



Contributions à la statistique des processus : estimation, prédiction et extrêmes

Olivier Wintenberger

► **To cite this version:**

Olivier Wintenberger. Contributions à la statistique des processus : estimation, prédiction et extrêmes. Statistiques [math.ST]. Université Paris Dauphine - Paris IX, 2012. tel-00757756

HAL Id: tel-00757756

<https://tel.archives-ouvertes.fr/tel-00757756>

Submitted on 27 Nov 2012

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

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

CEREMADE — UMR 7534
UNIVERSITÉ PARIS-DAUPHINE

MÉMOIRE D'HABILITATION À DIRIGER DES
RECHERCHES

Olivier WINTENBERGER

Contributions à la statistique des
processus : estimation, prédiction et
extrêmes

Soutenu le 23 novembre devant le jury :

M. JEAN-MARC BARDET
Mme. FABIENNE COMTE (PRÉSIDENTE)
M. PAUL DOUKHAN
M. THOMAS MIKOSCH
M. VINCENT RIVOIRARD (COORDINATEUR)
M. JOHAN SEGERS

Rapporteurs :

M. ISTVAN BERKES
Mme. FLORENCE MERLEVÈDE
M. ERIC MOULINES

Mémoire préparé au
CEREMADE
Université Paris-Dauphine
Place du Maréchal De Lattre De Tassigny
75 775 PARIS CEDEX 16 - FRANCE

Résumé

Ce mémoire d'habilitation traite de la statistique des processus à temps discret faiblement dépendants. Une première partie présente des résultats asymptotiques d'estimation pour les paramètres de modèles affines généraux. La méthode étudiée est la maximisation du critère de quasi-vraisemblance. Afin de traiter de possibles ruptures de stationnarité, nous pénalisons ce critère par le nombre de ruptures. Pour les modèles à volatilité comme le modèle EGARCH, cette procédure est instable et nous proposons de contraindre le critère au domaine dit d'inversibilité continue. Nous étudions le problème de la prédiction de processus faiblement dépendants dans une seconde partie. Les résultats obtenus sont des inégalités d'oracle non asymptotiques nécessitant l'étude préalable des propriétés de concentration gaussiennes de lois faiblement dépendantes. Pour ce faire nous utilisons une notion de transport faible et de nouvelles inégalités dites de transport conditionnel. Enfin, le comportement des extrêmes en présence de dépendance fait l'objet de la troisième partie. Nous introduisons un indice de *cluster* qui caractérise les lois limites α -stables dans le théorème de la limite centrale et les grandes déviations des sommes partielles à variation régulière. Nous traitons des exemples de processus à queues épaisses tels que les solutions des équations récurrentes stochastiques linéaires et le modèle GARCH. Nous appliquons ces résultats pour caractériser asymptotiquement les erreurs d'estimation des auto-covariances de processus à queues épaisses.

Mots-clefs : Processus faiblement dépendants, théorèmes limites, théorie des valeurs extrêmes, indice de *cluster*, inégalités de transport faible, inégalités d'oracle, quasi maximum de vraisemblance.

CONTRIBUTIONS TO TIME SERIES ANALYSIS: ESTIMATION, PREDICTION AND EXTREMES

Abstract

This habilitation manuscript presents my research work on statistics for weakly dependent processes. Asymptotical results for the Quasi Maximum Likelihood Estimator in general affine models are given in the first part. To detect stationarity breaks, we suggest to penalize the Quasi Likelihood criteria by the number of breaks. For some volatility models such as EGARCH model, the procedure is not stable and we suggest to constrain the criteria on the continuously invertible domain. Then we consider the one-step prediction of weakly dependent processes, establishing new oracle inequalities. Such non asymptotical results need to assert the gaussian concentration properties of weakly dependent measures. To this aim, we propose a notion of weak transport and new conditional transport inequalities. Finally, we introduce the cluster index to characterize the extremal behavior of regularly varying partial sums. We obtain limit properties such as the α -stable limits in the Central Limit Theorem or the large deviations in presence of dependent extremes. Solutions of linear stochastic recurrent equation and the GARCH model are examples of heavy tailed processes. We apply our results to characterize asymptotically the estimation errors of heavy tailed processes autocovariances.

Keywords : Weakly dependent processes, limit theorems, extreme value theory, cluster index, weak transport inequalities, oracle inequalities, Quasi Maximum Likelihood Estimation.

Remerciements

Le travail décrit dans ce mémoire a vu le jour grâce à de nombreuses personnes que je souhaite remercier ici.

