 1997, 1998) and Bernard and Gerlach (1998) were the first to estimate the probability that a recession occurs at time t with a simple probit model, using the spread between the three-month and ten-year yields and macroeconomic variables to improve the quality of the regression. The model takes the following form:

$$Pr_{t-1}(y_t = 1) = F(\pi_t) = F(x_{t-1}\beta) \text{ for } t = 1, 2, ..., T, \tag{3.2}$$

where T represents the number of time series observations. Dependent variable y_t is a $[t-1]$ vector, x_{t-1} is a $[(t-1) \times k]$ matrix that represents explanatory variables, k is the number of explanatory variables, β is a $[k]$ vector that contains the set of estimated coefficients, and F(.) is a transformation function.[6]

Chauvet and Potter (2005) proposed an improvement in estimating this relationship, adding a latent continuous stochastic process and a coefficient associated with the error term. This explanatory variable has two important purposes. First, as a new source of information, it can improve the quality of the estimation. However, it also allows a time-varying parameter by taking into account the dependence on the latent variable. Kauppi and Saikkonen (2008) added a lagged dependent variable. Indeed, if a country is

[6]F(.) is a Gaussian c.d.f for the probit model and a logistic c.d.f. for the logit model.

in recession at time t, because of persistence of the crisis, the probability of the country staying in recession during the following period must be impacted accordingly. Then, a dynamic dichotomous model takes the following form:

$$Pr_{t-1}(y_t = 1) = F(\pi_t) = F(x_{t-1}\beta + y_{t-1}\alpha + \pi_{t-1}\delta), \text{ for } t = 1, 2, ..., T, \tag{3.3}$$

The parameters are the same as those defined in equation (3.2). The innovation of Chauvet and Potter (2005) and Kauppi and Saikkonen (2008) consists of the addition of two variables – the lagged index variable π_{t-1} and the lagged dependent variable y_{t-1} – with their associated coefficients δ and α.

Bernard and Gerlach (1998) studied the possibility of using the yield curve to predict future economic activity in 8 countries. More recently, Chinn and Kucko (2015) reexamined this evidence using a dataset covering 9 countries. The above papers estimated this relationship for each country separately from the others. Instead, a panel approach is preferred that would estimate the impact of an explanatory variable on a set of countries. Moreover, a panel regression allows increasing the number of observations available to improve the estimation capacity. Candelon, Dumitrescu and Hurlin (2014) proposed extending the model of Kauppi and Saikkonen (2008) to panel data. This model is written as follows:

$$Pr_{t-1}(y_{i,t} = 1) = F(\pi_{i,t}) = F(\beta' x_{i,t-1} + \alpha y_{i,t-1} + \delta \pi_{i,t-1} + \eta_i),$$
$$\text{for } t = 1, 2, ..., T, \text{ and } i = 1, 2, ..., N, \tag{3.4}$$

where N is the number of countries in the panel, and η_i is a country fixed effect for the control of unobserved heterogeneity and potential bias.

The first innovation of this paper consists of an extension of approaches of Bernard and Gerlach (1998) and Chinn and Kucko (2015) with a dynamic logit panel model.[7] As far as we know, a balanced panel framework has never been used to study the relation-

[7] The logistic c.d.f. is preferred to a Gaussian c.d.f, as it is more appropriate for the study of extreme events such as crises.

ship between yield spreads and recessions.[8] Here, we propose extending such studies by evaluating a set of 13 countries over the period of 1975-2019, following the methodology proposed recently by Candelon, Dumitrescu and Hurlin (2014). Furthermore, we follow Kauppi and Saikkonen (2008), and we estimate four dichotomous models. The first model is a static logit model with two restrictions: $\alpha = \delta = 0$ (Model 1). In this case, only the exogenous macroeconomic variables affect the future occurrence of a crisis. The second and third models are dynamic and include either a lagged value of the binary variable y_{t-1} with a restriction on $\delta = 0$ (Model 2) or a lagged index π_{t-1} with a restriction on $\alpha = 0$ (Model 3). Finally, the last dynamic model combines the two preceding cases and includes both a lagged binary variable y_{t-1} and a lagged index π_{t-1} (Model 4). The model that minimizes the Bayesian information criterion (BIC) is chosen as the best model.

Candelon, Dumitrescu and Hurlin (2014) show that the four different alternatives of the model presented in equation (3.4) can be estimated under the same exact maximum likelihood (EML) framework. The log-likelihood function has the following general form:

$$\begin{aligned}
\mathrm{LogL}(\theta, \eta_i) &= \sum_{i=1}^{N} \mathrm{LogL}_i(\theta, \eta_i) && (3.5)\\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}[y_{i,t}\mathrm{log}(F(\pi_{i,t}(\theta, \eta_i))) + (1 - y_{it})\mathrm{log}(1 - F(\pi_{i,t}(\theta, \eta_i)))], && (3.6)
\end{aligned}$$

However, the panel approach may have some constraints. The assumption of homogeneity of all parameters can be too restrictive despite the presence of fixed effects that can capture heterogeneity. Berg, Candelon and Urbain (2008) explained that pooling all available countries into one panel model might not be the best approach and should be supplemented by studying the existence of a clusters. In the last few years, heterogeneous panels have been the main focus of attention in the literature. Many studies test slope homogeneity and poolability in the panel data. Within this framework, Blomquist and Westerlund (2013) have extended the test of Pesaran and Yamagata (2008). However, their test does not deal with the practically relevant case of cross-sectional dependence and does not allow a dependence between the set of predictors and unobservable errors.

[8] Ozturk and Pereira (2013) studied the power of the yield curve to predict recessions for 32 countries with a static and unbalanced panel dichotomous approach.

Ando and Bai (2015)proposed an alternative solution using the results of Bai (2009), Su and Chen (2013) and Ando and Bai (2014) by incorporating interactive fixed effects.

Thus, to improve upon the existing methods, the second innovation of this paper consists of the identification of subgroups with homogeneous slopes. Our purpose is to deal with the potential problem of heterogeneity in order to complete and validate the panel approach. Inspired by the approach proposed by Zhang, Wang and Zhu (2019), we have developed a two-stage approach involving a dichotomous dynamic model. In their paper, Zhang, Wang and Zhu (2019) proposed a method for panel data where fixed effects were estimated upstream. Such fixed effect, once estimated, is subtracted from the explanatory variable (Lin and Ng, 2012). Afterwards, the parameters of the regression are estimated by minimizing over all possible partitions of N units into G groups. In our case, since the dependent variable is a binary variable that only takes values of 1 or 0, we cannot simply subtract the fixed effect. Hence, to estimate our clustering logit model, we first estimate the logit model for each country in order to obtain the intercept for each of them. Subsequently, we estimate the group-specific parameters θ_g for $g = 1, ..., G$ while constraining the previously estimated fixed effects, where θ_g =[β_g α_g δ_g] and G is not fixed beforehand. Thus, without subtracting the fixed effects from the explanatory variable (Lin and Ng, 2012), this method allows us to offer an alternative for dichotomous models.

Let $\Theta = \{\theta_g : 1, ..., G\}$ be the set comprising all group-specific slopes and $\gamma = \{g_i, i = 1, ..., N\}$ be the set of group memberships for N units. Thus, $\gamma \in F_G$ denotes a particular partition of N units, where F_G is the set of all partitions of $\{1, ..., N\}$ into G groups. Let Ψ be a compact subset of $\mathbb{R}$ and Θ be a compact subset of $\mathbb{R}^p$. In the first stage, we fit the logit regression for each unit and estimate the fixed effect $\widetilde{\eta}_i$ by $\widetilde{\eta}_i$, where

$$(\widetilde{\theta}_i, \widetilde{\eta}_i) =_{\eta \in \Psi, \theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} (y_{i,t} - F(\theta' X_{i,t} + \eta)). \tag{3.7}$$

Next, following Zhang, Wang and Zhu (2019), we estimate the group memberships and the group-specific parameters by

$$(\widehat{\Theta}, \widehat{\gamma}) =_{\theta \in \Theta, \gamma \in F_G} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{i,t} - F(\theta_g' X_{i,t} + \widetilde{\eta}_i)) \tag{3.8}$$

where $\beta \subseteq \Theta$ and $\gamma \in F_G$.

In the second step (equation (3.8)), we select θ that minimizes Pearson residuals [9] over all possible partitions of N units into G groups and the group-specific parameters from a compact subset of $\mathbb{R}^p$. The estimation procedure is summarized in Appendix 1. This allows us to cluster our sample of countries into homogeneous country groups. Moreover, the existence of groups of several countries allows us to test the partial homogeneity of our sample. As G is not initially fixed, the number of groups obtained from our method will be the one that allows the Pearson residuals to be minimized.

3.3.2 Data

Our objective is to reexamine the empirical relationship between the slope of the yield curve and future recessions. Our contribution is twofold: we aim to investigate (i) this relationship across countries and over time, and (ii) the cross-country homogeneity of the results. To do so, a dynamic panel data is clearly the most appropriate framework, as we aim to combine both cross-sectional and time series dimensions. Specifically, this approach is relevant because we aim to control the impact of monetary policy on the predictive power of the yield curve as well as other recession risk factors. Furthermore, we choose a balanced panel framework. The motivation of this choice is that our specific contribution is to investigate the homogeneity of this relationship across countries. Indeed, the identification of subgroups of countries with heterogeneous predictive power of the yield spread is based on a regression-based clustering method that requires balanced panel data.

To match these data requirements, we introduce an extended and updated database that is, as far as we know, the largest balanced panel dataset for the yield spread - recession relationship. First, starting from the dataset of Chinn and Kucko (2015), we add four countries and update the entirety of data until March 2019. Specifically, our dataset includes 13 OECD countries: Australia, Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, New Zealand, Sweden, Switzerland, the UK and the US. The resulting dataset covers almost 45 years of historical data for each country at the monthly frequency. As to the yield curve, we choose to focus on the "10-year minus 3-month"

[9] As we are in a clustering logistic panel analysis, using Pearson residuals is preferable. See Agresti (2018), section 5.2.4, page 147-148.

interest rate spread because in the literature it is considered to be the most statistically significant predictor of recessions (Nyberg, 2010). However, our dataset differs from those discussed in the literature, as we do not focus on NBER and ECRI recession dummies for the US and other countries, respectively, but instead use OECD recession dummies.[10] This choice is motivated by the objective to cover as many countries as possible and the need to have similar measures for every country (see the methodology of Bry and Boschan (1971)). Following Wright (2006), we also include central bank policy rates as a control variable. Indeed, this variable enables us not only to control for monetary policy but also to disentangle the origins of yield spread variations that could be related to variations of short- and/or long-term yields. As to macroeconomic control ("macro-control") variables, we first include a set of recession risk factors as in Ng (2012), who introduces stock and housing markets' prices, credit and liquidity spreads and a macro-leading index. Instead of using a macro-leading indicator, we select a set of distinct macroeconomic variables. Specifically, we include proxies of (i) stock market returns as in Nyberg (2010), (ii) oil price returns as in Kilian and Vigfusson (2017), (iii) housing market returns as in Ng (2012), (iv) market sentiment as in Christiansen, Eriksen and Molleret (2014), (v) credit spread as in Ponka (2017), (vi) liquidity spread as in Ng (2012), (vii) economic policy uncertainty as in Karnizova and Li (2014) and (viii) volatility index (VIX) as in Adrian, Estrella and Shin (2010).

In summary, our final database covers 13 OECD countries from 1975 to 2019 at the monthly frequency (i.e., having a total of 6,890 observations). It includes country-level recession dummies, 10-y - 3-m yield spreads, central banks' rates and a set of recession risk factors selected from the recent literature. Tables 3.1 and 3.2 summarize the data and report the variables' respective names, descriptions, codes, frequency, sources and references.

[10]As a robustness check, we have performed the same estimations using the ECRI database. The results obtained using OECD and ECRI recession dummies are similar. We note that the ECRI database has a lower coverage than does the OECD database (in our case, the respective numbers of countries are 8 and 13). The results based on ECRI recession dummies are reported in Appendix 2.

Table 3.1: Description of Panel Data: 1975-2019

Variable	Description	Code	Freq.
Recession	Recession dummy indicator	REC	M
Yield spread	Difference between the long- and short-term government debt yields (10 Y minus 3 M)	$YSPR$	M
Central bank's rate	Refinancing interest rate	$CBAN$	M
Stock market	Stock market index	$STOM$	M
Crude oil market	Brent spot price (USD/bbl)	$OILM$	M
Housing market	Housing prices' indicator	$HOUM$	Q
Sentiment	Consumer sentiment and business trends' surveys: Comp. Indicators	$SENT$	M
Credit spread	Difference between the interest rates of inter-bank loans and short-term government debt	$CSPR$	M
Liquidity spread	Difference between short-term government debt and central bank's refinancing rate	$LSPR$	M
Uncertainty	EPU index of Baker, Bloom and Davis (2016)	EPU	Q
Volatility	CBOE Volatility Index	VIX	M

Notes: This table provides each country-level variable's name, description, code and frequency ("M" = monthly and "Q" = quarterly). The panel dataset covers 13 OECD countries over the period from January 1975 to March 2019.

Table 3.2: Sources of Panel Data

Variable	Sources	References
Recession	OECD and ECRI	Chinn and Kucko (2015)
Yield spread	OECD	Nyberg (2010)
Central banks' rates	BIS and OECD	Wright (2006)
Stock market	Bloomberg	Estrella and Mishkin (1998), Nyberg (2010)
Crude oil market	World Bank	Engemann, et al. (2011) and Kilian and Vigfusson (2017)
Housing market	OECD	Ng (2012)
Sentiment	OECD	Christiansen, Eriksen and Molleret (2014)
Credit spread	OECD and FRED	Ponka (2017)
Liquidity spread	OECD and FRED	Ng (2012), Erdogan, Bennett and Ozyildirim (2015)
Uncertainty	FRED	Karnizova and Li (2014)
Volatility	FRED	Adrian, Estrella and Shin (2010)

Notes: This table shows each country-level variable's name, source(s) and reference(s).

3.3.3 Empirical analysis

In this section, we aim to reexamine the predictive power of the term spread in two steps. First, we investigate the relationship between the yield spread and future recessions in an international balanced panel dataset. Afterwards, using the previously obtained results, we test the homogeneity of this relationship across countries via a cluster analysis. Consistently with previous studies, we use a binary model framework, preferring a logit model to a probit model as in Sensier et al. (2004), Moneta (2005) and more recently in Hwang (2019).[11] On this basis, we introduce some changes of the econometric specification. First, we have to adapt the binary model mostly used in a univariate fashion in the literature to a balanced panel framework using country fixed effects. Next, in regression estimation, we use the correction of Driscoll and Kraay (1998) for cross-sectional dependence.[12] Using this augmented logit model, we run several regressions to estimate various binary models' specifications: a static logit model (Model 1), a dynamic logit model including the lagged recession dummy y_{t-1} (Model 2), a dynamic logit model including the lagged index π_{t-1} (Model 3), and a dynamic logit model including both the lagged recession dummy and the lagged index (Model 4). The results are reported in Table 3.3.

Table 3.3: Estimation results of panel logit models – Monthly frequency – 1975-2019

Model	(1)	(2)	(3)	(4)
$YSPR_{-1}$	-0.0821**	-0.1648***	-0.0719**	-0.1833***
	(0.0378)	(0.0492)	(0.0317)	(0.0653)
REC_{-1}		6.4657***		7.4607***
		(0.1111)		(0.1911)
$Index_{-1}$			-0.2416	-0.2808
			(0.4383)	(0.4773)
	Relevant Statistics			
BIC	9,532.2	2,351.7	9,538.0	2,353.1
	Fixed Effects			
Country	Yes	Yes	Yes	Yes
No. Observations	6,890	6,890	6,890	6,890

Notes: This table reports the estimates obtained from static and dynamic logit models (1)-(4) for a panel of 13 countries covering the period from February 1975 to March 2019 at the monthly frequency with one lag. The dependent variable is the recession dummy. Results are computed using R 3.6.0 (R Core Team, 2020) and the *ews* (*v0.1.0*; Hasse and Lajaunie, 2020) package. The full reproducible code is available on CRAN. We report Bayesian (BIC) information criteria for each specification. Standard errors are reported in parentheses below the estimates. Labels $***$, $**$ and $*$ indicate significance at 99%, 95% and 90% levels, respectively.

The results in Table 3.3 indicate that the lagged yield spread and the lagged binary coefficients are both highly significant. Their coefficients are negative and positive, respectively, i.e., the yield spread is inversely related to the probability of future recessions, and the probability of being in recession at time $t-1$ is strongly related to the probability

[11]Probit and logit are both dichotomous models that exhibit very similar features. Considering the low ratio of ones to zeros that the recession dummy exhibits, logit models are preferable to probit models. See Ben Naceur, Candelon and Lajaunie (2019).

[12]Based on a univariate analysis, previous empirical studies use the correction of Newey and West (1987, 1994) instead.

of being in recession at time t. In a period covering 45 years of monthly observations across 13 OECD countries, the predictive power of yield spread appears to be strong and significant. This predictive power is unaltered by the presence of a lagged binary variable or/and a lagged index. According to the BIC criteria,[13] we select model (2) that includes an intercept, the lagged yield spread and the lagged binary variables. The next steps of this empirical study focus on this econometric specification.

The recent literature on the predictive power of the yield spread extensively explores the role of the macroeconomic environment. Following Nyberg (2010), Ng (2012), Christiansen, Eriksenand and Moller (2014), Karnizova and Li (2014), Engemann, Kliesen and Owyang (2011), Adrian, Estrella and Shin (2010), Kilian and Vigfusson (2017) and Ponka (2017), we extend the previously selected model (Model 2), adding several macro-control variables. These recession risk factors are added to Model 2 in different ways, depending on whether they are economic or financial variables. The results are reported in Table 3.4, where the two first columns indicate the regression results obtained with economic and financial control variables, respectively. Indeed, Ng and Wright (2013) document that recessions originate from monetary policy shocks or potentially in the financial markets. The last column reports the results of a regression of the augmented binary model with all macro-control variables.

The results in Table 3.4 indicate that the coefficient of the lagged yield spread remains negative and significant. Consistently with the recent literature, the predictive power of the yield spread is robust to the introduction of macro-control variables. Specifically, coefficients of crude oil and stock market returns are negative and significant as in Engemann, Kliesen and Owyang (2011) and Ng (2012), respectively. The results also confirm the empirical results of Christiansen, Eriksen and Moller (2014) and Ponka (2017), as both coefficients of sentiment and credit spread are positive and significant.

[13] In the literature, the pseudo-R^2 measure has been the criterion most frequently used as a goodness-of-fit measure guiding the choice of the model and the optimal lag orders for explanatory variables (Estrella and Mishkin, 1998; Kauppi and Saikkonen, 2008). Following Candelon, Dumitrescu and Hurlin (2012, 2014), we select the best econometric specification from BIC criteria.

Table 3.4: Estimation of a panel with macro-control variables – Monthly frequency – 1975-2019

Model	Eco	Fin	Global
$YSPR_{-1}$	-0.1691***	-0.1799***	-0.1886***
	(0.0191)	(0.0202)	(0.0201)
$SENT_{-1}$	0.0808		0.1203***
	(0.0497)		(0.0427)
$OILM_{-1}$	-1.1146***		-0.8227*
	(0.398)		(0.4782)
$STOM_{-1}$		-4.5252***	-4.4755***
		(0.4417)	(0.4321)
$CSPR_{-1}$		0.3732***	0.3907***
		(0.0674)	(0.0648)
REC_{-1}	6.5082***	6.4643***	6.5388***
	(0.0606)	(0.0429)	(0.0458)
Relevant Statistics			
BIC	2,365.4	2,335.8	2,348.7
Fixed Effects			
Country	Yes	Yes	Yes
No. Observations	6,890	6,890	6,890

Notes: This table reports the estimates obtained from a dynamic logit model with a lagged binary variable for a panel of 13 countries over the period of 1975 – 2019 with one lag, and monthly frequency. The dependent variable is the recession dummy. We report Bayesian (BIC) information criteria for each specification. Using generalized linear model (GLM), the correction of Driscoll and Kraay (1998) is applied so that standard errors are robust to heteroscedasticity and autocorrelation. Results are computed using R 3.6.0 (R Core Team, 2020). Standard errors are reported in parentheses below the estimates. Labels $***$, $**$ and $*$ indicate significance at 99%, 95% and 90% levels, respectively.

As a robustness check, we replicate this first empirical study in two steps. First, we use data at the quarterly frequency to test the predictive power of the yield spread in a longer run. Incidentally, using lower-frequency data enables us to enrich the model with two macro-control variables that are unavailable at the monthly frequency: housing market prices and economic policy uncertainty (see Table 3.5). Afterwards, we restrict the initial data sample to a temporal subsample from 1999, using lags of 1 and 3 (see Tables 3.6 and 3.6, respectively). Indeed, the recent literature reports structural breaks during the 1990s; additionally, the European Monetary Union adopted Euro in 1999. The results are reported in Tables 3.5, 3.6 and 3.7, respectively.

Table 3.5: Estimation of a panel with macro-control variables – Quarterly frequency – 1975-2019

Model	Eco	Fin	Global
$YSPR_{-1}$	-0.2138***	-0.2228***	-0.2434***
	(0.0239)	(0.0272)	(0.0336)
$SENT_{-1}$	0.1550**		0.2768***
	(0.0636)		(0.0755)
$OILM_{-1}$	-0.6416		-0.4927*
	(0.3931)		(0.2705)
EPU_{-1}	-0.5122		-0.3384
	(0.6729)		(0.6800)
$STOM_{-1}$		-3.7137***	-3.7778***
		(1.4334)	(1.4016)
$CSPR_{-1}$		0.2327***	0.2685***
		(0.0678)	(0.0716)
$HOUM_{-1}$		-7.9327***	-12.8603***
		(1.6013)	(3.2419)
REC_{-1}	4.2201***	4.063***	4.2297***
	(0.0611)	(0.0460)	(0.0841)
	Relevant Statistics		
BIC	1,716.5	1,686.1	1,692.2
	Fixed Effects		
Country	Yes	Yes	Yes
No. Observations	2,275	2,275	2,275

Notes: This table reports the estimates obtained from a dynamic logit model with a lagged binary variable for a panel of 13 countries at the quarterly frequency over the period of $1975-2019$ with one lag. The dependent variable is the recession dummy. We report Bayesian (BIC) information criteria for each specification. Using generalized linear model (GLM), the correction of Driscoll and Kraay (1998) is applied so that standard errors are robust to heteroscedasticity and autocorrelation. Results are computed using R 3.6.0 (R Core Team, 2020). Standard errors are reported in parentheses below the estimates. Labels $***$, $**$ and $*$ indicate significance at 99%, 95% and 90% levels, respectively.

Table 3.6: Estimation of a panel with macro-control variables – Monthly frequency – 1999-2019

Model	Eco	Eco Global	Fin	Fin Global	Global
$YSPR_{-1}$	-0.3493***	-0.2804***	-0.3994***	-0.4330***	-0.4163***
	(0.0857)	(0.0752)	(0.0857)	(0.0693)	(0.0710)
$CBAN_{-1}$		0.0692		-0.0244	-0.0371
		(0.0765)		(0.0767)	(0.0742)
$SENT_{-1}$	0.0531	0.0582			0.2302***
	(0.0565)	(0.06)			(0.0605)
$OILM_{-1}$	-1.4332	-1.5061			-1.0708
	(1.2637)	(1.3272)			(1.4529)
$STOM_{-1}$			-3.0940*	-3.0153*	-2.3735
			(1.7032)	(1.6762)	(1.7645)
$LSPR_{-1}$				-0.1422	-0.0712
				(0.4824)	(0.4395)
$CSPR_{-1}$			0.5261***	0.4145	0.4782
			(0.1279)	(0.4495)	(0.4034)
VIX_{-1}			0.0341**	0.0376***	0.0491***
			(0.0142)	(0.0131)	(0.0114)
REC_{-1}	6.5139***	6.531***	6.3971***	6.3859***	6.4821***
	(0.0769)	(0.0852)	(0.0562)	(0.0627)	(0.0621)
Relevant Statistics					
BIC	1,165.6	1,172.6	1,157.7	1,173.5	1,184.7
Fixed Effects					
Country	Yes	Yes	Yes	Yes	Yes
No. Observations	3,172	3,172	3,172	3,172	3,172

Notes: This table reports the estimates obtained from a dynamic logit model with a lagged binary variable for a panel of 13 countries over the period of 1999 – 2019 with one lag, and monthly frequency. The dependent variable is the recession dummy. We report Bayesian (BIC) information criteria for each specification. Using generalized linear model (GLM), the correction of Driscoll and Kraay (1998) is applied so that standard errors are robust to heteroscedasticity and autocorrelation. Results are computed using R 3.6.0 (R Core Team, 2020). Standard errors are reported in parentheses below the estimates. Labels $***$, $**$ and $*$ indicate significance at 99%, 95% and 90% levels, respectively.

Table 3.7: Estimation of a panel with macro-control variables – Quarterly frequency – 1999-2019

Model	Eco	Eco Global	Fin	Fin Global	Global
$YSPR_{-1}$	-0.2844***	-0.1876***	-0.2905***	-0.2249***	-0.1806***
	(0.0856)	(0.0683)	(0.0756)	(0.0546)	(0.0536)
$CBAN_{-1}$		0.1014		0.0159	-0.01206
		(0.0796)		(0.0747)	(0.0612)
$SENT_{-1}$	0.2012***	0.2121***			0.5172***
	(0.0604)	(0.0656)			(0.0556)
$OILM_{-1}$	-2.0676***	-2.1877***			-0.4707
	(0.4961)	(0.5484)			(1.3111)
EPU_{-1}	-2.2981**	-1.1563			-1.0933**
	(1.0894)	(1.3624)			(0.5359)
$STOM_{-1}$			-8.8147***	-9.0266***	-8.1991***
			(0.8977)	(0.9178)	(0.8160)
$LSPR_{-1}$				0.8268***	1.0324***
				(0.2668)	(0.2290)
$CSPR_{-1}$			0.4089**	1.0534***	1.2680***
			(0.1844)	(0.3204)	(0.2725)
$HOUM_{-1}$			-13.3674***	-13.6075***	-23.3889***
			(3.2317)	(3.5788)	(2.7381)
VIX_{-1}			-0.0054	-0.0157	0.0067
			(0.0128)	(0.0140)	(0.0126)
REC_{-1}	4.2541***	4.2642***	4.1242***	4.1681***	4.4197***
	(0.0752)	(0.0769)	(0.0704)	(0.0764)	(0.0843)
Relevant Statistics					
BIC	844.8	850.3	810.4	821.2	818.7
Fixed Effects					
Country	Yes	Yes	Yes	Yes	Yes
No. Observations	1,040	1,040	1,040	1,040	1,040

Notes: This table reports the estimates obtained from a dynamic logit model with a lagged binary variable for a panel of 13 countries at the quarterly frequency over a period of 1999 − 2019 with one lag.The dependent variable is the recession dummy. We report Bayesian (BIC) information criteria for each specification. Using generalized linear model (GLM), the correction of Driscoll and Kraay (1998) is applied so that standard errors are robust to heteroscedasticity and autocorrelation. Results are computed using R 3.6.0 (R Core Team, 2020). Standard errors are reported in parentheses below the estimates. Labels $***$, $**$ and $*$ indicate significance at 99%, 95% and 90% levels, respectively.

Compared to the previous results for the predictive power of the yield spread and the role of macro-control variables, the results in Tables 3.5, 3.6 and 3.7 are qualitatively similar. As in Table 3.4, we observe that the lagged yield spread's coefficient remains negative and significant in each case. The predictive power of the term spread is altered neither if the lag is changed from 1 to 3 months nor by the introduction of two more macro-control variables (i.e., economic policy uncertainty (EPU), housing market prices and stock market volatility (VIX)). Moreover, the results for these recession risk factors reported in Table 3.5 are consistent with the literature. On the one hand, the lagged housing market prices' coefficient is negative and highly significant as in Ng (2012). Indeed, in the past 45 years, housing prices often declined before the start of most recessions, as did equity prices. On the other hand, the lagged EPU coefficient is not significant. This result is similar to those of Karnizova and Li (2014), who empirically show that EPU is only significant beyond 5 quarters. Hence, economic policy uncertainty does not seem to have a significant role in recessions in the short run (from 1 to 3 months). By the way, the VIX is not significant as in Adrian, Estrella and Shin (2010). The results in Table 6 that reports regressions on a temporal subsample are similar to those described previously. Focusing on the period from 1999 to 2019, the results indicate that the predictive power of the yield spread is robust to time-sampling. The coefficient of the lagged yield spread is still highly significant and negative; however, we note that it is slightly smaller than it is for the entire sample at the quarterly frequency. This result supports the widespread idea that the predictive power of the yield curve has deteriorated during the 1990s. According to Chinn and Kucko (2015), this phenomenon stems from (i) changing links between interest rates and output, (ii) a failure of long-term interest rates to rise along with the short-term policy rate, and (iii) the zero lower bound (ZLB) implying that central banks try to lower long-term interest rates instead of lowering short-term rates. To deal with this issue, we add another macro-control variable: the lagged central banks' rates as in Wright (2006). The purpose is to regress the yield spread, controlling for (i) the monetary policy stance and (ii) the level of short- and long-term yields at the same time.[14] In addition to central banks' rates, we also add the lagged liquidity spread as in Ng (2012). The results

[14] Referring to Wright (2006), Chinn and Kucko (2015) choose to use the 3-month yield instead of central banks' rate. Using this short-term yield enables the authors to easily control the level of yields to distinguish a rise in short-term yields and a drop in long-term yields, and vice versa.

indicate that neither central bank rates nor liquidity risks are significant. The monetary policy stance, proxied by central bank rates, appears to have no short-term impact on the predictive power of the term spread.

predictive power of the yield curve. Indeed, Berg, Candelon and Urbain (2008) focused on the poolability issue for a panel, and recommend to construct country clusters. To do so, we adapt the methodology of Zhang, Wang and Zhu (2019) to compute optimal clusters of entities in a balanced panel dataset. The purpose of this cluster analysis is to indirectly assess homogeneity of the predictive power of the yield spread across countries. If the optimal number of

In summary, the predictive power of the term spread has been reexamined using a panel dataset of 13 OECD countries over a period of more than 45 years. The results indicate that the yield spread is a valuable predictor of future recession in the short run. Specifically, the predictive power of the term spread is robust to several econometric specifications, time sampling and, last but not the least, a set of eight macro-control variables chosen in accordance with recent studies. Our approach is different from the literature due to using balanced panel data. Using these empirical results that are consistent with those of previous studies, we now aim to investigate the homogeneity of the predictive power of the yield curve. Indeed, Berg, Candelon and Urbain (2008) focused on the poolability issue for a panel, and recommend to construct country clusters. To do so, we adapt the methodology of Zhang, Wang and Zhu (2019) to compute optimal clusters of entities in a balanced panel dataset. The purpose of this cluster analysis is to indirectly assess homogeneity of the predictive power of the yield spread across countries. If the optimal number of clusters is equal to the number of countries, then this cluster analysis would indicate that our results for the predictive power of the yield spread are highly heterogeneous across countries. If, on the contrary, the optimal number of clusters is equal to one, then we could presume that our results are highly homogeneous. Using the optimal number and composition of clusters, we further analyze the predictive power of the yield spread in each cluster. The results are reported in Table 3.8 and illustrated in Appendix 3. [15]

[15] Tables 3.11 and 3.12 in Appendix 5 also include all the estimations obtained from dynamic logit model (2) for each of the 13 countries. These results complement our panel approach to see the consistency of

Table 3.8: Estimation of the clustering panel – Monthly frequency – 1999-2019

Groups	Group 1	Group 2	Group 3	Group 4
	SWE	UK	ITA	CHE
	NLD	AUS	FRA	
	BEL	NZL		
	JPN	DEU		
	CAN			
	USA			
$YSPR_{-1}$	-0.4274***	-0.0662	-0.5641***	-0.2851
	(0.0800)	(0.1018)	(0.1168)	(0.5386)
$CBAN_{-1}$	0.1536	0.0449	-0.0364	-0.1120
	(0.0938)	(0.0313)	(0.1916)	(0.2841)
REC_{-1}	6.6000***	6.3589***	6.6098***	6.4305***
	(0.1366)	(0.1021)	(0.0.3330)	(0.6954)
		Fixed Effects		
Country	Yes	Yes	Yes	Yes
No. Observations	1,470	980	490	245

Notes: This table reports the estimates obtained from a dynamic logit model with a lagged binary variable for a panel of 13 countries at the monthly frequency over the period of 1999 − 2019 with one lag. The dependent variable is the recession dummy. Using generalized linear model (GLM), the correction of Driscoll and Kraay (1998) is applied so that standard errors are robust to heteroscedasticity and autocorrelation. Results are computed using R 3.6.0 (R Core Team, 2020). Standard errors are reported in parentheses below the estimates. Labels $***$, $**$ and $*$ indicate significance at 99%, 95% and 90% levels, respectively.

At the global level, the results in Table 3.8 indicate that the predictive power of the yield spread across our sample of 13 OECD countries is partially homogeneous. Indeed, the cluster analysis reveals that the optimal combination consists of four clusters only. Moreover, among clusters of this optimal combination, 2 clusters include 10 countries out of 13. Next, at the cluster level, some differences are observed about the relationship between the term spread and future recessions. The lagged yield spread coefficient is significant for 2 clusters only (Group 1: Belgium, Canada, Japan, the Netherlands, Sweden and USA; Group 3: France and Italy). Furthermore, the coefficient is smaller for Group 3 than for Group 1, while central banks' rates are not significant in any cluster. As other variables are significant in each cluster, the optimal set of clusters appears to be computed from yield spread coefficients only. In summary, these results highlight that more than the half of countries have experienced recessions related to a yield spread decrease, and that, at first sight, monetary policy seems to have no role in the short run. Incidentally, countries belonging to the European Monetary Union since 1999 are not grouped in a single cluster. However, we shall approach the conclusions with caution, as central banks' policy rates are only one proxy of monetary policy stance. Investigating further shows that

clustering results.

the latter results highlight a distinct feature of monetary policy across clusters. Indeed, the cluster distribution may be related to the monetary policy target and to the use of unconventional monetary policy tools. For instance, Groups 1 and 3 include proportionally few countries that have adopted inflation targeting over the period from 1999 to 2019. In contrast, Group 2 includes a majority of countries that have officially adopted inflation targeting before 1999. Last, we note that countries engaged in some form of quantitative easing (QE) and that have reached the Zero Lower Bound (ZLB) tend to be in the same clusters. These similarities are reported in Table 3.9.

Table 3.9: Clusters, monetary policy and the yield curve

Country	Cluster	Pred. power	Monetary policy target	Date	QE	Dates
Belgium	A	Yes	hybrid		Yes	2015
Canada	A	Yes	inflation targeting	1991	No	
Japan	A	Yes	inflation targeting	2013	Yes	2001
Netherlands	A	Yes	hybrid		Yes	2015
Sweden	A	Yes	inflation targeting	1993	Yes	2015
United States	A	Yes	inflation targeting	2012	Yes	2008
Australia	B	No	inflation targeting	1993	No	
Germany	B	No	hybrid		Yes	2015
New Zealand	B	No	inflation targeting	1990	No	
United Kingdom	B	No	inflation targeting	1992	Yes	2009
France	C	Yes	hybrid		Yes	2015
Italy	C	Yes	hybrid		Yes	2015
Switzerland	D	No	hybrid		Yes	2012

Notes: This table reports the clusters' countries and the associated results for the predictive power of the yield curve, the monetary policy target and date of the last change as well as the Quantitative Easing (QE) and the first launch date. Data on monetary policy targets are from the IMF Annual Report on Exchange Arrangements and Exchange Restrictions (2018).

Based on this empirical analysis, we confirm that monitoring the yield curve should be useful for forecasting recessions in most industrialized countries. At the global level, the predictive power of the yield spread is confirmed in a panel dataset covering a period of over 45 years of monthly observations and including 13 OECD countries. Controlling for a set of 8 recession risk factors selected from the empirical literature, the predictive power of the yield spread is also robust to several econometric specifications and time-sampling in the short run (from 1 month to 1 quarter ahead). Investigating the potential homogeneity of the predictive power of the yield spread across countries, we perform a cluster analysis on the results from panel logit regressions. The results indicate that the relationship between the yield spread and the probability of future recession is partially homogeneous. Specifically, we provide empirical evidence that, controlling for central banks' rates, the yield spread is a useful tool for more than a half of countries. Last, at the global level, central banks' policy rates have no impact on the predictive power of the yield curve, while the results at the cluster level seem to indicate that the predictive power could be related to monetary policy target. These mixed results about the impact of conventional monetary policy on the predictive power of the yield curve are not completely consistent with the analytical and empirical results of Estrella (2005) and Wright (2006), respectively. However, our results are in line with Chinn and Kucko (2015): we find that the short-term rate parameter is statistically insignificant and the predictive power of the yield spread seems to be impacted by the ZLB. Hence, these results call into question the structural interpretation of the relationship between the term spread and future recession.

3.4 Conclusions

In this paper, we reexamine the predictive power of the yield curve across countries and over time. Our purpose is to confirm the predictive power of the yield spread and to investigate its homogeneity across countries. To this end, we adapt the univariate modeling approach of Kauppi and Saikkonen (2008) to a balanced panel framework. We also adapt the clustering methodology of Zhang, Wang and Zhu (2019) for quantile regression to dichotomous models. Afterwards, we build a unique database to estimate the predictive power of the term spread, controlling for central banks' official rates (Wright, 2006), stock market returns (Nyberg, 2010), housing markets' returns and liquidity spread (Ng, 2012), stock market volatility (VIX) (Adrian, Estrella and Shin, 2010), sentiment (Christiansen, Eriksen and Molleret, 2014), economic policy uncertainty (Karnizova and Li, 2014), crude oil market returns (Engemann, Kliesen and Owyang, 2011; Kilian and Vigfusson, 2017) and credit spread (Ponka, 2017). Our results confirm the predictive power of the yield spread in most countries and indicate its partial homogeneity across countries. Our empirical findings, except those for the central bank policy rates, are consistent with the recent literature. Contrary to the analytical and empirical results of Estrella (2005) and Wright (2006), respectively, our findings indicate that central bank policy rates have no impact on the predictive power of the yield curve. However, the mixed results of the cluster analysis indicate that the predictive power of the term spread could be related to monetary policy frameworks (inflation targeting or alternative policy frameworks), as argued by Estrella (2005) or unconventional monetary policy tools, as argued by Chinn and Kucko (2015). Our results are robust to several econometric specifications, time-sampling and macro-control variables.

In summary, our contribution to the literature is threefold. First, we extend the database of Chinn and Kucko (2015), proposing a unique database including a set of eight country-level recession risk factors, and covering 13 industrialized countries over 45 years at the monthly frequency. Second, we confirm the predictive power of the yield spread in a new balanced panel framework and provide empirical evidence of its partial homogeneity via an innovative cluster analysis. Third, we show that monitoring of the yield curve evolution should be extended to countries other than the US. The impact of conventional monetary policy on the predictive power of the yield spread appears to be weaker than expected. These empirical findings support a wider use of the yield curve for monitoring business cycles.

References

Adrian, T., A. Estrella, and H. S. Shin (2010). “Monetary cycles, financial cycles and the business cycle”. In: *FRB of New York Staff Report* 421.

Agresti, A. (2018). *An introduction to categorical data analysis*. John Wiley & Sons.

Ahrens, R. (2002). “Predicting recessions with interest rate spreads: a multicountry regime-switching analysis”. In: *Journal of international Money and Finance* 21.4, pp. 519–537.

Ando, T. and J. Bai (2015a). “A simple new test for slope homogeneity in panel data models with interactive effects”. In: *Economics Letters* 136, pp. 112–117.

— (2015b). “Asset pricing with a general multifactor structure”. In: *Journal of Financial Econometrics* 13.3, pp. 556–604.

Bai, J. (2009). “Panel data models with interactive fixed effects”. In: *Econometrica* 77.4, pp. 1229–1279.

Bellégo, C. and L. Ferrara (2009). *Forecasting Euro-Area Recessions Using Time-Varying Binary Response Models for Financial Markets*. Banque de France Working Paper.

Bernard, H. and S. Gerlach (1998). “Does the term structure predict recessions? The international evidence”. In: *International Journal of Finance & Economics* 3.3, pp. 195–215.

Blomquist, J. and J. Westerlund (2013). “Testing slope homogeneity in large panels with serial correlation”. In: *Economics Letters* 121.3, pp. 374–378.

Bry, G. and C. Boschan (1971). “Programmed selection of cyclical turning points”. In: *Cyclical analysis of time series: Selected procedures and computer programs*. NBER, pp. 7–63.

Candelon, B., E.-I. Dumitrescu, and C. Hurlin (2012). “How to evaluate an early-warning system: Toward a unified statistical framework for assessing financial crises forecasting methods”. In: *IMF Economic Review* 60.1, pp. 75–113.

— (2014). “Currency crisis early warning systems: Why they should be dynamic”. In: *International Journal of Forecasting* 30.4, pp. 1016–1029.

Chauvet, M. and S. Potter (2002). “Predicting a recession: evidence from the yield curve in the presence of structural breaks”. In: *Economics Letters* 77.2, pp. 245–253.

— (2005). “Forecasting recessions using the yield curve”. In: *Journal of forecasting* 24.2, pp. 77–103.

Chinn, M. and K. Kucko (2015). “The predictive power of the yield curve across countries and time”. In: *International Finance* 18.2, pp. 129–156.

Christiansen, C., J. N. Eriksen, and S. V. Møller (2014). “Forecasting US recessions: The role of sentiment”. In: *Journal of Banking & Finance* 49, pp. 459–468.

Dimand, R. W. and R. Gomez Betancourt (2012). “Retrospectives: Irving Fisher’s Appreciation and Interest (1896) and the Fisher Relation”. In: *Journal of Economic Perspectives* 26.4, pp. 185–96.

Dotsey, M. (1998). “The predictive content of the interest rate term spread for future economic growth”. In: *FRB Richmond Economic Quarterly* 84.3, pp. 31–51.

Driscoll, J. C. and A. C. Kraay (1998). “Consistent covariance matrix estimation with spatially dependent panel data”. In: *Review of economics and statistics* 80.4, pp. 549–560.

Duarte, A., I. A. Venetis, and I. Paya (2005). “Predicting real growth and the probability of recession in the Euro area using the yield spread”. In: *International Journal of Forecasting* 21.2, pp. 261–277.

Dueker, M. J. (1997). “Markov switching in GARCH processes and mean-reverting stock-market volatility”. In: *Journal of Business & Economic Statistics* 15.1, pp. 26–34.

Engemann, K. M., K. L. Kliesen, and M. T. Owyang (2011). “Do oil shocks drive business cycles? Some US and international evidence”. In: *Macroeconomic Dynamics* 15.S3, pp. 498–517.

Estrella, A. (2005). “Why does the yield curve predict output and inflation?” In: *The Economic Journal* 115.505, pp. 722–744.

Estrella, A. and G. A. Hardouvelis (1991). “The term structure as a predictor of real economic activity”. In: *The journal of Finance* 46.2, pp. 555–576.

Estrella, A. and F. S. Mishkin (1996). “The yield curve as a predictor of US recessions”. In: *Current issues in economics and finance* 2.7.