Paul Doukhan m'a soutenu dès la première heure. C'est grâce à lui que le domaine de recherche des statistiques des processus est aussi développé en France. Je profite énormément de ses travaux précurseurs ainsi que de son cycle thématique qui attire la communauté internationale de ce domaine scientifique à Paris. Son association avec Jean-Marc Bardet pour diriger ma thèse m'a énormément apportée. Jean-Marc est un aventurier et avec lui je me suis lancé dans de nouvelles thématiques. Je lui serai aussi toujours reconnaissant de me permettre d'intervenir au Bénin pour donner des cours de statistique. Cette expérience humaine est très importante pour moi.

Since my post doctoral position in Copenhagen, it has been a real pleasure to be in touch with Thomas Mikosch. I really benefit of his experience that he shares with me during pleasant discussions. It influences me deeply, and not only professionally. The time I spend in Copenhagen is always great and I hope our collaboration will long for many years. Our work has been influenced by ideas developed by Adam Jakubowski and I am proud to collaborate with Adam too. I would also thank Istvan Berkes for doing a report of my thesis. The wideness of his research work is a model for me.

Vincent Rivoirard a joué un grand rôle dans ce travail en acceptant d'être mon coordinateur. Je lui en serai toujours reconnaissant. J'exprime aussi ma gratitude à Florence Merlevède pour la justesse de ses conseils scientifiques. Eric Moulines a quant à lui joué un rôle moteur en initiant cette habilitation au détour d'une conversation, il y a de cela quelques mois. Je lui dois énormément, aussi pour avoir pris de son temps si précieux pour rapporter ce mémoire. Le jury ne serait pas complet sans Fabienne Comte et Johan Segers qui sont venus apporter leur expertise, Fabienne dans le domaine de l'apprentissage statistique et Johan dans le domaine des valeurs extrêmes.

Il y a cinq ans que je suis enseignant-chercheur au CEREMADE. Je remercie l'ensemble des membres du laboratoire pour créer un environnement idéal pour la recherche. En particulier, ma collègue Katia Meziani apporte beaucoup de bonne humeur à notre bureau B513. Je remercie aussi toute la bande du cinquième étage, en particulier Anthony, FX, Guillaume, Julien et Pierre avec qui j'ai partagé des cafés et des discussions mathématiques. Je partage aussi de bons moments avec l'équipe du sixième, en particulier les membres du Ceremade Football Club Boris, Bruno, François B., Matthieu, Philippe, Rémi et Vincent, ainsi qu'avec mon professeur de squash François S. Notre recherche bénéficie du support administratif apporté par Isabelle, sans qui je n'aurais pas de lecteur CD, Marie, avec qui je partage la passion de la course à pied, Gilles, qui a toujours de bons conseils tant en informatique qu'en bricolage, et surtout César, en charge de mes missions et de ma dose de caféine journalière indispensable. Enfin je tenais à remercier José Trashorras avec qui j'ai collaboré à grand renfort de chocolats et de pauses "Coca".

Je suis co-responsable du master actuariat de l'université Paris Dauphine avec Romuald Elie à qui je dois énormément. Je le remercie tout particulièrement pour égayer nos réunions grâce à son humour et son rire communicatif. Un grand merci aussi à l'ensemble de l'administration du MIDO, avec une mention spéciale pour la secrétaire du

master Judith, qui gère tous les dossiers de main de maître et avec le sourire, Jean-Paul, le seul à me complimenter pour mes pulls flashy, et Patricia qui corrige l'orthographe de tous les rapports avec une patience inouïe.

Je suis aussi depuis un an affilié au LFA-CREST. Jean-Michel Zakoïan et Christian Francq sont devenus des collègues particulièrement patients à l'écoute de mes divagations scientifiques. J'apprécie leurs rigueurs scientifiques et leurs conseils avisés. J'ai toujours aimé travailler en équipe et j'ai la chance d'avoir des collaborateurs qui sont tous des amis. Merci à Irène Gannaz avec qui je partage un article, une traversée des dômes de Miages ainsi que d'autres bons moments en montagne. Je remercie énormément Pierre Alquier pour nos escapades scientifiques et amicales au CIRM et à St Flour. Merci à William Kengne pour sa générosité, en particulier dans le travail, et à Xiaoyin Li pour sa gentillesse. Je pense aussi aux personnes qui m'ont conseillé et avec qui j'espère collaborer dans le futur comme Yannig Goude et Sylvain Arlot sur la prédiction, Patrice Bertail et Stéphan Cléménçon sur la régénération, Nathael Gozlan et Paul-Marie Samson sur le transport, Jérôme Dedecker sur le couplage, Philippe Soulier sur les extrêmes et Marc Hoffmann sur les applications financières.