— (1997). “The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European Central Bank”. In: *European economic review* 41.7, pp. 1375–1401.

— (1998). “Predicting US recessions: Financial variables as leading indicators”. In: *Review of Economics and Statistics* 80.1, pp. 45–61.

Estrella, A., A. P. Rodrigues, and S. Schich (2003). “How stable is the predictive power of the yield curve? Evidence from Germany and the United States”. In: *Review of Economics and Statistics* 85.3, pp. 629–644.

Fama, E. F. (1986). “Term premiums and default premiums in money markets”. In: *Journal of Financial Economics* 17.1, pp. 175–196.

Harvey, C. R. (1988). “The real term structure and consumption growth”. In: *Journal of Financial Economics* 22.2, pp. 305–333.

Hasse, J.-B. and Q. Lajaunie (2020). “EWS: Early Warning System”. In: *R package version 0.1. 0*.

Hwang, Y. (2019). “Forecasting recessions with time-varying models”. In: *Journal of Macroeconomics* 62, p. 103153.

Karnizova, L. and J. C. Li (2014). “Economic policy uncertainty, financial markets and probability of US recessions”. In: *Economics Letters* 125.2, pp. 261–265.

Kauppi, H. and P. Saikkonen (2008). “Predicting US recessions with dynamic binary response models”. In: *The Review of Economics and Statistics* 90.4, pp. 777–791.

Kessel, R. A. (1971). “The cyclical behavior of the term structure of interest rates”. In: *Essays on Interest Rates, Volume 2*. NBER, pp. 337–390.

Kilian, L. and R. J. Vigfusson (2017). “The role of oil price shocks in causing US recessions”. In: *Journal of Money, Credit and Banking* 49.8, pp. 1747–1776.

King, T. B., A. T. Levin, and R. Perli (2007). *Financial market perceptions of recession risk*. FEDS Working Paper.

Lin, C.-C. and S. Ng (2012). “Estimation of panel data models with parameter heterogeneity when group membership is unknown”. In: *Journal of Econometric Methods* 1.1, pp. 42–55.

Moneta, F. (2005). “Does the yield spread predict recessions in the Euro area?” In: *International Finance* 8.2, pp. 263–301.

Naceur, S. B., B. Candelon, and Q. Lajaunie (2019). “Taming financial development to reduce crises”. In: *Emerging Markets Review* 40, p. 100618.

Newey, W. K. and K. D. West (1987). “Hypothesis testing with efficient method of moments estimation”. In: *International Economic Review*, pp. 777–787.

— (1994). “Automatic lag selection in covariance matrix estimation”. In: *The Review of Economic Studies* 61.4, pp. 631–653.

Ng, E. C. (2012). “Forecasting US recessions with various risk factors and dynamic probit models”. In: *Journal of Macroeconomics* 34.1, pp. 112–125.

Ng, S. and J. H. Wright (2013). “Facts and challenges from the great recession for forecasting and macroeconomic modeling”. In: *Journal of Economic Literature* 51.4, pp. 1120–54.

Nyberg, H. (2010). “Dynamic probit models and financial variables in recession forecasting”. In: *Journal of Forecasting* 29.1-2, pp. 215–230.

Ozturk, H. and L. F. V. Pereira (2013). “Yield curve as a predictor of recessions: Evidence from panel data”. In: *Emerging Markets Finance and Trade* 49.sup5, pp. 194–212.

Park, B. U., L. Simar, and V. Zelenyuk (2020). “Forecasting of recessions via dynamic probit for time series: replication and extension of Kauppi and Saikkonen (2008)”. In: *Empirical Economics* 58.1, pp. 379–392.

Pesaran, M. H. and T. Yamagata (2008). “Testing slope homogeneity in large panels”. In: *Journal of econometrics* 142.1, pp. 50–93.

Ponka, H. (2017). “The role of credit in predicting US recessions”. In: *Journal of Forecasting* 36.5, pp. 469–482.

Rosenberg, J. V. and S. Maurer (2008). “Signal or noise? Implications of the term premium for recession forecasting”. In: *Federal Reserve Bank of New York-Economic Policy Review* 14.1.

Rudebusch, G. D. and J. C. Williams (2009). “Forecasting recessions: the puzzle of the enduring power of the yield curve”. In: *Journal of Business & Economic Statistics* 27.4, pp. 492–503.

Sensier, M. et al. (2004). “Domestic and international influences on business cycle regimes in Europe”. In: *International Journal of Forecasting* 20.2, pp. 343–357.

Su, L. and Q. Chen (2013). “Testing homogeneity in panel data models with interactive fixed effects”. In: *Econometric Theory* 29.6, pp. 1079–1135.

Van den Berg, J., B. Candelon, and J.-P. Urbain (2008). “A cautious note on the use of panel models to predict financial crises”. In: *Economics Letters* 101.1, pp. 80–83.

Wheelock, D. C., M. E. Wohar, et al. (2009). “Can the term spread predict output growth and recessions? A survey of the literature”. In: *Federal Reserve Bank of St. Louis Review* 91.5 Part 1, pp. 419–440.

Wright, J. H. (2006). *The yield curve and predicting recessions*. FEDs Working Paper.

Zhang, Y., H. J. Wang, and Z. Zhu (2019). “Quantile-regression-based clustering for panel data”. In: *Journal of Econometrics* 213.1, pp. 54–67.

Appendices

Appendix 1 - Clustering panel procedure for dichotomous models

In this appendix, we detail the procedure used to compute logit regression-based clustering for panel data. Our approach is inspired by the method of Zhang, Wang and Zhu (2019) for quantile regression-based clustering for panel data.

For clarity, we restate the dichotomous panel model being considered:

$$P_{t-1}(y_{i,t}=1)=F(\pi_{i,t})=F(\beta^{'}x_{i,t-1}+\alpha y_{i,t-1}+\delta\pi_{i,t-1}+\eta_i),$$
$$\text{for } t=1,2,...,T, \text{ and } i=1,2,...,N,$$

where N is the number of countries in the panel, $P_{t-1}(y_{i,t}=1)$ is the conditional probability of observing a recession at time t in country i, and π_t is the index at time $t-1$. F is the logistic c.d.f. Variable η_i is a country fixed effect for the control of unobserved heterogeneity and potential bias. We define θ as the vector of the estimated coefficients $[\beta, \alpha^T, \delta^T]^T$.

The goal of this estimation is twofold: first, to identify the subgroup membership in order to test partial homogeneity of our sample, and second, to obtain accurate estimation of group-specific parameters.
The computation of logit regression-based clustering consists of performing the following steps:

1. Fit a logit regression for each country and estimate the fixed effect η_i by $\widetilde{\eta}_i$ with

$$\text{LogL}(\widetilde{\theta}_i,\widetilde{\eta}_i)=\sum_{t=1}^{T}[y_{it}\text{log}(F(\pi_{i,t}(\theta_i,\eta_i)))+(1-y_{it})\text{log}(1-F(\pi_{i,t}(\theta_i,\eta_i)))].$$

2. Consider each country in the sample to be associated with a value from 1 to N, where N is the number of countries. Draw a set of data from 1 to N without replacement, denoted by Ω.

3. Take the first value of Ω, denoted by Ω_1, and create the first group with the associated country. Next, initialize parameters $\theta_g = \theta_1$ as follows:

$$\text{LogL}(\widetilde{\theta}_1) = \sum_{t=1}^{T}[y_{(\Omega_1,t)}\log(F(\pi_{(\Omega_1,t)}(\theta_1, \widetilde{\eta}_{\Omega_1}))) + (1 - y_{(\Omega_1,t)})\log(1 - F(\pi_{(\Omega_1,t)}(\theta_1, \widetilde{\eta}_{\Omega_1})))].$$

4. Iterate and consider each subsequent value Ω_i in Ω for $i = 2$, ... , N one-by-one. For each value, test the country's membership in the already existing groups by measuring a new θ_g, as well as the possibility that the country is in a new group alone. To this end, for each new country being added, take the solution that minimizes the Pearson residuals, proceeding as follows:

   For $i = 2$, ... , N:

   (a) Consider the number G of existing groups. Note that in the beginning, for $i = 2$, the number of groups is equal to 1, and the only group contains country Ω_1. Add the subsequent country associated with Ω_i, and estimate for $g = 1$, ... , $G + 1$ the new θ_g:

   $$\text{LogL}(\widetilde{\theta}) = \sum_{i=1}^{G}\sum_{t=1}^{T}[y_{it}\log(F(\pi_{i,t}(\theta_g, \widetilde{\eta}_i))) + (1 - y_{it})\log(1 - F(\pi_{i,t}(\theta_g, \widetilde{\eta}_i)))].$$

   (b) Vector θ associated with each group is estimated with a constrained value of $\widetilde{\eta}$. To allow convergence of the estimators, estimate a new η_i for each country by fixing θ based on the last estimated $\widetilde{\theta}$:

   $$\text{LogL}(\widetilde{\eta}_i) = \sum_{t=1}^{T}[y_{i,t}\log(F(\pi_{i,t}(\widetilde{\theta}_g, \eta_i))) + (1 - y_{i,t})\log(1 - F(\pi_{i,t}(\widetilde{\theta}_g, \eta_i)))],$$

   where country i belongs to group g.

   (c) Repeat (a) and (b) a sufficiently large number of times to observe convergence of estimator $\widetilde{\theta}_g$ and η_i for $g = 1$, ... , $G + 1$ and for $i = 1$, ... , N.

   (d) Among $G + 1$ estimations, select the estimation that minimizes the Pearson residuals:

   $$Residual \quad Sum =_{g\in[1,G+1]} \sum_{j=1}^{g}\frac{Y_j - F(\pi_j)}{F(\pi_j)(1 - F(\pi_j))},$$

   where *Residual Sum* is the minimum of the sum of Pearson residuals.

5. Repeat items 2-4 for $s = 1$, ... , S, where S is the total number of simulations. Compare the final *Residual Sum* obtained in 4.(d) for each simulation, and save the value corresponding to the minimum over all simulations.

6. For all groups obtained in steps 1-5, re-estimate θ_g and η_i, the fixed effect of each country that belongs to the associated group g. A Driscoll-Kraay (1998) correction is implemented to avoid bias due to cross-sectional dependence for all groups.

$$\begin{aligned}\text{LogL}(\theta_g, \eta_i) &= \sum_{i=1}^{N} \text{LogL}_i(\theta_g, \eta_i) \\ &= \sum_{i=1}^{N} \sum_{t=1}^{T} [y_{it}\text{log}(F(\pi_{i,t}(\theta_g, \eta_i))) + (1 - y_{it})\text{log}(1 - F(\pi_{i,t}(\theta_g, \eta_i)))].\end{aligned}$$

Appendix 2 - Replicating results using ECRI data - NBER recessions

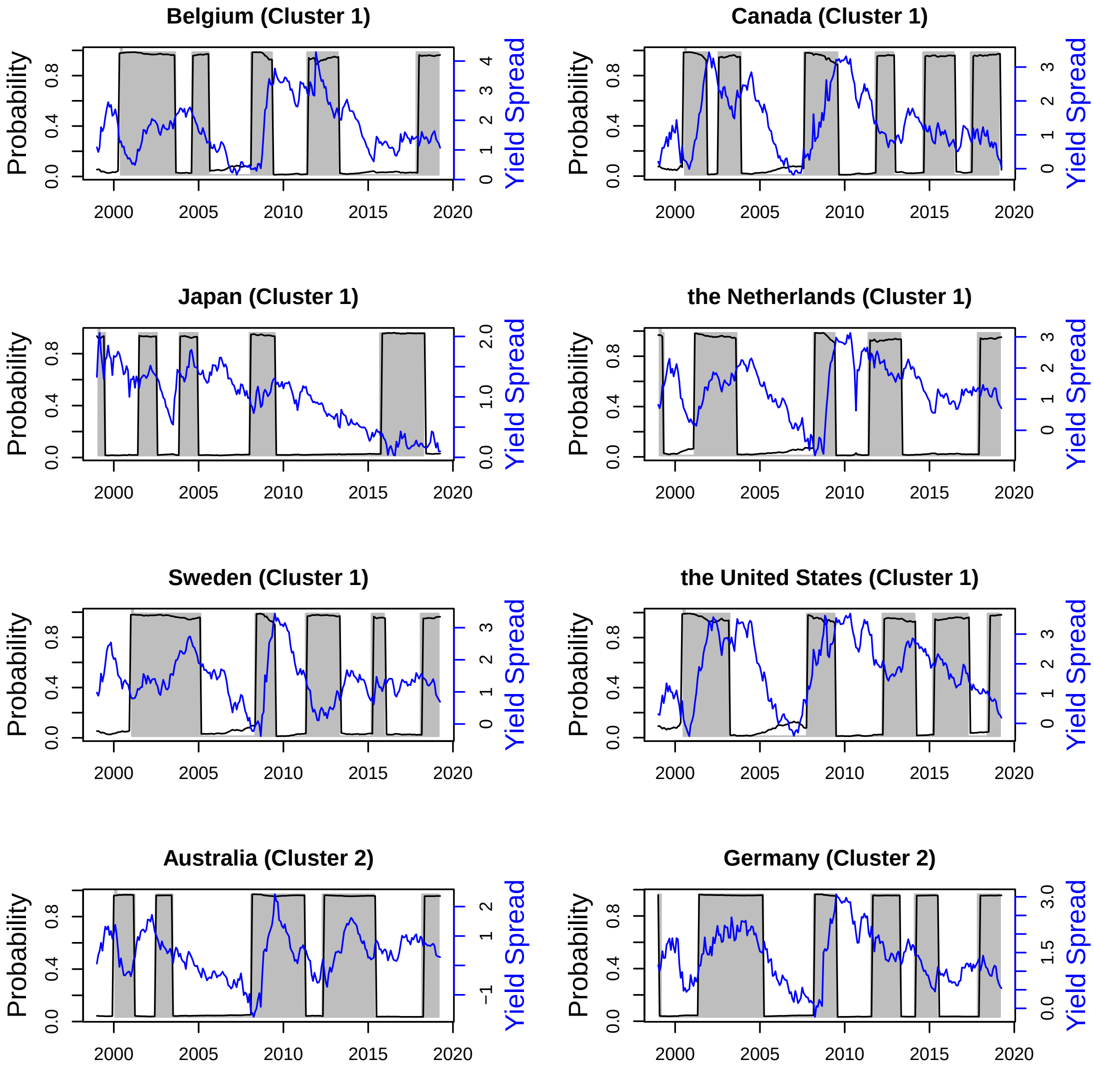
Table 3.10: Estimation results of panel logit models – Monthly frequency – 1975-2016

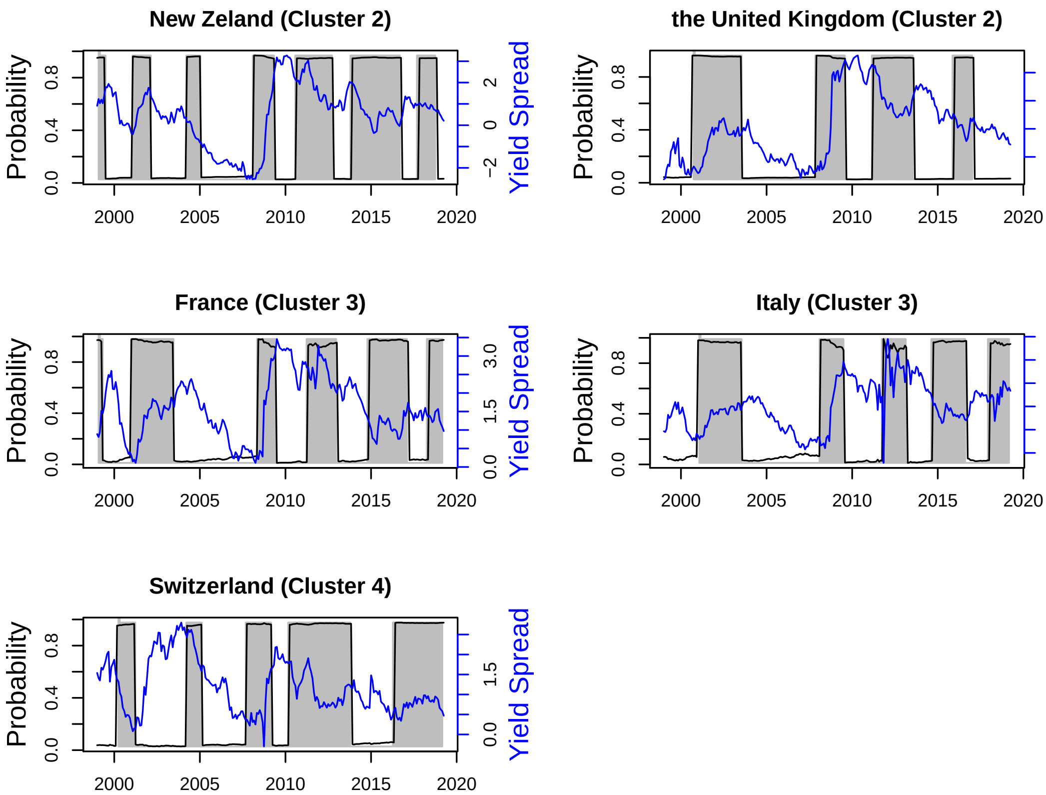
Model	(1)	(2)	(3)	(4)
$YSPR_{-1}$	-0.3329***	-0.3771***	-0.1645***	-0.3641***
	(0.0565)	(0.0658)	(0.0171)	(0.0647)
REC_{-1}		7.3903***		7.3345***
		(0.1895)		(0.4392)
$Index_{-1}$			-1.0336***	-1.0449***
			(0.1421)	(0.5642)
		Relevant Statistics		
BIC	5,172.2	1,057.33	5,077.4	1,066.0
		Fixed Effects		
Country	Yes	Yes	Yes	Yes
No. Observations	5,424	5,424	5,424	5,424

Notes: This table reports the estimates obtained from static and dynamic logit models (1)-(4) for a panel of 11 countries covering the period from February 1975 to March 2016 at the monthly frequency with one lag. The dependent variable is the recession dummy extracted from NBER database for the United States and from ECRI database for the others. We report Bayesian (BIC) information criteria for each specification. Results are computed using R 3.6.0 (R Core Team, 2020) and the *ews* (*v0.1.0*; Hasse and Lajaunie, 2020) package. The full reproducible code is available on CRAN. Standard errors are reported in parentheses below the estimates. Labels $***$, $**$ and $*$ indicate significance at 99%, 95% and 90% levels, respectively.

Appendix 3 - The predictive power of the yield spread

Figure 3.1: The predictive power of the yield curve from 1999 to 2019





Notes: This figure plots the yield curves (blue curves), observed and fitted recessions (grey areas and black curves respectively) from 1999 to 2019. Results are estimated using a dynamic logit model. Country-level results indicate that the yield spread signals recessions. Country clusters highlight common features about the predictive power of the yield curve.

Appendix 4 - EWS Package and Documentation

Package ‘EWS’

April 7, 2020

Type Package

Title Early Warning System

Version 0.1.0

Description Early Warning Systems (EWS) are a toolbox for policymakers to prevent or attenuate the impact of economic downturns. Modern EWS are based on the econometric framework of Kauppi and Saikkonen (2008) <doi:10.1162/rest.90.4.777>. Specifically, this framework includes four dichotomous models, relying on a logit approach to model the relationship between yield spreads and future recessions, controlling for recession risk factors. These models can be estimated in a univariate or a balanced panel framework as in Candelon, Dumitrescu and Hurlin (2014) <doi:10.1016/j.ijforecast.2014.03.015>. This package provides both methods for estimating these models and a dataset covering 13 OECD countries over a period of 45 years. This package constitutes a useful toolbox (data and functions) for scholars as well as policymakers.

Depends R (>= 2.10)

License GPL-3

Encoding UTF-8

LazyData true

Imports numDeriv

NeedsCompilation no

Author Quentin Lajaunie [cre],
Jean-Baptiste Hasse [aut]

Maintainer Quentin Lajaunie <quentin_lajaunie@hotmail.fr>

Repository CRAN

Date/Publication 2020-04-07 15:00:22 UTC

R topics documented:

`data_panel` *Historical data for 13 OECD countries*

Description

data_USA contains: - OECD based Recession Indicators for 13 OECD countries from the Peak through the Trough from 1975:03 to 2019:05 - Yield Spread (10Years TB minus 3Months TB) for 13 OECD countries from 1975:03 to 2019:05

List of countries: Australia, Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, New Zealand, Sweden, Switzerland, the United Kinkdom, the United States.

Usage

```
data("data_panel")
```

Format

A data frame with 6903 observations on the following 4 variables.

`country` List of countries.

`Date` Vector of dates.

`YIESPR` historical yield spread for the 13 OECD countries.

`OECD_Recession` Historical binary variable related to historical recessions for the 13 OECD countries.

Source

https://fred.stlouisfed.org/

`data_USA` *Historical data for the United States*

Description

data_USA contains: - NBER based Recession Indicators for the United States from 1975:03 to 2019:05 - Yield Spread (10Years TB minus 3Months TB) for the United States from 1975:03 to 2019:05

Usage

```
data("data_USA")
```

Format

A data frame with 531 observations on the following 4 variables.

`country` USA.

`Date` Vector of dates.

`YIESPR` Historical yield spread.

`NBER` Historical binary variable related to historical recessions.

Source

https://fred.stlouisfed.org/

`Logistic_Estimation` *Logistic Estimation for Dichotomous Analysis*

Description

This function provides methods for estimating the four dichotomous models as in Kauppi & Saikkonen (2008). Based on a logit approach, models are estimated in a univariate or a balanced panel framework as in Candelon, Dumitrescu and Hurlin (2014). This estimation has been used in recent papers such in Ben Naceur, Candelon and Lajaunie (2019) and Hasse and Lajaunie (2020).