Enfin, d'autres personnes ont joué un rôle beaucoup moins professionnel mais tout aussi important. Charles Bouveyron, Cécile Hardouin et l'ensemble de la grande famille du SAMM(OS) m'ont très fortement encouragé, sur un mur d'escalade et ailleurs. Merci à Aurélien Garivier qui a relu ce manuscrit et m'a fait part de fins commentaires, pas seulement mathématiques. Merci aussi à toute ma famille, en particulier ma mère pour ses conseils en français et Pascale pour ses conseils en anglais, mon père pour le vin blanc et mon beau père pour le vin rouge. Enfin, je dédie ce mémoire à ma compagne Nathalie qui l'a relu plusieurs fois pour corriger les fautes d'orthographe. Je la remercie pour cette belle preuve d'amour et pour notre vie commune heureuse et fructueuse en tant que parents d'Alix et de sa petite soeur d'ici peu.

Table des matières

Présentation	8
Parcours personnel	8
Organisation du mémoire	8
I Utilisation du Quasi Maximum de Vraisemblance dans les modèles à volatilité	10
1 L'estimation de ruptures dans les modèles affines	11
1.1 Les modèles affines généraux	11
1.2 Le problème de ruptures de stationnarité	11
1.3 Estimation par Quasi-Maximum de Vraisemblance (QMV)	11
1.4 Hypothèses de Lipschitz uniforme et de dépendance faible	12
1.5 Résultats asymptotiques	13
1.6 Calibration pratique par l'heuristique de pente	14
1.7 Application sur données réelles	14
2 L'estimateur du QMV dans le modèle EGARCH	15
2.1 Les modèles à volatilité stochastique	15
2.2 Inversibilité du modèle	16
2.3 Modèles continûment inversibles	17
2.4 Résultats asymptotiques pour des modèles continûment inversibles	18
2.5 Normalité asymptotique dans le modèle EGARCH	18
2.6 Le modèle log-GARCH	19
II La prédiction de séries Γ-faiblement dépendantes	22
3 Apprentissage statistique et transport de lois faiblement dépendantes	23
3.1 Le cadre d'apprentissage statistique	23
3.2 Inégalités de transport faible	25
3.3 La Γ -faible dépendance	26
3.4 Transport faible pour les lois Γ -faiblement dépendantes	28
3.5 Inégalités exponentielles	28

4	Inégalités d'oracle	29
4.1	Transport faible conditionnel	29
4.2	Inégalités d'oracle exactes à vitesses rapides	31
4.3	Inégalités d'oracle non exactes à vitesses rapides	32
4.4	Inégalités d'oracle exactes à vitesses lentes	32
4.5	Etude de Monte-Carlo dans le cas de prévisions linéaires	33
III	Théorie asymptotique des agrégats de valeurs extrêmes	34
5	L'indice de <i>cluster</i>	35
5.1	Grandes déviations dans le cas iid	35
5.2	Les processus à variation régulière	35
5.3	Condition de dérive des chaînes de Markov	36
5.4	Définition de l'indice de <i>cluster</i>	37
5.5	Lien avec l'indice extrémal	38
5.6	Calcul de l'indice de <i>cluster</i> sur des exemples	39
6	Théorie asymptotique des sommes partielles dépendantes et à variation régulière	40
6.1	Théorème de la limite centrale gaussienne	40
6.2	Théorème de la limite centrale α -stable	40
6.3	Principe uniforme de grandes déviations	42
6.4	Exemples de processus à queues épaisses	42
6.5	Comportement asymptotique des auto-covariances empiriques	43
	Conclusion et perspectives	46
	Références	48