Usage

```
Logistic_Estimation(Dicho_Y, Exp_X, Intercept, Nb_Id, Lag, type_model)
```

Arguments

`Dicho_Y`	Vector of the binary time series.
`Exp_X`	Vector or Matrix of explanatory time series.
`Intercept`	Boolean value: TRUE for an estimation with intercept, and FALSE otherwise.
`Nb_Id`	Number of individuals studied for a panel approach. Nb_Id=1 in the univariate case.
`Lag`	Number of lags used for the estimation.
`type_model`	Model number: 1, 2, 3 or 4.

-> 1 for the static model:

$$P_{t-1}(Y_t) = F(\pi_t) = F(\alpha + \beta' X_t)$$

-> 2 for the dynamic model with lag binary variable:

$$P_{t-1}(Y_t) = F(\pi_t) = F(\alpha + \beta' X_t + \gamma Y_{t-l})$$

-> 3 for the dynamic model with lag index variable:

$$P_{t-1}(Y_t) = F(\pi_t) = F(\alpha + \beta' X_t + \eta \pi_{t-l})$$

-> 4 for the dynamic model with both lag binary variable and lag index variable:

$$P_{t-1}(Y_t) = F(\pi_t) = F(\alpha + \beta' X_t + \eta \pi_{t-l} + \gamma Y_{t-l})$$

Value

A list with:

Estimation	a dataframe containing the coefficients of the logitic estimation, the Standard Error for each coefficient, the Z-score and the associated critical probability
AIC	a numeric vector containing the Akaike information criterion
BIC	a numeric vector containing the Bayesian information criterion
R2	a numeric vector containing the Pseudo R Square
LogLik	a numeric vector containing the Log likelihood value of the estimation
VCM	a numeric matrix of the Variance Covariance of the estimation

Note

For the panel estimation, data must be stacked one after the other for each country or for each individual.

Author(s)

Jean-Baptiste Hasse and Quentin Lajaunie

References

Candelon, Bertrand, Elena-Ivona Dumitrescu, and Christophe Hurlin. "Currency crisis early warning systems: Why they should be dynamic." International Journal of Forecasting 30.4 (2014): 1016-1029.

Hasse, Jean-Baptiste, Lajaunie Quentin. "Does the Yield Curve Signal Recessions? New Evidence from an International Panel Data Analysis." (2020)

Kauppi, Heikki, and Pentti Saikkonen. "Predicting US recessions with dynamic binary response models." The Review of Economics and Statistics 90.4 (2008): 777-791.

Naceur, Sami Ben, Bertrand Candelon, and Quentin Lajaunie. "Taming financial development to reduce crises." Emerging Markets Review 40 (2019): 100618.

Examples

```
# First Example: univariate analysis of the predictive power of the yield spread

# NOT RUN {


# Import data
data("data_USA")

# Data process
Var_Y <- as.vector(data_USA$NBER)
Var_X <- as.vector(data_USA$Spread)
```

```
# Estimate the logit regression
results <- Logistic_Estimation(Dicho_Y = Var_Y, Exp_X = Var_X, Intercept = TRUE,
                      Nb_Id = 1, Lag = 1, type_model = 4)

# print results
results


# }


# Second Example: panel analysis of the predictive power of the yield spread

# NOT RUN {


# Import data
data("data_panel")

# Data process
Var_Y <- as.vector(data_panel$OCDE)
Var_X <- as.vector(data_panel$Spread)

# Estimate the logit regression
results <- Logistic_Estimation(Dicho_Y = Var_Y, Exp_X = Var_X, Intercept = TRUE,
                        Nb_Id = 13, Lag = 1, type_model = 4)

# print results
results


# }
```

Matrix_lag *Matrix Lag - data processing*

Description

Compute a lagged version of a time series, shifting the time base back by a given number of observations defined by the user. The user must enter three parameters for this function: the matrix, the number of lags, and of boolean variable calls 'beginning'. If 'beginning'=TRUE, then the lag will be applied at the beginning of the matrix whereas if 'beginning'=FALSE, then the lag will be applied at the end of the matrix.

Usage

```
Matrix_lag(Matrix_target, Nb_lag, beginning)
```

Arguments

`Matrix_target`	Initial Matrix
`Nb_lag`	Number of lag
`beginning`	Boolean variable. If 'place'=TRUE, the lag is applied at the beginning of the matrix. If 'place'=FALSE, the lag is applied at the end of the matrix.

Value

A numeric Matrix.

Examples

```
# Initialize the following matrix
Matrix_example <- matrix(data=(1:10), nrow=5, ncol=2)

# Use Matrix_lag
new_matrix <- Matrix_lag(Matrix_target = Matrix_example, Nb_lag = 2, beginning = TRUE)

new_matrix

# Results:
#> new_matrix
#     [,1] [,2]
#[1,]    2    7
#[2,]    3    8
#[3,]    4    9
#[4,]    5   10
```

`Vector_lag` *Vector lag - data processing*

Description

Compute a lagged version of a time series, shifting the time base back by a given number of observations defined by the user. The user must enter three parameters for this function: the vector, the number of lags, and a boolean variable named 'beginning'. If 'beginning'=TRUE, then the lag will be applied at the beginning of the vector whereas if 'beginning'=FALSE, then the lag will be applied at the end of the vector.

Usage

```
Vector_lag(Vector_target, Nb_lag, beginning)
```

Arguments

`Vector_target`	Initial vector
`Nb_lag`	Number of lag
`beginning`	Boolean variable. If 'beginning'=TRUE, the lag is applied at the beginning of the vector. If 'beginning'=FALSE, the lag is applied at the end of the vector.

Value

A numeric Vector.

Examples

```
# Initialize the following vector
vector_example <- as.vector(1:10)

# Use Vector_lag
new_vector <- Vector_lag(Vector_target = vector_example, Nb_lag = 2, beginning = TRUE)

new_vector
# Results:
#> new_vector
#[1]  3  4  5  6  7  8  9 10
```

Index

Appendix 5-1 - Estimations obtained from dynamic logit model (2) for each of the 13 countries

Table 3.11: Estimation results for each country – Monthly frequency – 1999-2019

	$Intercept$	-3.308***		$Intercept$	-2.419***
		(0.540)			(0.473)
	$YSPR_1$	-0.084		$YSPR_1$	-0.709***
Australia		(0.157)	Netherlands		(0.213)
	REC_1	6.460***		REC_1	6.553***
		(0.457)			(0.504)
	$CBAN_1$	0.002		$CBAN_1$	-0.029
		(0.052)			(0.042)
	$Intercept$	-4.046***		$Intercept$	-3.096***
		(0.736)			(0.356)
	$YSPR_1$	0.100		$YSPR_1$	0.094
Belgium		(0.250)	New Zealand		(0.101)
	REC_1	6.411***		REC_1	6.351***
		(0.461)			(0.451)
	$CBAN_1$	0.124*		$CBAN_1$	0.011
		(0.074)			(0.088)
	$Intercept$	-2.445***		$Intercept$	-4.297***
		(0.569)			(0.708)
	$YSPR_1$	-0.564***		$YSPR_1$	-0.328**
Canada		(0.200)	Sweden		(0.167)
	REC_1	6.207***		REC_1	6.667***
		(0.454)			(0.524)
	$CBAN_1$	-0.062		$CBAN_1$	0.300**
		(0.062)			(0.125)
	$Intercept$	-2.239***		$Intercept$	-2.790***
		(0.674)			(0.524)
	$YSPR_1$	-0.775***		$YSPR_1$	-0.398**
France		(0.252)	Switzerland		(0.167)
	REC_1	6.592***		REC_1	6.385***
		(0.508)			(0.465)
	$CBAN_1$	-0.075		$CBAN_1$	-0.03
		(0.145)			(0.045)

Notes: This table reports the estimates obtained from dynamic logit model (2) for each of the 13 countries in the panel covering the period from January 1999 to March 2019 at the monthly frequency with one lag. The dependent variable is the recession dummy. Results are computed using R 3.6.0 (R Core Team, 2020) and the *ews* (*v0.1.0*; Hasse and Lajaunie, 2020) package. The full reproducible code is available on CRAN. Standard errors are reported in parentheses below the estimates. Labels $***$, $**$ and $*$ indicate significance at 99%, 95% and 90% levels, respectively.

Appendix 5-2 - Estimations obtained from dynamic logit model (2) for each of the 13 countries

Table 3.12: Estimation results for each country – Monthly frequency – 1999-2019

Germany	$Intercept$	-2.475***	United Kingdom	$Intercept$	-3.148***
		(0.608)			(0.477)
	$YSPR_1$	-0.671***		$YSPR_1$	-0.436***
		(0.234)			(0.169)
	REC_1	6.369***		REC_1	6.486***
		(0.480)			(0.492)
	$CBAN_1$	-0.023		$CBAN_1$	-0.007
		(0.101)			(0.059)
Italy	$Intercept$	-3.328***	United States	$Intercept$	-2.968***
		(0.556)			(0.642)
	$YSPR_1$	-0.174***		$YSPR_1$	-0.497**
		(0.156)			(0.242)
	REC_1	6.453***		REC_1	6.755***
		(0.465)			(0.519)
	$CBAN_1$	0.017		$CBAN_1$	0.085
		(0.068)			(0.074)
Japan	$Intercept$	-3.502***			
		(0.575)			
	$YSPR_1$	0.166			
		(0.241)			
	REC_1	6.314***			
		(0.455)			
	$CBAN_1$	0.009			
		(0.050)			

Notes: This table reports the estimates obtained from dynamic logit model (2) for each of the 13 countries in the panel covering the period from January 1999 to March 2019 at the monthly frequency with one lag. The dependent variable is the recession dummy. Results are computed using R 3.6.0 (R Core Team, 2020) and the *ews* (*v0.1.0*; Hasse and Lajaunie, 2020) package. The full reproducible code is available on CRAN. Standard errors are reported in parentheses below the estimates. Labels $***$, $**$ and $*$ indicate significance at 99%, 95% and 90% levels, respectively.

Chapter 4

Nonlinear Impulse Response Function for Dichotomous Models

Author: Quentin Lajaunie

About this chapter

The author thanks Bertrand Candelon, Jean-Baptiste Hasse, Christophe Hurlin and Yannick Le Pen for helpful comments. This research was performed as part of a research program titled "Risk Management, Investment Strategies and Financial Stability" under the aegis of the Europlace Institute of Finance, a joint initiative with insti7. The usual disclaimer applies.

Abstract

In this paper, I propose a generalized impulse response function (GI) for dichotomous models. Building on Kauppi and Saikkonen (2008), I develop the exact form of the response functions for each specification of their binary model. Using a block-bootstrap method, I compute robust confidence intervals for these response functions. I illustrate the usefulness of this analytical result for static and dynamic dichotomous models of U.S. recessions. According to the different specifications, I empirically find that the persistence of an impact of an exogenous shock to the U.S. economy is between one to five quarters.

Keywords: Impulse response functions ; Dichotomous model ; Recession prediction.

4.1 Introduction

The aim of this paper is to disentangle the impulse and propagation mechanisms in a univariate framework for dichotomous models. A systematic analysis of the response functions is proposed for this class of models. Specifically, a generalized impulse response function (GI) is developed for each specification of the four binary response models of Kauppi and Saikkonen (2008). Using a dichotomous model to estimate the relationship between yield spreads and future recessions, we assess the impact of an exogenous shock to U.S. business cycles.

The seminal paper of Sims (1980) introduced the impulse-response methodology and the vector autoregressive (VAR) representation. This new tool makes it possible to analyze the propagation mechanisms of an identified shock in a dynamic system of equations. Impulse-response analysis has been refined by Doan, Litterman and Sims (1984) and others. An impulse response function makes it possible to investigate the time profile of a shock or innovation on the behavior of a series. The analysis is done throughout a horizon h, conditional on the information available at time $t-1$. A common approach consists of considering a shock of delta size that occurs at time t. Next, two time series should be compared: one that is affected by this shock and another that is not. For such an investigation, the assumption that any other shock occurs between periods t and $t+h$ is made for both time series.

The methodology defined above has been widely used in many empirical studies. However, it should be noted that the analysis of the propagation of the shock depends both on the model used and on the value of the initial shock. For instance, among the early studies, Campbell and Mankiw (1987) and Diebold and Rudebusch (1989) studied the impact of an unexpected positive change in real GNP for the U.S. using an ARMA process. They concluded that a shock of one percent was very persistent. Considering two shocks of opposite sign, Beaudry and Koop (1993) provided analytical evidence of asymmetric shock persistence. The results indicated that negative innovation is less persistent than positive innovation. Potter (1995) found similar evidence of asymmetric response using a nonlinear SETAR model. As a consequence, Gallant, Rossi and Tauchen (1993) (GRT hereafter), Koop Pesaran and Potter (1996) (KPP hereafter) and Potter (2000) provided a unified approach to impulse response analysis. This approach can be used for both linear and nonlinear models and is called the generalized impulse response function (GI hereafter).

This new class of GI is constructed from the difference between two conditional means. One is the average of the expected vector conditional on history and the current shock, and the other is the baseline, which is conditional only on history. The GRT and KPP approaches are quite similar. However, GRT consider a delta shock of fixed size, specified by the investigator, while KPP directly generate this shock from the empirical distribution of innovations in the time series. The framework given by the GI makes it possible to analyze the persistence of a shock. However, this new tool differs depending on the model on which it is used. For example, Hafner and Herwartz (2006) proposed introducing a new concept of response function from the form proposed by KPP. Their response function is related to the conditional variance, unlike KPP, who focused on the conditional mean.

Impulse response analysis for dichotomous models has yet to be investigated. These models have been widely used and improved for the prediction of recessions. Estrella and Hardouvelis (1991) introduced the use of a binary model in the U.S. This resulted in numerous works using probit models (Dotsey, 1998; Estrella, Rodrigues and Schich, 2003; Rosenberg and Maurer, 2008) and logit models (Sensier et al., 2004; Moneta, 2005) in a univariate and static framework. Chauvet and Potter (2002 and 2005) proposed a first improvement by adding a dependency with the lagged latent variable.[1] Kauppi

[1] See also Dueker (2005).

and Saikkonen (2008) (KS hereafter) proposed a more general form in which the lagged binary variable can also be taken into account to improve the prediction of recessions.[2] Specifically, they proposed four different specifications. The first is a static approach, in which only exogenous variables are taken into account. The second and third specifications are dynamic and include either lagged values of the binary variable or lagged values of the dependent variable. Finally, the last specification combines the two preceding cases and includes both lagged binary variables and the lagged index.

We provide an impulse response for these four specifications using the method defined by KPP (1996). However, there is an important difference. For dichotomous models, the shock has an impact on the latent variable of the model. However, the binary variable is calculated conditional on this latent variable. Thus, the conditional expectation using the mean of the response vector conditional on history and a present shock compared to a baseline can be applied only for the latent variable. The response function of the binary variable must be calculated simultaneously and conditional on the values of the latent variables with and without the shock. Specifically, the response function for dichotomous models must be formalized as a system of two equations. The first equation is focused on the latent variable of the model, and the second is focused on the conditional binary variable. From this system of two equations, I propose to define the response functions specific to each model proposed by KS.

Note that the dynamics of the impulse response for the model containing both latent variables and binary variables are much more complex. Indeed, if the shock to the latent variable causes a change in the binary variable, the next latent variable could be impacted by both its autoregressive value and the lagged binary variable. Thus, to facilitate understanding, a finite form of the response function for this model is proposed and demonstrated. On the other hand, we propose to estimate the confidence intervals using a block-bootstrap method. The method of Hall, Horowitz and Jing (1995) is used to determine the size of the blocks. Then, we follow KPP and directly generate shocks from the empirical distributions of innovations in each simulation. Finally, these simulations allow me to evaluate confidence intervals.

This methodology is applied to study the impact of an exogenous shock in the U.S.

[2] Candelon, Dumitrescu and Hurlin (2013) proposed a multivariate approach estimated with an exact maximum-likelihood approach.

and compare the results of the four specifications defined above. This makes it possible to observe the persistence of the shock in different configurations. Whichever specification is used, the results show that the U.S. economy will fall into recession in the next quarter. As KS, we use the Akaike information criterion (AIC) and the Schwarz Bayesian criterion (BIC) to select the best models. The second and fourth specifications downplay these criteria. Concerning the second specification that includes a lagged value of the binary variable, the shock is not persistent, and the recession lasts only one quarter. However, the shock is much more persistent for the last specification that includes both a lagged binary variable and a lagged index. In this case, the recession would last 5 quarters.

The remainder of this paper is organized as follows. The frameworks employed for the impulse response function and binary time series models are introduced in Section 4.2. In Section 4.3, the impulse response function of each specification is formalized. The estimation and the moving block-bootstrap procedure for confidence intervals are described in Section 4.4. In Section 4.5, after having estimated the 4 specifications proposed by KS, we discuss the empirical results of the consequences of an exogenous shock to business cycles in the United States. Section 6 concludes the paper.

4.2 The model

Gallant Rossi and Tauchen (1993) define the generalized impulse response function (GI) as the difference between the expected value of a time series that has been hit by a shock at time h minus the expected value of the same time series without any shock for a given horizon h. They treat the shock as a hypothetical value of size δ. Another definition of the GI for both linear and nonlinear econometric models was proposed by Koop et al. (1996) (KPP) and Potter (2000). They consider that the shock is directly generated from the empirical distribution of innovations in the time series. We use the definition proposed by KPP to analyze the time profile of an exogenous shock in the dichotomous model framework. In our case, we assume that the shock enters in an additive manner.

$$GI_Y(h, \nu_t = \delta, \Omega_{t-1}) = E(Y_{t+h}|\nu_t, \Omega_{t-1}) - E(Y_{t+h}|\Omega_{t-1}) \tag{4.1}$$

To the best of our knowledge, the GI for the dichotomous model has not yet been developed. First it is important to recall the definition of a dichotomous model.

Definition 1 *The dichotomous model defines the probability that $y_t = 1$ is the value of the cumulative distribution function F(.) for π_t:*

$$P_{t-1}(y_t = 1) = F(\pi_t|\Omega_{t-1}),$$

where π_t is denoted as the index or the latent variable below and Ω_{t-1} is the information set available at time $t-1$.

In the dichotomous framework, exogenous explanatory variables allow us to estimate the probability that the binary variable takes value 1 for a given period. Since the seminal paper of Estrella and Hardouvelis (1991) introduced the use of dichotomous models to forecast recessions, these models have been greatly improved. Among these improvements, two in particular are noteworthy. The first is that developed by Chauvet and Potter (2005). In their model, they included the value of the lagged index in the vector of explanatory variables. This allows a richer dynamic to estimate the process of the dependent variable and a noticeable improvement in forecasts. Kauppi and Saikkonen (2008) (KS hereafter) extended this new dichotomous model by adding the lagged binary variable in the estimation. Thus, the general dichotomous model can be written as:

$$P_{t-1}(y_t = 1) = F(\pi_t) = F(\beta^{'} X_{t-1} + \sum_{j=1}^{p} \gamma_j y_{t-j} + \sum_{j=1}^{q} \eta_j \pi_{t-j}), \tag{4.2}$$

with

$$y_t = \begin{cases} 1 & \text{if } \pi_t > 0 \\ 0 & \text{if } \pi_t \leq 0 \end{cases},$$

where y_t is the dichotomous variable at time t, π_t is the latent variable at time t, and X_{t-1} is a $k \times 1$ vector of k exogenous variables. β is a $k \times 1$ vector that corresponds to the estimated slopes associated with each explanatory variable of X_{t-1}. Moreover, γ_j with $j = 1...p$ and η_j with $j = 1...q$ are the estimated slopes associated with the lagged y_{t-i} and π_{t-j}, respectively. p is the number of lags for the dichotomous variable, and q is the number of lags for the index variable. $F(.)$ is a function of R which takes value in [0,1] and monotonic increasing or strictly monotonic increasing.[3]

[3]The most popular dichotomous models are the probit and logit; see, for example, Estrella et al. (1991), Chauvet and Potter (2002, 2005), Dueker (2005), Kauppi and Saikkonen (2008), and Candelon et al. (2013, 2014), Naceur et al. (2019) and Hasse and Lajaunie (2020). F(.) is a Gaussian c.d.f for the probit model, and F(.) is a logistic c.d.f for the logit model.

In their paper, KS decomposed equation (4.2) into four different specifications. The first specification is static, with two restrictions: $\gamma = \eta = 0$ (Model 1). The second and third specifications are dynamic and include either one or more lagged values of the binary variable y_{t-1} with a restriction on $\eta = 0$ (Model 2) or one or more lagged indices π_{t-1} with a restriction on $\gamma = 0$ (Model 3). Finally, the last dynamic model combines the two preceding cases and includes both lagged binary variables y_{t-1} and a lagged index π_{t-1} (Model 4). The relationship between the index and the binary variable varies from one model to another. The response to an exogenous shock is therefore highly variable. We will study each GI's time profile associated with these models to estimate the different dynamics that arise from them. We define the 4 dichotomous models as follows:

Definition 2 *Let us define Model i for i = 1, 2, 3 or 4 as the dichotomous specification:*

$P_{t-1}(y_t^{(1)} = 1) = F(\pi_t^{(1)}) = F(\beta' X_{t-1})$ *for Model 1,*

$P_{t-1}(y_t^{(3)} = 1) = F(\pi_t^{(2)}) = F(\beta' X_{t-1} + \sum_{j=1}^{p} \gamma_j y_{t-j}^{(2)})$ *for Model 2,*

$P_{t-1}(y_t^{(2)} = 1) = F(\pi_t^{(3)}) = F(\beta' X_{t-1} + \sum_{j=1}^{q} \eta_j \pi_{t-j}^{(3)})$ *for Model 3, and*

$P_{t-1}(y_t^{(4)} = 1) = F(\pi_t^{(4)}) = F(\beta' X_{t-1} + \sum_{j=1}^{p} \gamma_j y_{t-j}^{(4)} + \sum_{j=1}^{q} \eta_j \pi_{t-j}^{(4)})$ *for Model 4.*

If one or more $|\eta_j|$ is greater than 1, this can lead to a constant increase in the latent variable. This would result in a permanent regime or a regime that oscillates perpetually. To avoid this scenario, as a sufficient condition, we make the following assumption:

Assumption 1 $|\eta_j| < 1 \ \forall \ j \in [1, ..., q]$.

Remark 1 *We want to avoid the case where the latent variable would increase indefinitely, leading to the appearance of an absorbent state. This hypothesis guarantees the non-explosion of the latent variable, but can be relaxed and replaced by an less restrictive hypothesis alternative that guarantees the stationarity of the index. However, this has no impact on the response function formulations presented below.*

The dichotomous model developed by KS seeks to improve crisis forecasting in the U.S. They estimate whether the binary variable will take value one at a given horizon conditional on the value of the index and on crossing a threshold c. This threshold is generally equal to 0. Indeed, when $\pi = 0$, the associated probability is 0.5. Candelon Dumitrescu and Hurlin (2014) review some methods where the threshold is estimated endogenously.

These methods take into account two types of error: misidentified crisis and false signals. Misidentified crisis, also called type I error, arises when the probability is below the cutoff but a crisis occurs. False signals, also called type II error, arise when the probability is above the cutoff but no crisis occurs. The threshold c is estimated at time t and can be used as an indicator of a potential crisis. The different estimations focus on type I or type II error. Table 4.1 illustrates these two errors. On the other hand, the definition of the accuracy measure (Candelon Dumitrescu and Hurlin, 2014), called c_{AM}, is preferred in this paper. Indeed, c_{AM} makes it possible to take into account both type I error and type II error.

Table 4.1: Type I and type II error - EWS

	Crisis Period	Calm Period
$\widehat{Pr} > c$	-	Type II Error
$\widehat{Pr} \leq c$	Type I Error	-

Sources: Kaminsky et. al (1998) and Candelon et. al (2014)

The GI evaluates the time profile of the effect of shocks at a given horizon. In the area of the dichotomous model, this definition is appropriate for the index. Concerning the binary variable, it is determined conditional on this index and a threshold. The complexity of such a model, as presented in equation (4.2), lies in the interactions that exist between π and y. The shock that occurs at time t has a direct impact on the index. A sufficiently large shock can lead to a nonlinear transmission of the shock to the binary variable. The GI presented in equation (4.1) has to be written as a system of two equations for the dichotomous model. The first equation is GI for the index of the model, and the second equation is GI for the dichotomous variable.

Let $GI_\pi(h, \nu_t = \delta, \Omega_{t-1})$ and $GI_y(h, \nu_t = \delta, \Omega_{t-1})$ be the impulse responses of the index and the dichotomous variable for the horizon h, when a shock of size δ occurs at time t. Ω_{t-1} denotes the known history available up to time $t-1$. We can write the GI of the dichotomous model as the following system:

$$\begin{cases} GI_{\pi}(h, \nu_t = \delta, \Omega_{t-1}) = E(\pi_{t+h}|\nu_t, \Omega_{t-1}) - E(\pi_{t+h}|\Omega_{t-1}) \\ GI_{y}(h, \nu_t = \delta, \Omega_{t-1}) = E(y_{t+h}|E(\pi_{t+h}|\nu_t, \Omega_{t-1}), c) - E(y_{t+h}|E(\pi_{t+h}|\Omega_{t-1}), c) \end{cases}, \quad (4.3)$$

where c represents the threshold above which the binary variable takes the value of 1.

The first line of the system in equation (4.3) allows us to study the time profile of the shock on the index by calculating the difference in expectations between the series for which the δ shock occurs, and the series without shock. The binary variable is computed conditionally at the index level. Its value also depends on the threshold c. Thus, the second line of the system represents the response function of the binary variable based on the conditional expectations of the latent variable π, and of c.

Remark 2 *In the system presented in equation (4.3), the equation for the binary variable y can only take values of -1, 0 or 1.*

To study the implications of the system of equations (4.3), we first analyze the expected values of the index and the binary variable at time t when the exogenous shock occurs. Thereafter, we will study the forms of each response function specific to the specifications presented in Definition 2.

When the exogenous shock occurs in period t, the expected value of $GI_{\pi}(0, \nu_t = \delta, \Omega_{t-1})$ is always equal to δ. Concerning the binary variable, it is determined conditional on the value of the latent variable π. Thus, the consequences for this variable may vary depending on the sign and size of the shock. Occasionally, we can observe a nonlinear transmission of the shock with a change in the value of the dichotomous variable if δ is large enough. Since the dichotomous variable is equal to 1 if the index is greater than 0, a change in the dichotomous variable can only be observed if π_t and δ have opposite signs. On the other hand, when π_t and δ have the same sign, $GI_y(h, \nu_t = \delta, \Omega_{t-1})$ is always equal to 0.

For all $\delta > 0$ and for $h = 0$, the $GI(.)$ system can be written as follows:

$$\begin{cases} GI_\pi(h=0,\delta,\Omega_{t-1}) = \delta \\ \begin{cases} GI_y(h=0,\delta,\Omega_{t-1}) = 1 \text{ , if } E(\pi_t|\nu_t,\Omega_{t-1}) > F^{-1}(c) \text{ and } E(F(\pi_t|\Omega_{t-1}) \leq F^{-1}(c) \\ GI_y(h=0,\delta,\Omega_{t-1}) = 0 \text{ , if } E(\pi_t|\nu_t,\Omega_{t-1}) > F^{-1}c \text{ and } E(\pi_t|\Omega_{t-1}) > F^{-1}(c) \\ GI_y(h=0,\delta,\Omega_{t-1}) = 0 \text{ , if } E(\pi_t|\nu_t,\Omega_{t-1}) \leq F^{-1}(c) \text{ and } E(\pi_t|\Omega_{t-1}) \leq F^{-1}(c) \end{cases} \end{cases} \tag{4.4}$$

Similarly, for all $\delta < 0$, we have:

$$\begin{cases} GI_\pi(h=0,\delta,\Omega_{t-1}) = \delta \\ \begin{cases} GI_y(h=0,\delta,\Omega_{t-1}) = -1 \text{ , if } E(\pi_t|\nu_t,\Omega_{t-1}) \leq F^{-1}(c) \text{ and } E(\pi_t|\Omega_{t-1}) > F^{-1}(c) \\ GI_y(h=0,\delta,\Omega_{t-1}) = 0 \text{ , if } E(\pi_t|\nu_t,\Omega_{t-1}) \leq F^{-1}(c) \text{ and } E(\pi_t|\Omega_{t-1}) \leq F^{-1}(c) \\ GI_y(h=0,\delta,\Omega_{t-1}) = 0 \text{ , if } E(\pi_t|\nu_t,\Omega_{t-1}) > F^{-1}(c) \text{ and } E(\pi_t|\Omega_{t-1}) > F^{-1}(c) \end{cases} \end{cases} \tag{4.5}$$

Remark 3 *The results of equations (4.4) and (4.5) are applicable to all models presented in Definition 2.*

4.3 Impulse Response Function

In this section, we evaluate *GI* for specifications 2, 3 and 4. An exact form of shock propagation is developed for each model. Since specification 1 is static, an exogenous shock will not be persistent, regardless of the effects of the shock on the system at the period when it last occurs. The study of the latter is therefore of no interest for a horizon h. As presented in equations (4.4) and (4.5), the shock may cause a change in the binary variable. However, for any $h > 0$, the value of the response function will be zero on both π and y. Specifications 2 and 4 that include the lagged binary variable are more complex. For this reason, we also explain the propagation mechanism in detail. In particular, we will explain how a shock to the index can lead to a nonlinear transmission to the explanatory binary variable and the impact that this transmission may have on the latent variable in the following period. To define these various response functions, we introduce the following notation:

Notation 1 *Let us denote by* $GI^{(i)}_\pi(h,\delta,\Omega_{t-1})$ *and* $GI^{(i)}_y(h,\delta,\Omega_{t-1})$ *the impulse response* *function of Model i for the index and the binary variable, respectively, with i = 1, 2, 3, or 4.*

We begin by studying the second specification, called $GI^{(2)}(.)$. The lagged binary variables are included in the estimation of the index. Thus, GI depends on the last p values of y_t. As previously described, when δ is large enough, a nonlinear transmission of the shock can occur. Therefore, the expectation of such an impulse response has a value different from 0 for a given horizon h when at least one of the p last regimes of the shocked series has been modified by δ. The shock can only be propagated thanks to the regime changes that previously occurred. Considering that such a nonlinear transmission takes place at period t, the value of the response function associated with the index will then take the value of gamma. This means that the latent value of the time series that has been hit by δ is equal to the benchmark to which gamma is added. If γ_1 is negative, the propagation of the shock will be made through oscillation. This could lead to a new variation in the binary variable in the opposite direction as the previous change. If gamma is positive, the shock will persist in the same direction. In other words, we have noted that a positive shock can cause a change of regime from state 0 to state 1 (and from 1 to 0 for a negative shock) in period t. However, from $h > 0$, whatever the sign of δ, a change of regime can be observed and return from state 1 to state 0. Therefore, the expectation of the response function depends on the sign of γ_j for $j = 1, ..., p$. The $GI^{(2)}(.)$ can be written as follows:

$$\begin{cases} GI_{\pi}^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) = \sum_{j=1}^{p} \gamma_j \left(E(y_{t+h-j}^{(2)} | E(\pi_{t+h} | \nu_t, \Omega_{t-1})) - E(y_{t+h-j}^{(2)} | E(\pi_{t+h} | \Omega_{t-1})) \right) \\ \begin{cases} GI_y^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) = 1 \\ \quad , \text{ if } E(\pi_{t+h}^{(2)} | \Omega_{t-1}) + GI_{\pi}^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) > F^{-1}(c) \text{ and } E(\pi_{t+h}^{(2)} | \Omega_{t-1}) \leq F^{-1}(c) \\ GI_y^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) = -1 \\ \quad , \text{ if } E(\pi_{t+h}^{(2)} | \Omega_{t-1}) + GI_{\pi}^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) \leq F^{-1}(c) \text{ and } E(\pi_{t+h}^{(2)} | \Omega_{t-1}) > F^{-1}(c) \\ GI_y^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) = 0 \\ \quad , \text{ if } E(\pi_{t+h}^{(2)} | \Omega_{t-1}) + GI_{\pi}^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) \leq F^{-1}(c) \text{ and } E(\pi_{t+h}^{(2)} | \Omega_{t-1}) \leq F^{-1}(c) \\ GI_y^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) = 0 \\ \quad , \text{ if } E(\pi_{t+h}^{(2)} | \Omega_{t-1}) + GI_{\pi}^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) > F^{-1}(c) \text{ and } E(\pi_{t+h}^{(2)} | \Omega_{t-1}) > F^{-1}(c) \end{cases} \end{cases} \tag{4.6}$$

Remark 4 *The value of* $GI_y^{(2)}(h, \nu_t = \delta, \Omega_{t-1})$ *is conditional on the value of* $E(\pi_{t+h}^{(2)}|\Omega_{t-1})$*. We can therefore rewrite equation (4.6) in two sub-cases*

$$\left\{\begin{array}{l} GI_\pi^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) = \sum_{j=1}^{p} \gamma_j \Big(E(y_{t+h-j}^{(2)}|E(\pi_{t+h}|\nu_t, \Omega_{t-1})) - E(y_{t+h-j}^{(2)}|E(\pi_{t+h}|\Omega_{t-1}))\Big) \\ \left\{\begin{array}{l} GI_y^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) = 1 \ , \ \text{if } E(\pi_{t+h}^{(2)}|\Omega_{t-1}) + GI_\pi^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) > F^{-1}(c) \\ GI_y^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) = 0 \ , \ \text{if } E(\pi_{t+h}^{(2)}|\Omega_{t-1}) + GI_\pi^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) \leq F^{-1}(c) \end{array}\right. \end{array}\right. , \tag{4.7}$$

for $E(\pi_{t+h}^{(2)}|\Omega_{t-1}) \leq F^{-1}(c)$.

$$\left\{\begin{array}{l} GI_\pi^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) = \sum_{j=1}^{p} \gamma_j \Big(E(y_{t+h-j}^{(2)}|E(\pi_{t+h}|\nu_t, \Omega_{t-1})) - E(y_{t+h-j}^{(2)}|E(\pi_{t+h}|\Omega_{t-1}))\Big) \\ \left\{\begin{array}{l} GI_y^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) = -1, if E(\pi_{t+h}^{(2)}|\Omega_{t-1}) + GI_\pi^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) \leq F^{-1}(c) \\ GI_y^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) = 0 \ , \ \text{if } E(\pi_{t+h}^{(2)}|\Omega_{t-1}) + GI_\pi^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) > F^{-1}(c) \end{array}\right. \end{array}\right. , \tag{4.8}$$

for $E(\pi_{t+h}^{(2)}|\Omega_{t-1}) > F^{-1}(c)$.

The third specification includes the lagged value of the dependent variable. Lagged values are added linearly and thereby allow for richer dynamics. The impulse response $GI_\pi^{(3)}(.)$ is quite similar to the impulse response function of an $AR(q)$, where q is the number of lags associated with π. Moreover, we have assumed that $|\eta_j| < 1 \ \forall \ j \in [1, ..., q]$ (Assumption (1)). Thus, the result of such GI converges to 0 when the horizon approaches infinity, meaning that $\lim\limits_{h\to\infty} E(\pi_{t+h}^{(3)}|\nu_t, \Omega_{t-1}) = \lim\limits_{h\to\infty} E(\pi_{t+h}^{(3)}|\Omega_{t-1})$. Similarly, $GI_y^{(3)}(.)$ converges to 0 when the horizon approaches infinity because of its dependence on the index value.

For example, if there is only one lag with a negative root associated with the latent variable, the effect of the shock will oscillate and converge towards 0. On the other hand, for the same shock, if the root is positive, then the response function will be monotonic and decreasing. Regarding the binary variable, changes can be observed throughout the convergence. Indeed, it is calculated conditional on the value of the index and on the value of the threshold c. Moreover, if a nonlinear transmission of the shock is observed for a given period, the following period is not directly impacted by it. Indeed, the index only depends on its last q values, and the lagged dichotomous variable is not integrated into its estimation.

We noted that for $h = 0$, a change in the binary variable is determined conditional not only on the previous variable of the dichotomous variable but also on the size and sign of δ. This means that we can observe a change from 0 to 1 for the dichotomous variable only if the shock is large enough, positive, and when the previous regime was 0. However, for all $h > 0$, whatever the sign of the shock, a change from regime 0 to regime 1 or from regime 1 to regime 0 can be observed. Indeed, the results of $GI_y^{(3)}(.)$ are also impacted by the sign of the slopes associated with the lagged index.

To write $GI_\pi^{(3)}(.)$ and $GI_y^{(3)}(.)$, we introduce the following notation:

$$z_t = \begin{bmatrix} \pi_t \\ \pi_{t-1} \\ \vdots \\ \pi_{t-q+1} \end{bmatrix}, \ \Lambda = \begin{bmatrix} \eta_1 & \eta_2 & \cdots & \eta_{q-1} & \eta_q \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}, \ and \ \Upsilon = \begin{bmatrix} \nu_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Then, such a time series underlying the index can be written as:

$$z_t = \Lambda . z_{t-1}.$$

From this notation, we obtain the following results:

$$GI_\pi^{(3)}(h, \nu_t = \delta, \Omega_{t-1}) = E(\pi_{t+h}^{(3)} | \nu_t, \Omega_{t-1}) - E(\pi_{t+h}^{(3)} | \Omega_{t-1}), \tag{4.9}$$

$$GI_\pi^{(3)}(h, \nu_t = \delta, \Omega_{t-1}) = [\Lambda^h \times \Upsilon]_{(1,1)} = [\Lambda^h]_{(1,1)} . \nu_t, \tag{4.10}$$

where $[.]_{(1,1)}$ corresponds to the first term of the matrix $[.]$.

Then, the system of GI that describes the propagation mechanisms of an exogenous shock for the third specification can be summarized as follows:

$$\begin{cases} GI_\pi^{(3)}(h,\nu_t=\delta,\Omega_{t-1})=(\Lambda^h)_{(1,1)}.\nu_t \\ \begin{cases} GI_y^{(3)}(h,\nu_t=\delta,\Omega_{t-1})=1 \\ \quad , \text{ if } E(\pi_{t+h}^{(3)}|\Omega_{t-1})+GI_\pi^{(3)}(h,\nu_t=\delta,\Omega_{t-1})>F^{-1}(c) \text{ and } E(\pi_{t+h}^{(3)}|\Omega_{t-1})\leq F^{-1}(c) \\ GI_y^{(3)}(h,\nu_t=\delta,\Omega_{t-1})=-1 \\ \quad , \text{ if } E(\pi_{t+h}^{(3)}|\Omega_{t-1})+GI_\pi^{(3)}(h,\nu_t=\delta,\Omega_{t-1})\leq F^{-1}(c) \text{ and } E(\pi_{t+h}^{(3)}|\Omega_{t-1})> F^{-1}(c) \\ GI_y^{(3)}(h,\nu_t=\delta,\Omega_{t-1})=0 \\ \quad , \text{ if } E(\pi_{t+h}^{(3)}|\Omega_{t-1})+GI_\pi^{(3)}(h,\nu_t=\delta,\Omega_{t-1})\leq F^{-1}(c) \text{ and } E(\pi_{t+h}^{(3)}|\Omega_{t-1})\leq F^{-1}(c) \\ GI_y^{(3)}(h,\nu_t=\delta,\Omega_{t-1})=0 \\ \quad , \text{ if } E(\pi_{t+h}^{(3)}|\Omega_{t-1})+GI_\pi^{(3)}(h,\nu_t=\delta,\Omega_{t-1})>F^{-1}(c) \text{ and } E(\pi_{t+h}^{(3)}|\Omega_{t-1})> F^{-1}(c) \end{cases} \end{cases} \tag{4.11}$$

Remark 5 *As previously for $GI_y^{(2)}(h,\nu_t=\delta,\Omega_{t-1})$, the value of $GI_y^{(3)}(h,\nu_t=\delta,\Omega_{t-1})$ is conditional on the value of $E(\pi_{t+h}^{(3)}|\Omega_{t-1})$. We can therefore rewrite equation (4.11) in two sub-cases*

$$\begin{cases} GI_\pi^{(3)}(h,\nu_t=\delta,\Omega_{t-1})=(\Lambda^h)_{(1,1)}.\nu_t \\ \begin{cases} GI_y^{(3)}(h,\nu_t=\delta,\Omega_{t-1})=1 \ , \ \textit{if } E(\pi_{t+h}^{(3)}|\Omega_{t-1})+GI_\pi^{(3)}(h,\nu_t=\delta,\Omega_{t-1})>F^{-1}(c) \\ GI_y^{(3)}(h,\nu_t=\delta,\Omega_{t-1})=0 \ , \ \textit{if } E(\pi_{t+h}^{(3)}|\Omega_{t-1})+GI_\pi^{(3)}(h,\nu_t=\delta,\Omega_{t-1})\leq F^{-1}(c) \end{cases} \end{cases} , \tag{4.12}$$

for $E(\pi_{t+h}^{(3)}|\Omega_{t-1})\leq F^{-1}(c)$.

$$\begin{cases} GI_\pi^{(3)}(h,\nu_t=\delta,\Omega_{t-1})=(\Lambda^h)_{(1,1)}.\nu_t \\ \begin{cases} GI_y^{(3)}(h,\nu_t=\delta,\Omega_{t-1})=-1 \ , \ \textit{if } E(\pi_{t+h}^{(3)}|\Omega_{t-1})+GI_\pi^{(3)}(h,\nu_t=\delta,\Omega_{t-1})\leq F^{-1}(c) \\ GI_y^{(3)}(h,\nu_t=\delta,\Omega_{t-1})=0 \ , \ \textit{if } E(\pi_{t+h}^{(3)}|\Omega_{t-1})+GI_\pi^{(3)}(h,\nu_t=\delta,\Omega_{t-1})> F^{-1}(c) \end{cases} \end{cases} , \tag{4.13}$$

for $E(\pi_{t+h}^{(3)}|\Omega_{t-1})> F^{-1}(c)$.

After proposing a formulation of the response function for models 2 and 3, it is important to note that a shock will be more persistent in the model that accounts for lagged latent variables (model 3). The autoregressive structure of this model has a much more progressive absorption of the shock, which will converge to 0 for a large enough horizon h. In Model 2, only the lagged binary variables affected by the shock have an impact on the $GI_\pi^{(2)}(.)$. We can speak of a threshold effect in this case. Indeed, as soon as the index of the time series that has been hit by by the shock and the index benchmark are both above or below the threshold c, the shock will tend to disappear.

After studying the response functions for specifications 2 and 3, we focus on the analysis of the fourth $GI^{(4)}(.)$ for $h > 0$. In such a case, the underlying dichotomous model studied corresponds to the general form and includes either the lagged values of the binary variable and the lagged value of the dependent variable.
In this model, the shock propagation mechanism is more complex. It relies on the interaction between the binary variable and the index. A shock that occurs in period t will directly impact the index, as described above. This shock can also be transmitted to the binary variable and cause a change in the latter. In the following period, the index may be impacted twice. The first impact will come from the potential nonlinear transmission of the shock to the dichotomous variable, as in Model 2. The second impact will come from the variation in the latent variable in the previous period, as in Model 3. Thus, a change in the binary variable for a given period has an impact on the p future index values. A variation in the index has an impact on q future index values. A shock may be more persistent in such a specification.