Présentation

Parcours personnel

Après mon doctorat à l'Université Paris 1 sous la direction de J.-M. Bardet et P. Doukhan, j'ai poursuivi mon travail de recherche dans le domaine de la statistique des processus en post doctorat à l'Université de Copenhague. Sous la supervision de T. Mikosch, j'y ai étudié la théorie des valeurs extrêmes et la problématique associée de quantification des risques liés à des événements rares de valeurs extrêmes. De retour en France, j'ai poursuivi mon travail de recherche au sein du laboratoire du CEREMADE en tant que maître de conférences à l'université Paris Dauphine. J'y apprécie les discussions enrichissantes avec les collègues du laboratoire sur des sujets variés de mathématiques appliquées. En particulier, j'ai eu l'opportunité de participer au groupe de travail EVOL sur la thématique du transport de lois. J'ai aussi pris la co-responsabilité, avec R. Elie, du master actuariat de l'université Paris Dauphine. Dans ce cadre, j'ai encadré de nombreux mémoires de fin d'étude sur les méthodes pratiques de quantification des risques en assurance et en finance. J'ai de plus participé officieusement à l'encadrement de la thèse de W. Kengne et je co-encadre la thèse de S. Cai avec P. Doukhan et J.-L. Prigent. Enfin, depuis un an, je suis aussi chercheur affilié au sein du Laboratoire de Finance et Assurances au CREST au côté de C. Francq et J.-M. Zakoïan.

Organisation du mémoire

Ce mémoire expose mes contributions à la statistique des processus faiblement dépendants. Alors que le cadre statistique classique est celui de données indépendantes et identiquement distribuées (iid), ce cadre dépendant se veut plus réaliste : les modèles sont introduits pour tenir compte d'une possible interaction entre les données observées. J'étudie des processus à temps discrets, appelés aussi séries temporelles, qui ont des données passées, présentes et futures possiblement dépendantes. On parle de dépendance faible lorsque les résultats statistiques restent de même nature que ceux obtenus dans le cadre iid. Parmi les thèmes classiques de ce domaine de recherche, je traite dans mon travail de doctorat de l'estimation paramétrique [W4], non paramétrique [W1, W5] et du théorème de la limite centrale gaussienne [W2]. J'obtiens des résultats similaires au cas iid sous des conditions de dépendance faible les plus générales possibles. Les modèles traités sont, par exemple, les chaînes à mémoire infinie introduites dans [W3].

Ce mémoire présente les solutions que j'apporte à trois problèmes distincts : l'estimation dans les modèles instables, le traitement non asymptotique de la prédiction statistique et les théorèmes limites en présence de valeurs extrêmes. Pour résoudre ces problèmes, j'ai développé des conditions de dépendance faible spécifiques :

- Pour l'estimation dans les modèles à volatilité stochastique, j'utilise des conditions dites d'inversibilité continue qui assurent la stabilité de la procédure statistique.
- Afin d'obtenir des propriétés de concentration gaussiennes des lois faiblement dépendantes, j'introduis des nouveaux coefficients de dépendance faible notés Γ . J'en déduis des résultats d'apprentissage statistique pour la prédiction.
- Je définis l'indice de *cluster* dans un cadre markovien pour quantifier les agrégats de valeurs extrêmes pour des processus à variation régulière.

Donnons plus précisément le contenu de ce mémoire chapitre par chapitre.

Je traite dans le chapitre 1 des ruptures de stationnarité dans les modèles affines. En collaboration avec W. Kengne et J.-M. Bardet, nous utilisons le critère de Quasi Maximum de Vraisemblance (QMV) pénalisé par le nombre de ruptures [W9]. Sous des conditions de dépendance faible, nous obtenons la normalité asymptotique de l'estimateur des paramètres. Nous proposons de calibrer la méthode en suivant l'heuristique de pente.

Dans le chapitre 2, j'introduis l'inversibilité continue pour traiter des modèles, comme le modèle EGARCH, pour lesquels la procédure du QMV est instable. Je montre dans [W11] que l'estimation par QMV est asymptotiquement normale lorsqu'on restreint le critère au domaine d'inversibilité continue. En collaboration avec C. Francq et J.-M. Zakoïan, nous étudions le modèle log-GARCH comme alternative au modèle EGARCH pour laquelle la procédure statistique par QMV est toujours stable [W16].

Le chapitre 3 décrit l'approche d'apprentissage statistique en prédiction des processus de [W8]. Afin d'obtenir des résultats non asymptotiques appelés inégalités d'oracle, j'introduis les conditions de Γ -faible dépendance [W14]. J'en déduis des inégalités de transport faible et des extensions d'inégalités exponentielles classiques au cadre Γ -faiblement dépendant. Les résultats présentés ici sont aussi précis que dans le cas iid, améliorant ainsi les résultats obtenus précédemment dans [W6].

Dans le chapitre 4, les inégalités d'oracle pour la prédiction sont obtenues par une nouvelle approche dite de transport conditionnel. J'en déduis des bornes non asymptotiques du risque de prédiction de l'estimateur des moindres carrés dans les modèles linéaires. Ces résultats sont nouveaux, même dans le cas iid. Suivant l'approche PAC-bayésienne, en collaboration avec P. Alquier nous étudions les estimateurs de Gibbs comme une alternative aux méthodes de sélection de modèles en prédiction [W13].