In an empirical application of such a model, proceeding by iteration makes it possible to understand the mechanism for the propagation of an exogenous shock. For this purpose, we must first measure the impact on the index for a given period. Then, we must calculate the binary variable associated with the time series hit by the shock and the benchmark. Then, we repeat this procedure during the next period until horizon h.

The exact form of the $GI^{(4)}(.)$ on the index is demonstrated by recurrence in Appendix 1. As with $GI^{(2)}(.)$ and $GI^{(3)}(.)$, the system of equations for this most general specification can be written conditionally on the value of $E(\pi_{t+h}^{(4)}|\Omega_{t-1})$ as follows:

$$\left\{ \begin{array}{l} GI_{\pi}^{(4)}(h,\nu_t=\delta,\Omega_{t-1}) = \sum_{i=0}^{h-1} GI_{\pi}^{(2)}(h-i,\nu_t=\delta,\Omega_{t-1})(\Lambda^i)_{(1,1)} + GI_{\pi}^{(3)}(h,\nu_t=\delta,\Omega_{t-1}) \\ \left\{ \begin{array}{l} GI_{y}^{(4)}(h,\nu_t=\delta,\Omega_{t-1}) = 1 \text{ , if } E(\pi_{t+h}^{(4)}|\Omega_{t-1}) + GI_{\pi}^{(4)}(h,\nu_t=\delta,\Omega_{t-1}) > F^{-1}(c) \\ GI_{y}^{(4)}(h,\nu_t=\delta,\Omega_{t-1}) = 0 \text{ , if } E(\pi_{t+h}^{(4)}|\Omega_{t-1}) + GI_{\pi}^{(4)}(h,\nu_t=\delta,\Omega_{t-1}) \leq F^{-1}(c) \end{array} \right. \end{array} \right. , \tag{4.14}$$

for $E(\pi_{t+h}^{(4)}|\Omega_{t-1}) \leq F^{-1}(c)$.

$$\begin{cases} GI_{\pi}^{(4)}(h, \nu_t = \delta, \Omega_{t-1}) = \sum_{i=0}^{h-1} GI_{\pi}^{(2)}(h-i, \nu_t = \delta, \Omega_{t-1})(\Lambda^i)_{(1,1)} + GI_{\pi}^{(3)}(h, \nu_t = \delta, \Omega_{t-1}) \\ \begin{cases} GI_{y}^{(4)}(h, \nu_t = \delta, \Omega_{t-1}) = -1 \text{ , if } E(\pi_{t+h}^{(4)}|\Omega_{t-1}) + GI_{\pi}^{(4)}(h, \nu_t = \delta, \Omega_{t-1}) \leq F^{-1}(c) \\ GI_{y}^{(4)}(h, \nu_t = \delta, \Omega_{t-1}) = 0 \text{ , if } E(\pi_{t+h}^{(4)}|\Omega_{t-1}) + GI_{\pi}^{(4)}(h, \nu_t = \delta, \Omega_{t-1}) > F^{-1}(c) \end{cases} \end{cases}, \tag{4.15}$$

for $E(\pi_{t+h}^{(4)}|\Omega_{t-1}) > F^{-1}(c)$.

Note that for the $GI_{\pi}^{(4)}(.)$ measured in the general dichotomous model case, for all $h \in \mathbb{Z}_+$, the equation resembles a combination of $GI_{\pi}^{(2)}(.)$ and $GI_{\pi}^{(3)}(.)$ when adding an extra term. Indeed, we can write equation (1R) as:

$$GI_{\pi}^{(4)}(h, \nu_t = \delta, \Omega_{t-1}) = GI_{\pi}^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) + \sum_{i=1}^{h-1} GI_{\pi}^{(2)}(h-i, \nu_t = \delta, \Omega_{t-1})(\Lambda^i)_{(1,1)} + GI_{\pi}^{(3)}(h, \nu_t = \delta, \Omega_{t-1}). \tag{4.16}$$

The left part of this equation entirely concerns the variation in the p last dichotomous variables. The right part relates to this index and is equal to the impulse response function of an AR(q), as we wrote it for *Model* 3. Finally, the additional term corresponds to the interaction that may exist between the index and the dichotomous variable.

4.4 Estimation procedure

In this section, we detail the procedure used to compute the bootstrapped confidence intervals for the GI of dichotomous models. This procedure can be followed for all the specifications defined above. A block-bootstrapped approach is used to generate these confidence intervals.

We introduce some notations for the following procedure. $\mathbf{Z}$ denotes a matrix of dimension $T \times (K+1)$, representing the sample studied, where T is the length of the time series and $K+1$ corresponds to the K explanatory variables plus the dependent variable. $\mathbf{Z}_t$ is a vector of dimension K, equal to the value of $\mathbf{Z}$ in period t. It can be represented as follows:

$$Z_t = \begin{bmatrix} Y_t & X_{1,t} & X_{2,t} & \cdots & X_{k-1,t} & X_{k,t} \end{bmatrix}$$

We also denote the size of a block by **b**. For a time series with T periods, we can generate $T-b+1$ overlapping blocks as follows:

$$\underbrace{Z_1,\cdots,Z_b}_{1\ block\ of\ length\ b}\ ;\ Z_2,\cdots,Z_{b+1}\ ;\ \cdots\ ;\ Z_{T-b+1},\cdots,Z_T$$

The computation of the bootstrapped confidence intervals for the GI of a dichotomous model consists of applying the following steps:

1. Estimate the size of the blocks to carry out a resampling. To determine **b**, follow the measurement proposed by Hall, Horowitz and Jing (1995):

$$b = T^{1/5}$$

2. Resample from $\mathbf{Z}$ a new sample, designated by $\mathbf{Z}^{(s)}$. To do this, replace the first **b** values of $\mathbf{Z}$ with the values of a block randomly drawn from the $T-b+1$ overlapping blocks.

3. Estimate one of the specifications from equation (4.2) by maximizing the log-likelihood function, using the new sample $\mathbf{Z}^{(s)}$.

$$\text{LogL}(\theta^{(s)}) = \sum_{t=1}^{T}\Big[y_t\log\big(F(\pi_t(\theta))\big) + (1-y_t)\log\big(1-F(\pi_t(\theta))\big)\Big]. \tag{4.17}$$

4. Initialize the threshold **c** in an arbitrary way (0.5 for example).[4] The non-linear transmission of a shock is conditional on crossing this threshold by $F(\pi_t)$.

5. Recover in a vector $\mathbf{V}_\varepsilon$ the set of residuals of the estimation of step 3. Then, calculate the shock δ by drawing the desired percentile ρ in the distribution $\mathbf{V}_\varepsilon$, where ρ can be set arbitrarily.

6. Then recursively for each horizon h from period 0 to period H, measure the GI as follows:

[4] **c** can also be calculated to take into account the type I and type II errors presented in Table 1. In the following section, for our empirical analysis, we use the threshold $\mathbf{c}_{AM}$ that is defined as the *Accuracy Measure* threshold in Candelon et, al. (2014). The formula of $\mathbf{c}_{AM}$ is presented in Appendix 2.

(a) Calculate the index π_{t+h} conditional on history available up to time $t-1$, $\Omega_{t-1}^{(s)}$, and the current shock ν of size δ. And the other index, the baseline, only conditional on history $\Omega_{t-1}^{(s)}$.

(b) From the two indexes calculated previously, calculate the GI as follows:

$$GI_{\pi}(h, \nu_t = \delta, \Omega_{t-1}) = E(\pi_{t+h}|\nu_t, \Omega_{t-1}^{(s)}) - E(\pi_{t+h}|\Omega_{t-1}^{(s)})$$

(c) Measure the two values of the binary variables that depends on both the value of the index and the threshold $\mathbf{c}$ as follows:

$$\begin{cases} E(y_{t+h}|E(\pi_{t+h}|\nu_t, \Omega_{t-1}^{(s)})) = 1 \text{ if } E(\pi_{t+h}|\nu_t, \Omega_{t-1}^{(s)}) > F^{-1}(c) \\ 0 \text{ otherwise} \end{cases}$$

$$\begin{cases} E(y_{t+h}|E(\pi_{t+h}|\Omega_{t-1}^{(s)})) = 1 \text{ if } E(\pi_{t+h}|\Omega_{t-1}^{(s)}) > F^{-1}(c) \\ 0 \text{ otherwise} \end{cases}$$

7. Repeat items 2-6 for $s = 1, ..., S$, where S is the total number of replications. Then, construct a 68% confidence interval around the GI by taking the 16th and 84th percentile of the distribution of the S replicas.

4.5 Empirical Analysis

Our aim in this section is to study the consequences of an exogenous shock for the probability of U.S.recessions. To evaluate the impact of such a shock, we proceed in two steps. First, we run four regressions to estimate each specification proposed by Kauppi and Saikkonen (2008). Next, using the previously obtained results, we measure the persistence of a shock for the four GIRFs as defined in Section 4.2, and we follow the procedure above to estimate the confidence intervals.

Following KS, we study the future recession using the predictive power of the yield curve. We propose to extend their study by evaluating the period of 1953:Q1-2020-Q1. As in previous studies, the recession variable, y_t, is obtained from the National Bureau of Economic Research (NBER). Regarding the explanatory variable, x_t, we take the interest rate spread as the difference between the ten-year Treasury bond rate and the three-month Treasury bill rate. We focus on the predictive power of each specification for forecast horizons from one quarter ahead. The following model is estimated:

$$Pr_{t-1}(y_t = 1) = F(\pi_t) = F(\alpha + \beta^{'} S_{t-1} + \gamma_1 y_{t-1} + \eta_1 \pi_{t-1} + \epsilon_t). \quad (4.18)$$

From Definition 2, four different specifications are estimated from equation (4.18). The first model is a static dichotomous model with $\gamma_1 = \eta_1 = 0$ (Model 1). For Model 1, the spread, S_t, is the only exogenous variable that affects the future occurrence of a crisis. The second and third models (Model 2 and Model 3, respectively) are both dynamic, with $\eta_1 = 0$ for Model 2 and $\gamma_1 = 0$ for Model 3. These two models are dynamic in that Model 2 includes the value of the previous state y_{t-1} and Model 3 includes the value of the lagged index π_{t-1}, that is, the value associated with the previous state. Finally, the last model combines the two previous models and includes both the lagged binary variable y_{t-1} and the lagged index π_{t-1} (Model 4). These specifications can be estimated by maximizing the log-likelihood function. This function has the following general form:

$$\text{LogL}(\theta) = \sum_{t=1}^{T} [y_t \text{log}(F(\pi_t(\theta))) + (1 - y_t)\text{log}(1 - F(\pi_t(\theta)))]. \quad (4.19)$$

Concerning the transformation function F, the logistic c.d.f. is preferred to a Gaussian c.d.f., as it is more appropriate for the study of extreme events such as crises (Kumar et al., (2003); van den Berg et al. (2008)). [5] The results we obtained are summarized in Table 4.2. These results are close to those obtained by Kauppi and Saikkonen (2008) with one lag. Indeed, the spread is always significant regardless of the model used, and the model that minimizes the AIC and BIC is Model 2.

From the results obtained in Table 4.2, we propose a GIRF analysis for each specification. We analyze the impact of an exogenous shock on both the index and dichotomous variable. This shock can be compared to an oil shock, such as those of 1973 and 1979, or to a health shock, such as the COVID-19 pandemic in 2020. If this shock is large enough, a nonlinear transmission can be observed, and the dichotomous variable can be affected. This nonlinear transmission is conditional on the index π_t crossing a threshold c. It can be observed when the shock occurs at time t but also at a more distant horizon h. Note that the persistence of the shock concerns only the three dynamic models. Thus, to measure the response functions, it is first necessary to calculate the value of the index at time t and

[5] The share of quarters in crisis is approximately 13.4% across our entire sample.

Table 4.2: Estimation results of the logit models - Quarterly frequency - 1953-2020

Model	Model 1	Model 2	Model 3	Model 4
$Intercept, \alpha$	-1.219***	-2.158	0.083	-2.502
	(0.322)	(0.355)	(0.101)	(0.576)
$Spread, S_{t-1}$	-0.524***	-1.075***	-0.323***	-1.111***
	(0.206)	(0.229)	(0.047)	(0.257)
$Recession, y_{t-1}$		4.846***		5.355***
		(0.482)		(0.611)
$Index, \pi_{t-1}$			0.851***	-0.125
			(0.004)	(0.078)
		Relevant Statistics		
AIC	202.209	109.404	164.088	109.800
BIC	211.384	123.165	177.850	128.149
$PseudoR^2$	0.033	0.468	0.218	0.471
$\#Observations$	268	268	268	268

Notes: This table reports the estimates obtained from the static and dynamic logit models (1) to (4) for the U.S. from 1953 to 2020 at a quarterly frequency with 1 lag. The dependent variable is the recession dummy. The results are computed using R 3.6.0 (R Core Team, 2020) and the *ews* (*v0.1.0*; Hasse and Lajaunie, 2020) package. The full reproducible code is available on CRAN. We report the Akaike (AIC) and Bayesian (BIC) information criterion for each specification. Standard errors are reported in parentheses below the estimates. Labels $***$, $**$ and $*$ indicate significance at the 99%, 95% and 90% levels, respectively.

define the threshold c. From Table 4.2 and from the value of the explanatory variables at time t, we measure the index of each model. Moreover, it is important to note that in the case in which the dichotomous variable is equal to 1, a positive shock can never lead to nonlinear transmission when it occurs at time t. This means that the initial state or the state that the economy is in at time t is important. Concerning the threshold, we estimate the c_{AM}, defined as the $Accuracy\ Measure$ in Candelon et al. (2014). This threshold arbitrates between type I and type II errors. Finally, with regard to the size of the shock, we initialize it from the residuals of our estimation as described in Section 4.4. The size of this shock is initialized at the 95th percentile of the residual distribution. Table 4.3 reports the value of the index π_t, the initial state y_t, the size of the shock δ and the threshold c_{AM} for each specification.

Table 4.3: Initial state and Index value - 2020:Q1

Model	Model 1	Model 2	Model 3	Model 4
$IndexValue, \pi_t$	-1.251	-2.222	-1.099	-2.298
$InitialState, y_t$	0	0	0	0
$Shock, \delta$	0.887	0.277	0.840	0.285
c_{AM}	0.112	0.147	0.177	0.132

Notes: This table reports the index value π_t, the initial state y_t, the shock δ, and the threshold c_{AM} for each specification.

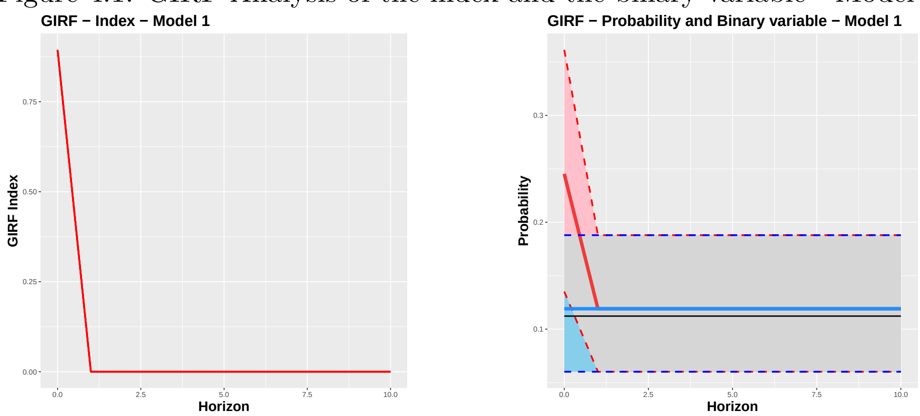
We follow the procedure described in Section 4.4, and we estimate a confidence interval for each specification. The results of our GIRF analysis are presented in the four tables and four figures below. The exogenous shock enters in an additive manner and is measured by a positive change in the conditional probability of a recession. Therefore, we consider the yield spread to be constant and equal to its value in t over the forecast horizon h.

Table 4.4: GIRF Analysis for Model 1

horizon	$GI_\pi(h, \nu_t = \delta, \Omega_{t-1})$	$P(E(\pi_{t+h}\|\delta, \Omega_{t-1}))$	$P(\pi_{t+h}\|\Omega_{t-1})$	c_{AM}	$E(Y_{t+h}\|\delta)$	$E(Y_{t+h})$
0	0.894 (0.892 / 0.894)	0.248 (0.129 / 0.399)	0.121 (0.057 / 0.213)	0.112	1	1
1	0.000 (0.000 / 0.000)	0.121 (0.057 / 0.213)	0.121 (0.057 / 0.213)	0.112	1	1
2	0.000 (0.000 / 0.000)	0.121 (0.057 / 0.213)	0.121 (0.057 / 0.213)	0.112	1	1
3	0.000 (0.000 / 0.000)	0.121 (0.057 / 0.213)	0.121 (0.057 / 0.213)	0.112	1	1
4	0.000 (0.000 / 0.000)	0.121 (0.057 / 0.213)	0.121 (0.057 / 0.213)	0.112	1	1
5	0.000 (0.000 / 0.000)	0.121 (0.057 / 0.213)	0.121 (0.057 / 0.213)	0.112	1	1
10	0.000 (0.000 / 0.000)	0.121 (0.057 / 0.213)	0.121 (0.057 / 0.213)	0.112	1	1
20	0.000 (0.000 / 0.000)	0.121 (0.057 / 0.213)	0.121 (0.057 / 0.213)	0.112	1	1
30	0.000 (0.000 / 0.000)	0.121 (0.057 / 0.213)	0.121 (0.057 / 0.213)	0.112	1	1

Notes: This table reports the value of the generalized impulse response function of the index for Model 1 in column 2, with the confidence intervals in brackets. The third column corresponds to the probability of observing a crisis conditional on a shock of size δ and all the information available at time $t-1$ (confidence intervals in brackets). Then, the probability of observing a crisis for the benchmark is in column 4 (confidence intervals in brackets). The threshold c_AM is in column 5. The value of the binary variables for both series hit and not hit by a shock are in the two last columns.

Figure 4.1: GIRF Analysis of the index and the binary variable - Model 1



Notes: The figure on the left plots the impulse response function associated with the index of Model 2. The figure on the right plots the impulse response function associated with the probability and the binary variable of Model 2. The shock is equal to the 95th percentile of the residual distribution. The confidence interval is estimated with 10,000 replications of the block-bootstrap procedure presented in Section 4. The red and blue shaded areas correspond to the 68% confidence intervals of the Moving-Block Bootstrap for the forecast of the probability with and without shock. The black line corresponds to the threshold c_{AM}.

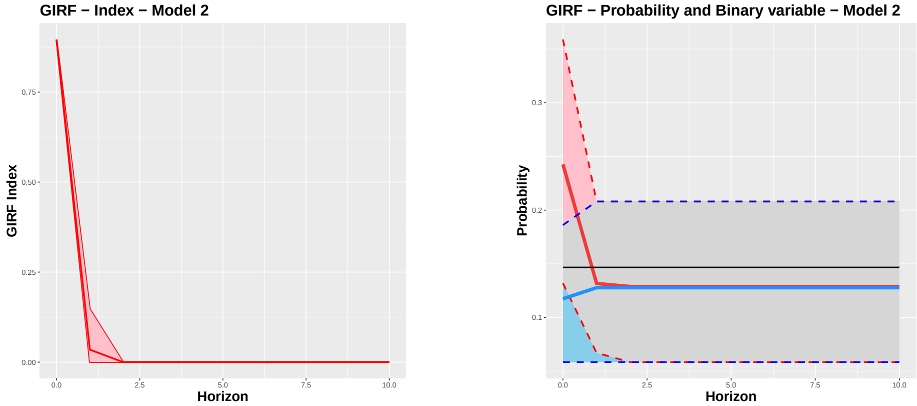
In $Model$ 1, the shock is not persistent. The value of the index response function is zero from period 2. This is explained by the static form of the model. The figure on the right illustrates that the difference between the conditional probability of the time series that has been hit by a shock of size δ in period t and the other conditional probability of the time series that has not been hit is also zero from period 2. The probability level remains in both cases above the threshold c_{AM}. For both time series, the binary variable is equal to 1 from period t, and a positive shock has no impact on its value. However, note that a part of the confidence interval of the time series that has not been hit by the shock is below this threshold, meaning that in some of the replications, the response function of the dichotomous variable is equal to 1 at time t, but convergence is achieved from the following period.

Table 4.5: GIRF Analysis for Model 2

horizon	$GI_\pi(h, \nu_t = \delta, \Omega_{t-1})$	$P(E(\pi_{t+h}\|\delta, \Omega_{t-1}))$	$P(\pi_{t+h}\|\Omega_{t-1})$	c_{AM}	$E(Y_{t+h}\|\delta)$	$E(Y_{t+h})$
0	0.896 (0.894 / 0.896)	0.243 (0.132 / 0.359)	0.117 (0.058 / 0.186)	0.147	1	0
1	0.035 (0.000 / 0.147)	0.132 (0.067 / 0.208)	0.128 (0.058 / 0.208)	0.147	0	0
2	0.000 (0.000 / 0.000)	0.129 (0.058 / 0.208)	0.128 (0.058 / 0.208)	0.147	0	0
3	0.000 (0.000 / 0.000)	0.129 (0.058 / 0.208)	0.128 (0.058 / 0.208)	0.147	0	0
4	0.000 (0.000 / 0.000)	0.129 (0.058 / 0.208)	0.128 (0.058 / 0.208)	0.147	0	0
5	0.000 (0.000 / 0.000)	0.129 (0.058 / 0.208)	0.128 (0.058 / 0.208)	0.147	0	0
10	0.000 (0.000 / 0.000)	0.129 (0.058 / 0.208)	0.129 (0.058 / 0.208)	0.147	0	0
20	0.000 (0.000 / 0.000)	0.129 (0.058 / 0.208)	0.129 (0.058 / 0.208)	0.147	0	0
30	0.000 (0.000 / 0.000)	0.129 (0.058 / 0.208)	0.129 (0.058 / 0.208)	0.147	0	0

Notes: This table reports the value of the generalized impulse response function of the index for Model 2 in column 2, with the confidence intervals in brackets. The third column corresponds to the probability of observing a crisis conditional on a shock of size δ and all the information available at time $t-1$ (confidence intervals in brackets). Then, the probability of observing a crisis for the benchmark is in column 4 (confidence intervals in brackets). The threshold c_AM is in column 5. The value of the binary variables for both series hit and not hit by a shock are in the two last columns.

Figure 4.2: GIRF Analysis of the index and the binary variable - Model 2



Notes: The figure on the left plots the impulse response function associated with the index of Model 2. The figure on the right plots the impulse response function associated with the probability and the binary variable of Model 2. The shock is equal to the 95th percentile of the residual distribution. The confidence interval is estimated with 10,000 replications of the block-bootstrap procedure presented in Section 4. The red and blue shaded areas correspond to the 68% confidence intervals of the Moving-Block Bootstrap for the forecast of the probability with and without shock. The black line corresponds to the threshold c_{AM}.

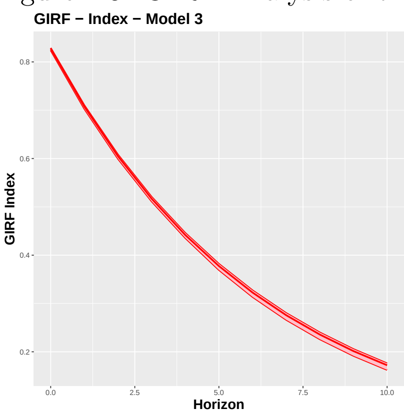
The dynamics of $Model$ 2 seem to lengthen the persistence of the shock compared to $Model$ 1. This is due to the use of the lagged dichotomous variable as one of the explanatory variables. Indeed, the index only fully converges from period $t+2$. The conditional probability calculated from the index of this model differs slightly from the conditional probability calculated in the previous model. Indeed, the time series that has not been hit by the shock remains below the threshold c_{AM}, unlike the time series that has been hit. Therefore, a nonlinear transmission of the shock is observed at time t. Note also that $Model2$ seems to be the model chosen by the AIC and BIC. Therefore, this scenario would be the most conceivable. Thus, if the United States suffered an exogenous shock in the second quarter of 2020, the persistence of this shock over the next quarter could be expected.

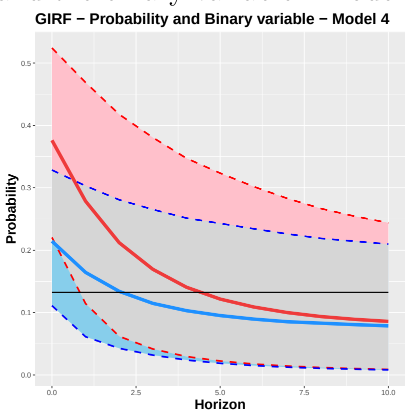
Table 4.6: GIRF Analysis for Model 3

horizon	$GI_{\pi}(h, \nu_t = \delta, \Omega_{t-1})$	$P(E(\pi_{t+h}\|\delta, \Omega_{t-1}))$	$P(\pi_{t+h}\|\Omega_{t-1})$	c_{AM}	$E(Y_{t+h}\|\delta)$	$E(Y_{t+h})$
0	0.828 (0.824 / 0.829)	0.346 (0.188 / 0.504)	0.192 (0.092 / 0.307)	0.177	1	1
1	0.708 (0.703 / 0.710)	0.287 (0.110 / 0.495)	0.171 (0.057 / 0.325)	0.177	1	0
2	0.605 (0.600 / 0.608)	0.246 (0.067 / 0.485)	0.157 (0.038 / 0.339)	0.177	1	0
3	0.517 (0.512 / 0.521)	0.217 (0.042 / 0.476)	0.148 (0.025 / 0.352)	0.177	1	0
4	0.442 (0.436 / 0.446)	0.196 (0.028 / 0.474)	0.141 (0.018 / 0.366)	0.177	1	0
5	0.378 (0.370 / 0.382)	0.180 (0.020 / 0.469)	0.136 (0.014 / 0.376)	0.177	1	0
10	0.172 (0.163 / 0.176)	0.141 (0.007 / 0.454)	0.124 (0.006 / 0.410)	0.177	0	0
20	0.036 (0.032 / 0.037)	0.122 (0.003 / 0.448)	0.119 (0.003 / 0.439)	0.177	0	0
30	0.007 (0.006 / 0.008)	0.119 (0.003 / 0.446)	0.118 (0.003 / 0.444)	0.177	0	0

Notes: This table reports the value of the generalized impulse response function of the index for Model 3 in column 2, with the confidence intervals in brackets. The third column corresponds to the probability of observing a crisis conditional on a shock of size δ and all the information available at time $t-1$ (confidence intervals in brackets). Then, the probability of observing a crisis for the benchmark is in column 4 (confidence intervals in brackets). The threshold $c_A M$ is in column 5. The value of the binary variables for both series hit and not hit by a shock are in the two last columns.

Figure 4.3: GIRF Analysis of the index and the binary variable - Model 3





Notes: The figure on the left plots the impulse response function associated with the index of Model 3. The figure on the right plots the impulse response function associated with the probability and the binary variable of Model 3. The shock is equal to the 95th percentile of the residual distribution. The confidence interval is estimated with 10,000 replications of the block-bootstrap procedure presented in Section 4. The red and blue shaded areas correspond to the 68% confidence intervals of the Moving-Block Bootstrap for the forecast of the probability with and without shock. The black line corresponds to the threshold c_{AM}.

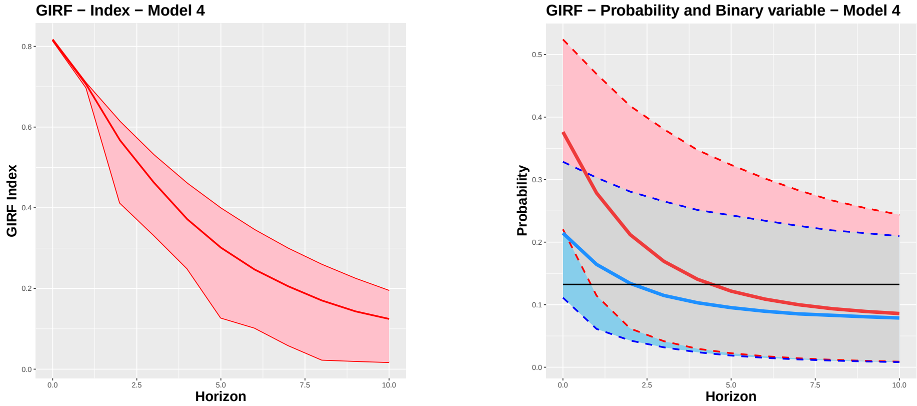
The shock is much more persistent in $Model$ 3. The index process is comparable to an AR (1). In Table 4.2, the result of our estimation indicates that the coefficient of the lagged index is equal to 0.851. This explains why convergence takes place more slowly. Note that the effect of the shock on the index is less than 0.05 from period $t + 18$. Concerning the conditional probability, the time series that has been hit by the shock at time t remains above the threshold c_{AM} until period $t + 5$, while the time series that has not been hit by the shock remains above the threshold only in period t. This means that an exogenous shock would increase the duration of the crisis, which was already forecast by this specification.

Table 4.7: GIRF Analysis for Model 4

horizon	$GI_{\pi}(h, \nu_t = \delta, \Omega_{t-1})$	$P(E(\pi_{t+h}\vert\delta, \Omega_{t-1)}))$	$P(\pi_{t+h}\vert\Omega_{t-1)})$	c_{AM}	$E(Y_{t+h}\vert\delta)$	$E(Y_{t+h})$
0	0.816 (0.815 / 0.817)	0.376 (0.220 / 0.524)	0.214 (0.111 / 0.329)	0.132	1	1
1	0.705 (0.698 / 0.708)	0.279 (0.114 / 0.469)	0.164 (0.061 / 0.303)	0.132	1	1
2	0.567 (0.412 / 0.613)	0.212 (0.062 / 0.418)	0.134 (0.043 / 0.281)	0.132	1	1
3	0.463 (0.333 / 0.531)	0.169 (0.042 / 0.381)	0.115 (0.032 / 0.265)	0.132	1	0
4	0.372 (0.249 / 0.460)	0.141 (0.030 / 0.347)	0.103 (0.024 / 0.251)	0.132	1	0
5	0.301 (0.128 / 0.398)	0.122 (0.022 / 0.324)	0.095 (0.019 / 0.243)	0.132	0	0
10	0.125 (0.017 / 0.194)	0.086 (0.009 / 0.244)	0.079 (0.008 / 0.210)	0.132	0	0
20	0.028 (0.003 / 0.047)	0.072 (0.005 / 0.194)	0.071 (0.004 / 0.187)	0.132	0	0
30	0.007 (0.001 / 0.011)	0.070 (0.004 / 0.183)	0.070 (0.004 / 0.182)	0.132	0	0

Notes: This table reports the value of the generalized impulse response function of the index for Model 4 in column 2, with the confidence intervals in brackets. The third column corresponds to the probability of observing a crisis conditional on a shock of size δ and all the information available at time $t - 1$ (confidence intervals in brackets). Then, the probability of observing a crisis for the benchmark is in column 4 (confidence intervals in brackets). The threshold c_AM is in column 5. The value of the binary variables for both series hit and not hit by a shock are in the two last columns.

Figure 4.4: GIRF Analysis of the index and the binary variable - Model 4



Notes: The figure on the left plots the impulse response function associated with the index of Model 4. The figure on the right plots the impulse response function associated with the probability and the binary variable of Model 4. The shock is equal to the 95th percentile of the residual distribution. The confidence interval is estimated with 10,000 replications of the block-bootstrap procedure presented in Section 4. The red and blue shaded areas correspond to the 68% confidence intervals of the Moving-Block Bootstrap for the forecast of the probability with and without shock. The black line corresponds to the threshold c_{AM}.

The last specification proposed by Kauppi and Saikkonen (2008) is much more dynamic. Indeed, both the lagged index and the lagged binary variable are added as explanatory variables. Note that the AIC and BIC are close to those of $Model$ 2. Therefore, the scenario proposed by this model remains to be considered. Concerning the index, the shock is slightly less persistent than it was for $Model$ 3. Nevertheless, it converges faster towards 0. The value of the response function for the index is less than 0.05 from period $t + 15$. The conditional probability of the time series that has been hit by the shock at time t and the conditional probability of the time series that has not been hit are above the threshold c_{AM} during the two quarters following the shock period. However, the shock led to a prolongation of the recession by 2 quarters and finally ended in $t + 4$, which is 1 year after the occurrence of the shock. For the United States, such a scenario would indicate that this exogenous shock would lead to a recession until the second quarter of 2021.

4.6 Conclusion

This paper introduces the generalized impulse response function (GI) for dichotomous models in a univariate framework. The definition of Koop Pesaran and Potter (1996) (KPP) is extended to a system of two impulse response functions to be adapted for such a model. The system of two equations allows us to study the consequences of an exogenous shock for both the latent variable and the binary variable. Then, the exact form of the response functions for each specification of Kauppi and Saikkonen (2008) (KS) is developed. Finally, a block-bootstrap method is also proposed to estimate the confidence intervals of this analysis. This provides a new framework for estimating the propagation of shocks in binary models.

An empirical illustration of this new methodology is applied to study the persistence and propagation mechanisms of an exogenous shock to the U.S. economy. As KPP suggest, the shocks are drawn from the historical distribution. I discuss the consequences of historical shocks on the four specifications of KS and their effect on the U.S. business cycle. The results differ from one specification to another. We focus on the results of the two best models regarding the Akaike information criterion (AIC) and the Schwarz Bayesian criterion (BIC) . Model 2, which includes a lagged value of the binary variable, indicates that an exogenous shock could be transmitted in a nonlinear fashion to the binary variable. Thus, it shows that the onset of a recession could occur in the second quarter of 2020. The shock does not seem to be persistent. For Model 4, the results indicate that a shock would be much more persistent. This specification includes both a lagged binary variable and a lagged index. An exogenous shock could lead to a recession lasting until the second quarter of 2021.

References

Beaudry, P. and G. Koop (1993). "Do recessions permanently change output?" In: *Journal of Monetary economics* 31.2, pp. 149–163.

Campbell, J. Y. and N. G. Mankiw (1987). "Are output fluctuations transitory?" In: *The Quarterly Journal of Economics* 102.4, pp. 857–880.

Candelon, B., E.-I. Dumitrescu, and C. Hurlin (2012). "How to evaluate an early-warning system: Toward a unified statistical framework for assessing financial crises forecasting methods". In: *IMF Economic Review* 60.1, pp. 75–113.

— (2014). "Currency crisis early warning systems: Why they should be dynamic". In: *International Journal of Forecasting* 30.4, pp. 1016–1029.

Candelon, B., E.-I. Dumitrescu, C. Hurlin, and F. C. Palm (2013). "Multivariate dynamic probit models: an application to financial crises mutation". In: *VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims*. Emerald Group Publishing Limited.

Chauvet, M. and S. Potter (2002). "Predicting a recession: evidence from the yield curve in the presence of structural breaks". In: *Economics Letters* 77.2, pp. 245–253.

— (2005). "Forecasting recessions using the yield curve". In: *Journal of forecasting* 24.2, pp. 77–103.

Diebold, F. X. and G. D. Rudebusch (1989). "Long memory and persistence in aggregate output". In: *Journal of monetary economics* 24.2, pp. 189–209.

Doan, T., R. Litterman, and C. Sims (1984). "Forecasting and conditional projection using realistic prior distributions". In: *Econometric reviews* 3.1, pp. 1–100.

Dotsey, M. (1998). "The predictive content of the interest rate term spread for future economic growth". In: *FRB Richmond Economic Quarterly* 84.3, pp. 31–51.

Dueker, M. (2005). "Dynamic forecasts of qualitative variables: a Qual VAR model of US recessions". In: *Journal of Business & Economic Statistics* 23.1, pp. 96–104.

Estrella, A. and G. A. Hardouvelis (1991). "The term structure as a predictor of real economic activity". In: *The journal of Finance* 46.2, pp. 555–576.

Estrella, A., A. P. Rodrigues, and S. Schich (2003). "How stable is the predictive power of the yield curve? Evidence from Germany and the United States". In: *Review of Economics and Statistics* 85.3, pp. 629–644.

Gallant, A. R., P. E. Rossi, and G. Tauchen (1993). “Nonlinear dynamic structures”. In: *Econometrica: Journal of the Econometric Society*, pp. 871–907.

Hafner, C. M. and H. Herwartz (2006). “Volatility impulse responses for multivariate GARCH models: An exchange rate illustration”. In: *Journal of International Money and Finance* 25.5, pp. 719–740.

Hall, P., J. L. Horowitz, and B.-Y. Jing (1995). “On blocking rules for the bootstrap with dependent data”. In: *Biometrika* 82.3, pp. 561–574.

Hasse, J.-B. and Q. Lajaunie (2020a). *Does the Yield Curve Signal Recessions? New Evidence from an International Panel Data Analysis*. AMSE Working Paper.

— (2020b). “EWS: Early Warning System”. In: *R package version 0.1. 0*.

Kaminsky, G., S. Lizondo, and C. M. Reinhart (1998). “Leading indicators of currency crises”. In: *Staff Papers* 45.1, pp. 1–48.

Kauppi, H. and P. Saikkonen (2008). “Predicting US recessions with dynamic binary response models”. In: *The Review of Economics and Statistics* 90.4, pp. 777–791.

Koop, G., M. H. Pesaran, and S. M. Potter (1996). “Impulse response analysis in nonlinear multivariate models”. In: *Journal of econometrics* 74.1, pp. 119–147.

Kumar, M., U. Moorthy, and W. Perraudin (2003). “Predicting emerging market currency crashes”. In: *Journal of Empirical Finance* 10.4, pp. 427–454.

Moneta, F. (2005). “Does the yield spread predict recessions in the Euro area?” In: *International Finance* 8.2, pp. 263–301.

Naceur, S. B., B. Candelon, and Q. Lajaunie (2019). “Taming financial development to reduce crises”. In: *Emerging Markets Review* 40, p. 100618.

Potter, S. M. (2000). “Nonlinear impulse response functions”. In: *Journal of Economic Dynamics and Control* 24.10, pp. 1425–1446.

Rosenberg, J. V. and S. Maurer (2008). “Signal or noise? Implications of the term premium for recession forecasting”. In: *Federal Reserve Bank of New York-Economic Policy Review* 14.1.

Sensier, M. et al. (2004). “Domestic and international influences on business cycle regimes in Europe”. In: *International Journal of Forecasting* 20.2, pp. 343–357.

Sims, C. A. (1980). “Macroeconomics and reality”. In: *Econometrica: journal of the Econometric Society*, pp. 1–48.

Van den Berg, J., B. Candelon, and J.-P. Urbain (2008). “A cautious note on the use of panel models to predict financial crises”. In: *Economics Letters* 101.1, pp. 80–83.

Appendices

Appendix 1: Proof for the IRF

Proof:

Proof: We will prove by induction that for all $h \in \mathbb{Z}_+$ and for $\pi_t = \alpha + \beta' X_{t-1} + \sum_{i=1}^{p} \gamma_i y_{t-i} + \sum_{j=1}^{q} \eta_j \pi_{t-j} + \epsilon_t$,

$$GI_\pi^{(4)}(h, \nu_t = \delta, \Omega_{t-1}) = \sum_{i=0}^{h-1} GI_\pi^{(2)}(h-i, \nu_t = \delta, \Omega_{t-1})(\Lambda^i)_{(1,1)} + GI_\pi^{(3)}(h, \nu_t = \delta, \Omega_{t-1}), \tag{4.A.1}$$

where,

$$GI_\pi^{(2)}(h, \nu_t = \delta, \Omega_{t-1}) = \sum_{j=1}^{p} \gamma_j \Big(E\big(y_{t+h-j} | E(\pi_{t+h} | \nu_t, \Omega_{t-1}), c\big) - E\big(y_{t+h-j} | E(\pi_{t+h} | \Omega_{t-1}), c\big) \Big),$$

$$GI_\pi^{(3)}(h, \nu_t = \delta, \Omega_{t-1}) = (\Lambda^h)_{(1,1)}.\nu_t,$$

and

$$\Lambda = \begin{bmatrix} \eta_1 & \eta_2 & \cdots & \eta_{q-1} & \eta_q \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}.$$

Base Case: From the equation (4.3), the IRF of the index for a given horizon can be written as the difference between the latent variable with a shock and the latent variable without any shock. So, when $h = 1$, we have:

$$\begin{aligned}
GI_{\pi}^{(4)}(1,\nu_t=\delta,\Omega_{t-1}) &= E(\pi_{t+1}|\nu_t,\Omega_{t-1}) - E(\pi_{t+1}|\Omega_{t-1}) \\
&= E(\alpha+\beta^{'}X_t+\sum_{i=1}^{p}\gamma_i E(y_{t+1-i}|E(\pi_{t+1-i}|\nu_t,\Omega_{t-1}),c)+\sum_{j=1}^{q}\eta_j E(\pi_{t+1-j}|\nu_t,\Omega_{t-1})) \\
&- E(\alpha+\beta^{'}X_t+\sum_{i=1}^{p}\gamma_i E(y_{t+1-i}|E(\pi_{t+1-i}|\Omega_{t-1}),c)+\sum_{j=1}^{q}\eta_j E(\pi_{t+1-j}|\Omega_{t-1})) \\
&= \sum_{i=1}^{p}\gamma_i[E(y_{t+1-i}|E(\pi_{t+1-i}|\nu_t,\Omega_{t-1}),c) - E(y_{t+1-i}|E(\pi_{t+1-i}|\Omega_{t-1}),c)] \\
&+ \sum_{j=1}^{q}\eta_j[E(\pi_{t+1-j}|\nu_t,\Omega_{t-1}) - E(\pi_{t+1-j}|\Omega_{t-1})]
\end{aligned}$$

Because the shock occurs at period t, $E(\pi_{t+1-j}|\nu_t,\Omega_{t-1}) = E(\pi_{t+1-j}|\Omega_{t-1})\ \forall\ j > h$. So, we get:

$$\begin{aligned}
GI_{\pi}^{(4)}(0,V_t=\delta,\Omega_{t-1}) &= \gamma_1[E(y_t|E(\pi_t|\nu_t,\Omega_{t-1}),c) - E(y_t|E(\pi_t|\Omega_{t-1}),c)] \\
&+ \eta_1[E(\pi_t|\nu_t,\Omega_{t-1}) - E(\pi_t|\Omega_{t-1})] \\
&= GI_{\pi}^{(2)}(0,\nu_t=\delta,\Omega_{t-1}) + \eta_1.\nu_t \\
&= GI_{\pi}^{(2)}(0,\nu_t=\delta,\Omega_{t-1}) + GI_{\pi}^{(3)}(0,\nu_t=\delta,\Omega_{t-1})(\Lambda^0)_{(1,1)}
\end{aligned}$$

So, (4.A.1) is true for $h = 1$.

Induction Step: Let $n \in \mathbb{Z}_+$ be given and suppose (4.A.1) is true for $h = n$. Then

$$\begin{aligned}
GI_{\pi}^{(4)}(n+1,\nu_t=\delta,\Omega_{t-1}) &= E(\pi_{t+n+1}|\nu_t,\Omega_{t-1}) - E(\pi_{t+n+1}|\Omega_{t-1}) \\
&= \sum_{i=1}^{p}\gamma_i[E(y_{t+n+1-i}|E(\pi_{t+n+1-i}|\nu_t,\Omega_{t-1}),c) - E(y_{t+n+1-i}|E(\pi_{t+n+1-i}|\Omega_{t-1}),c)] \\
&+ \sum_{j=1}^{q}\eta_j[E(\pi_{t+n+1-j}|\nu_t,\Omega_{t-1}) - E(\pi_{t+n+1-j}|\Omega_{t-1})] \\
&= GI_{\pi}^{(2)}(n+1,\nu_t=\delta,\Omega_{t-1}) + \sum_{j=1}^{q}\eta_j GI_{\pi}^{(4)}(n+1-j,\nu_t=\delta,\Omega_{t-1}) \\
&= GI_{\pi}^{(2)}(n+1,\nu_t=\delta,\Omega_{t-1}) \\
&+ \sum_{j=1}^{q}\eta_j(\sum_{i=0}^{n-1} GI_{\pi}^{(2)}(n-i,\nu_t=\delta,\Omega_{t-1})(\Lambda^i)_{(1,1)} + GI_{\pi}^{(3)}(n,\nu_t=\delta,\Omega_{t-1}))
\end{aligned}$$

(by induction hypothesis).

Note that,

$$\sum_{j=1}^{q} \eta_j GI_{\pi}^{(2)}(n, \nu_t = \delta, \Omega_{t-1}) = GI_{\pi}^{(2)}(n+1, \nu_t = \delta, \Omega_{t-1}) = (\Lambda^{n+1})_{(1,1)}.\nu_t,$$

by inference ,

$$\sum_{j=1}^{q} \eta_j (\Lambda^n)_{(1,1)}.\nu_t = (\Lambda^{n+1})_{(1,1)}.\nu_t.$$

We can deduce that,

$$\begin{aligned}\sum_{j=1}^{q} \eta_j (\sum_{i=0}^{n-1} GI_{\pi}^{(2)}(n-i, \nu_t = \delta, \Omega_{t-1})(\Lambda^i)_{(1,1)}) &= \sum_{i=0}^{n-1} GI_{\pi}^{(2)}(n-i, \nu_t = \delta, \Omega_{t-1})(\Lambda^{i+1})_{(1,1)} \\ &= \sum_{i=1}^{n} GI_{\pi}^{(2)}(n+1-i, \nu_t = \delta, \Omega_{t-1})(\Lambda^i)_{(1,1)}.\end{aligned}$$

So,

$$\begin{aligned}GI_{\pi}^{(4)}(n+1, V_t = \delta, \Omega_{t-1}) &= GI_{\pi}^{(2)}(n+1, \nu_t = \delta, \Omega_{t-1}) \\ &+ \sum_{i=1}^{n} GI_{\pi}^{(2)}(n+1-i, \nu_t = \delta, \Omega_{t-1})(\Lambda^i)_{(1,1)} + GI_{\pi}^{(3)}(n+1, \nu_t = \delta, \Omega_{t-1}) \\ &= \sum_{i=0}^{n} GI_{\pi}^{(2)}(n+1-i, \nu_t = \delta, \Omega_{t-1})(\Lambda^i)_{(1,1)} + GI_{\pi}^{(3)}(n+1, \nu_t = \delta, \Omega_{t-1}).\end{aligned}$$

Thus, (4.A.1) holds for $h = n+1$, and the proof of the induction step is complete.

Conclusion: By the principle of induction, (4.A.1) is true for all $h \in \mathbb{Z}_+$.

Appendix 2: Accuracy Measure threshold

$\mathbf{c}_{AM}$ is defined as the *Accuracy Measure* threshold in Candelon et al,. (2014). It takes into account both type I error and type II error. It is calculated as follows:

$$c_{AM} = \arg\max_{c\in[0;1]} \left[\frac{\sum_{t=1}^{T} \mathbb{1}_{(\hat{y}_t(c)=1)} \times \mathbb{1}_{(y_t(c)=1)}}{\sum_{t=1}^{T} \mathbb{1}_{(y_t(c)=1)}} + \frac{\sum_{t=1}^{T} \mathbb{1}_{(\hat{y}_t(c)=0)} \times \mathbb{1}_{(y_t(c)=0)}}{\sum_{t=1}^{T} \mathbb{1}_{(y_t(c)=0)}} - 1 \right],$$

where:

$\mathbb{1}_{(\hat{y}_t(c)=1)} \times \mathbb{1}_{(y_t(c)=1)}$ *takes value 1 if* $F(\pi_t) \geq c$ *and* $y_t = 1$ *and* $\mathbb{1}_{(y_t(c)=0)}$ *takes value 1 if* $y_t = 1$*,*

$\mathbb{1}_{(\hat{y}_t(c)=0)} \times \mathbb{1}_{(y_t(c)=0)}$ *takes value 1 if* $F(\pi_t) < c$ *and* $y_t = 0$ *and* $\mathbb{1}_{(y_t(c)=0)}$ *takes value 1 if* $y_t = 0$*.*

General conclusion

The four chapters of this thesis are original contributions with the common objective of studying questions of finance or macroeconomics using the econometric tools of nonlinear models.

In the first chapter of this thesis, we sought to show that the investment decisions of mutual funds classified as ethical are not always in accordance with the principles announced by their managers. We also empirically tested the impact of environmental, social and governance criteria on the financial performance of these funds. To do so, our study is based on the CAPM (Sharpe, 1964) and its extensions to 3 and 4 factors (Fama and French, 1993; Carhart, 1997) estimated using Hansen's (2000) nonlinear panel approach. This econometric specification constructed from a dichotomous variable makes it possible to study the potential difference in performance between ethical funds and conventional funds while taking into account the ESG scores associated with the investments made by these funds. Our results show that there are no significant differences in terms of extra-financial performance between conventional funds and funds classified as ethical by their managers. On the other hand, our results indicate that the impact of extrafinancial criteria on financial performance is negative. These results are consistent with the theoretical arguments of Bollen (2007) and Fama and French (2007): extrafinancial constraints reduce the investment universe and therefore can only have a negative effect on the risk/return ratio of a portfolio.

In the second chapter, we examine the relationship between financial development and banking crises. We use the dichotomous panel of Candelon, Dumitrescu and Hurlin (2014) on data from approximately 100 countries. The correction a la Carro (2007) is also implemented to correct possible biases related to the fixed effects of the panel approach. We use the financial development indicators extended by Svirydzenka (2016) decomposed into six sub-indices, allowing us to determine in detail the factors related to banking crises. Our results reveal heterogeneity in the relationship between the different subindices of financial development and banking crises across developed, emerging and low-income countries.

In the third chapter, we study the predictive power of the yield spread on business cycles for a panel of OECD countries. In this framework, the dichotomous panel of Candelon, Dumitrescu and Hurlin (2014) is also used. On the other hand, a cluster methodology is developed from the work of Zhang, Wang and Zhu (2019) to test the homogeneity of the relationship between the yield spread, monetary policy and the business cycle. The panel results confirm the predictive power of the yield spread over business cycles while being robust to different macroeconomic control variables. Concerning clusters, two main groups stand out, confirming the partial homogeneity of the panel. However, the coefficient of the central banks' key interest rate does not appear to be significant, regardless of the cluster. These results can be explained in particular by the unconventional monetary policy instruments used in recent years, as argued by Chinn and Kucko (2015).

Finally, in the last chapter, we determine the exact shape of the response function adapted to the four specifications of Kauppi and Saikkonen (2008). This formalization is based on the definition of the generalized impulse-response function of Koop, Pesaran and Potter (1996). This response function is then applied to study the impact of an exogenous shock on cycles in the United States. In this empirical part, we first study the relationship between the yield spread and cycles over the period 1953-2020. Then, based on these estimates, we analyze the impact of the shock on future cycles. Particular attention is paid to the two dynamic models that include the lagged binary variable. While the specification that takes into account only the lagged binary variable predicts a recession in the United States over one quarter, the specification that also uses the lagged underlying index predicts a recession over the next five quarters. This difference is explained in particular by the autoregressive structure of the latter.

These various contributions show that dichotomous models are useful for better understanding and forecasting business cycles and financial risks. The contribution of nonlinear econometrics is significant in both macroeconomics and finance. The extensions of my work that I consider relevant would be to study the asymptotic properties of the IRF to propose an exact form of the confidence interval. An extension of the IRF to the multivariate framework defined by Candelon, Dumitrescu, Hurlin and Palm (2013) could also be the subject of future research. On the other hand, Berg, Candelon and Urbain (2008) highlight the need to verify the possibility of processing panel data through geographic groupings. This point would need to be studied as a follow-up to our study of the predictive power of the yield spread on business cycles for the 13 OECD countries. Finally, Berg, Candelon and Urbain (2008) address the notion of spillover effects as a determinant in the transmission of financial crises. Constructing a control variable for each country that would take into account the proportion of countries in crisis within the panel and then adding it to our model could also be an avenue for development.

Bibliography

Bollen, N. P. (2007). "Mutual fund attributes and investor behavior". In: *Journal of Financial and Quantitative Analysis* 42.3, pp. 683–708.

Candelon, B., E.-I. Dumitrescu, and C. Hurlin (2014). "Currency crisis early warning systems: Why they should be dynamic". In: *International Journal of Forecasting* 30.4, pp. 1016–1029.

Candelon, B., E.-I. Dumitrescu, C. Hurlin, and F. C. Palm (2013). *Multivariate Dynamic Probit Models: An Application to Financial Crises Mutation', VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims (Advances in Econometrics, Volume 32).*

Carhart, M. M. (1997). "On persistence in mutual fund performance". In: *The Journal of finance* 52.1, pp. 57–82.

Carro, J. M. (2007). "Estimating dynamic panel data discrete choice models with fixed effects". In: *Journal of Econometrics* 140.2, pp. 503–528.

Chinn, M. and K. Kucko (2015). "The predictive power of the yield curve across countries and time". In: *International Finance* 18.2, pp. 129–156.

Fama, E. F. and K. R. French (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of financial economics* 33.1, pp. 3–56.

— (2007). "Disagreement, tastes, and asset prices". In: *Journal of financial economics* 83.3, pp. 667–689.

Hansen, B. E. (2000). "Sample splitting and threshold estimation". In: *Econometrica* 68.3, pp. 575–603.

Kauppi, H. and P. Saikkonen (2008). "Predicting US recessions with dynamic binary response models". In: *The Review of Economics and Statistics* 90.4, pp. 777–791.

Koop, G., M. H. Pesaran, and S. M. Potter (1996). “Impulse response analysis in nonlinear multivariate models”. In: *Journal of econometrics* 74.1, pp. 119–147.

Sharpe, W. F. (1964). “Capital asset prices: A theory of market equilibrium under conditions of risk”. In: *The journal of finance* 19.3, pp. 425–442.

Svirydzenka, K. (2016). *Introducing a new broad-based index of financial development.* International Monetary Fund.

Van den Berg, J., B. Candelon, and J.-P. Urbain (2008). “A cautious note on the use of panel models to predict financial crises”. In: *Economics Letters* 101.1, pp. 80–83.

Zhang, Y., H. J. Wang, and Z. Zhu (2019). “Quantile-regression-based clustering for panel data”. In: *Journal of Econometrics* 213.1, pp. 54–67.

Conclusion générale

Les quatre chapitres de cette thèse sont des contributions originales ayant en commun d'étudier des questions de la finance ou de la macroéconomie à l'aide des instruments de l'économétrie des modèles non linéaires.

Dans le premier chapitre de cette thèse, nous avons cherché à montrer que les décisions d'investissement des fonds ouverts classés comme éthiques ne sont pas toujours en accord avec les principes annoncés par leurs gérants. Nous avons également empiriquement testé l'impact des critères Environnementaux, Sociaux, et de Gouvernance sur la performance financière de ces fonds. Pour ce faire, notre étude est basée sur le CAPM (Sharpe, 1964) et ses extensions à 3 et 4 facteurs (Fama et French, 1993 ; Carhart, 1997) estimés via l'approche en panel non-linéaire d'Hansen (2000). Cette spécification économétrique construite à partir d'une variable dichotomique permet d'étudier la potentielle différence de performance entre les fonds éthiques et les fonds conventionnels, tout en prenant en compte les notes ESG associées aux investissements réalisés de ces fonds. Nos résultats montrent qu'il n'y a pas de différences significatives en termes de performances extra-financières entre les fonds conventionnels et les fonds classés comme éthiques par leurs gérants. D'autre part, nos résultats indiquent que l'impact de critères extra-financiers sur la performance financière est négatif. Ces résultats sont cohérents avec les arguments théoriques Bollen (2007) et Fama et French (2007): les contraintes extra-financières réduisent l'univers d'investissement et donc ne peuvent avoir qu'un effet négatif sur le couple rendement-risque d'un portefeuille.

Dans le deuxième chapitre, nous nous intéressons à la relation entre le développement financier et les crises bancaires. Nous utilisons le panel dichotomique de Candelon, Dumitrescu et Hurlin (2014) sur les données d'une centaine de pays. La correction á la Carro (2007) est également implémentée pour corriger d'éventuels biais liés aux effets fixes de l'approche en panel. Nous utilisons les indicateurs de développement financier étendus par Svirydzenka (2016) décomposés en six sous-indices, nous permettant de déterminer de façon détaillée les facteurs liés aux crises bancaires. Nos résultats montrent une hétérogénéité dans la relation entre les différents sous-indices du développement financier et les crises bancaires selon que les pays soient développés, émergents ou à faible revenu.

Dans le troisième chapitre, nous étudions le pouvoir prédictif du *yield spread* sur les cycles économiques pour un panel de pays de l'OCDE. Dans ce cadre, le panel dichotomique de Candelon, Dumitrescu et Hurlin (2014) est également utilisé. D'autre part, une méthodologie de cluster est développée à partir des travaux de Zhang, Wang et Zhu (2019) pour tester l'homogénéité de la relation entre le *yield spread*, la politique monétaire et le cycle économique. Les résultats du panel confirment le pouvoir prédictif du *yield spread* sur les cycles économiques, tout en étant robuste à différentes variables de contrôle macroéconomiques. Concernant les clusters, deux groupes principaux ressortent, confirmant ainsi une homogénéité partielle du panel. Toutefois, le coefficient du taux directeur des banques centrales n'apparaît pas comme significatif, quelque soit le cluster. Ces résultats peuvent notamment s'expliquer par les instruments de politique monétaire non conventionels utilisés ces dernières années, comment le soutiennent Chinn et Kucko (2015).

Enfin, dans le dernier chapitre, nous déterminons la forme exacte de la fonction de réponse adaptée aux quatre spécifications de Kauppi et Saikkonen (2008). Cette formalisation s'appuie sur la définition de la fonction impulsion-réponse généralisée de Koop, Pesaran et Potter (1996). Cette fonction de réponse est ensuite appliquée pour étudier l'impact d'un choc exogène sur les cycles aux Etats-Unis. Dans cette partie empirique, nous étudions d'abord la relation entre le *yield spread* et les cycles sur la période 1953-2020. Puis, à partir de ces estimations, nous analysons l'impact du choc sur les cycles futurs. Une attention particulière est accordée aux deux modèles dynamiques intégrant la variable binaire retardée. Alors que la spécification qui ne tient compte que de la variable binaire retardée prédit une récession aux Etats-Unis sur un trimestre, la spécification qui utilise également l'indice sous-jacent retardé prédit une récession sur les cinq trimestres à venir. Cette différence s'explique notamment par la structure autorégressive de cette dernière.

Ces différentes contributions montrent que les modèles dichotomiques sont utiles pour mieux comprendre et prévoir les cycles économiques et les risques financiers. L'apport de l'économétrie non-linéaire est significatif en macroéconomie comme en finance. Les prolongements de mes travaux qui me paraissent pertinents seraient d'étudier les propriétés asymptotiques des IRF afin de proposer une forme exacte de l'intervalle de confiance. Une extension de l'IRF au cadre multivarié défini par Candelon, Dumitrescu, Hurlin et Palm (2013) pourrait également faire l'objet de recherches futures. D'autre part, Berg, Candelon et Urbain (2008) mettent en avant la nécessité de vérifier la possibilité de traiter les données en panel à travers des regroupements géographiques. Ce point nécessiterait d'être étudié pour faire suite à notre étude du pouvoir prédictif du *yield spread* sur les cycles économiques pour les 13 pays de l'OCDE. Enfin, Berg, Candelon et Urbain (2008) abordent la notion d'effets d'entraînement comme déterminant dans la transmission des crises financières. Construire une variable de contrôle pour chaque pays qui prendrait en compte la proportion des pays en crise au sein du panel, puis l'ajouter dans notre modèle, pourrait également être une piste de développement.

Bibliography

Bollen, N. P. (2007). “Mutual fund attributes and investor behavior”. In: *Journal of Financial and Quantitative Analysis* 42.3, pp. 683–708.

Candelon, B., E.-I. Dumitrescu, and C. Hurlin (2014). “Currency crisis early warning systems: Why they should be dynamic”. In: *International Journal of Forecasting* 30.4, pp. 1016–1029.

Candelon, B., E.-I. Dumitrescu, C. Hurlin, and F. C. Palm (2013). *Multivariate Dynamic Probit Models: An Application to Financial Crises Mutation', VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims (Advances in Econometrics, Volume 32).*

Carhart, M. M. (1997). “On persistence in mutual fund performance”. In: *The Journal of finance* 52.1, pp. 57–82.

Carro, J. M. (2007). “Estimating dynamic panel data discrete choice models with fixed effects”. In: *Journal of Econometrics* 140.2, pp. 503–528.

Chinn, M. and K. Kucko (2015). “The predictive power of the yield curve across countries and time”. In: *International Finance* 18.2, pp. 129–156.

Fama, E. F. and K. R. French (1993). “Common risk factors in the returns on stocks and bonds”. In: *Journal of financial economics* 33.1, pp. 3–56.

— (2007). “Disagreement, tastes, and asset prices”. In: *Journal of financial economics* 83.3, pp. 667–689.

Hansen, B. E. (2000). “Sample splitting and threshold estimation”. In: *Econometrica* 68.3, pp. 575–603.

Kauppi, H. and P. Saikkonen (2008). “Predicting US recessions with dynamic binary response models”. In: *The Review of Economics and Statistics* 90.4, pp. 777–791.

Koop, G., M. H. Pesaran, and S. M. Potter (1996). "Impulse response analysis in nonlinear multivariate models". In: *Journal of econometrics* 74.1, pp. 119–147.

Sharpe, W. F. (1964). "Capital asset prices: A theory of market equilibrium under conditions of risk". In: *The journal of finance* 19.3, pp. 425–442.

Svirydzenka, K. (2016). *Introducing a new broad-based index of financial development.* International Monetary Fund.

Van den Berg, J., B. Candelon, and J.-P. Urbain (2008). "A cautious note on the use of panel models to predict financial crises". In: *Economics Letters* 101.1, pp. 80–83.

Zhang, Y., H. J. Wang, and Z. Zhu (2019). "Quantile-regression-based clustering for panel data". In: *Journal of Econometrics* 213.1, pp. 54–67.

RÉSUMÉ

Cette thèse sur articles est composée de quatre chapitres autonomes, contribuant au domaine de l'économétrie non-linéaire. Le premier chapitre s'intéresse à l'apport de l'économétrie non-linéaire à travers la mesure de la performance financière en utilisant une variable dichotomique comme variable indépendante. Les trois chapitres suivants sont basés sur les modèles de régression non-linéaire où la variable dichotomique est la variable dépendante de l'équation. Compte tenu des liens entre le risque financier et le contexte macroéconomique, cette partie est liée au thème de l'allocation optimale via l'étude des crises et récessions. Cette classe de modèle (probit/logit) est utilisée dans le second chapitre pour étudier empiriquement le rôle du développement financier dans la probabilité d'occurrence de crises bancaires. Ensuite, les deux derniers chapitres se concentrent sur le cadre méthodologique développé par Kauppi et Saikkonen (2008) et Candelon, Dumitrescu et Hurlin (2012 ; 2014) au sujet de la prévision des cycles économiques à partir de modèles probit/logit. Ainsi, le troisième chapitre étudie la relation empirique liant l'évolution du spread de taux et la probabilité future d'expansion / récession dans un panel de données élargi tout en testant l'homogénéité de cette relation. Enfin, le quatrième chapitre propose une contribution théorique en dérivant les fonctions de réponse des modèles probit/logit à partir de l'approche de Kauppi et Saikkonen (2008). Ces fonctions de réponse sont ensuite utilisées dans un cadre empirique afin d'estimer l'impact d'un choc exogène sur le cycle expansion / récession.

MOTS CLÉS

Econométrie Non-Linéaire, Modèle Dichotomique, Fonction de Réponse, Cycles Economiques

ABSTRACT

This paper-based thesis is composed of four autonomous chapters and contributes to the field of nonlinear econometrics. The first chapter focuses on the contribution of nonlinear econometrics through the measurement of financial performance using a dichotomous variable as an independent variable. The next three chapters are based on nonlinear regression models where the dichotomous variable is the dependent variable in the equation. Given the links between financial risk and the macroeconomic context, this section is linked to the theme of optimal allocation through the study of crises and recessions. This class of model (probit/logit) is used in the second chapter to empirically study the role of financial development in the probability of the occurrence of banking crises. Then, the last two chapters focus on the methodological framework developed by Kauppi and Saikkonen (2008) and Candelon, Dumitrescu and Hurlin (2012; 2014) concerning the forecasting of business cycles using probit/logit models. Thus, the third chapter examines the empirical relationship linking the evolution of the interest rate spread and the future probability of expansion/recession in an extended data panel while testing the homogeneity of this relationship. Finally, the fourth chapter proposes a theoretical contribution by deriving the response functions of probit/logit models from the approach of Kauppi and Saikkonen (2008). These response functions are then used in an empirical framework to estimate the impact of an exogenous shock on the expansion/recession cycle.

KEYWORDS

Non-Linear Econometrics, Dichotomous Model, Impulse Response Function, Economic Cycles