J'introduis dans le chapitre 5 l'indice de *cluster* quantifiant les valeurs extrêmes de processus faiblement dépendants. En collaboration avec T. Mikosch nous étudions le comportement limite des valeurs extrêmes formant des agrégats (*cluster* en anglais) dans les sommes de données dépendantes à variation régulière. Dans un contexte markovien, nous caractérisons l'indice de *cluster* en fonction du processus spectral de queue [W15]. L'indice est calculé pour la solution de l'équation récurrente stochastique linéaire et le modèle GARCH.

Le chapitre 6 présente de nouveaux théorèmes limites caractérisés par l'indice de *cluster*. Avec T. Mikosch, je détermine les lois limites α -stables des théorèmes centraux obtenus dans [W7] en fonction de l'indice de *cluster*. Cet indice quantifie aussi le principe de grandes déviations des sommes partielles de variables à variation régulière dépendantes [W10]. Nous montrons que l'indice de *cluster* et l'indice extrémal sont tous deux indispensables à l'étude des processus à queues de distribution épaisses. Nous appliquons ces résultats pour caractériser asymptotiquement les erreurs d'estimation des auto-covariances de processus à queues épaisses.

Je fournis quelques illustrations de ces résultats de statistique mathématique sur simulations et sur données réelles.

Première partie

Utilisation du Quasi Maximum de Vraisemblance dans les modèles à volatilité

Sommaire

1	L'estimation de ruptures dans les modèles affines	11
1.1	Les modèles affines généraux	11
1.2	Le problème de ruptures de stationnarité	11
1.3	Estimation par Quasi-Maximum de Vraisemblance (QMV)	11
1.4	Hypothèses de Lipschitz uniforme et de dépendance faible	12
1.5	Résultats asymptotiques	13
1.6	Calibration pratique par l'heuristique de pente	14
1.7	Application sur données réelles	14
2	L'estimateur du QMV dans le modèle EGARCH	15
2.1	Les modèles à volatilité stochastique	15
2.2	Inversibilité du modèle	16
2.3	Modèles continûment inversibles	17
2.4	Résultats asymptotiques pour des modèles continûment inversibles	18
2.5	Normalité asymptotique dans le modèle EGARCH	18
2.6	Le modèle log-GARCH	19

1 L'estimation de ruptures dans les modèles affines

Avec J-M Bardet et W. Kengne j'ai étudié le problème de détection de ruptures dans [W9] à partir des résultats asymptotiques obtenus durant ma thèse sur l'estimateur du Quasi Maximum de Vraisemblance (QMV) [W4].

1.1 Les modèles affines généraux

Depuis le prix Nobel d'économie décerné en 2003 à MM Engle et Granger, une attention particulière est portée sur l'étude des modèles ARMA-GARCH et leurs généralisations [24]. Ces modèles affines sont utilisés en statistique car l'estimation de leurs paramètres est simple. Par ailleurs, ils partagent certaines propriétés observées sur des données réelles ayant des valeurs extrêmes dépendantes (ils prennent en compte le phénomène de *cluster* des valeurs extrêmes présenté dans la partie III).

Nous traitons de l'estimation paramétrique de modèles affines de la forme suivante :

Définition 1.1. Soit T un intervalle de \mathbb{Z} . Le processus à valeurs réelles $X = (X_t)_{t \in T}$ suit un modèle affine généralisé noté $\mathcal{M}_T(\sigma_\theta, f_\theta)$ s'il satisfait la relation :

$$X_t = \sigma_\theta((X_{t-i})_{i \geq 1})Z_t + f_\theta((X_{t-i})_{i \geq 1}) \quad \text{pour tout } t \in T, \quad (1)$$

où $(Z_t)_{t \in \mathbb{Z}}$ est le processus des innovations iid telles que $\mathbb{E}Z_0 = 0$ et $\mathbb{E}Z_0^2 = 1$ et si (X_t) est non anticipatif (X_t indépendant de $(Z_{t+1}, Z_{t+2}, \dots)$ pour tout $t \in T$). Les paramètres du modèle θ sont dans un ensemble Θ de dimension d : $\Theta \subseteq \mathbb{R}^d$.

A titre d'exemples, les modèles suivants entrent dans ce cadre : les modèles linéaires ARMA inversibles ($\sigma_\theta = 1$), les modèles à volatilité stochastique ARCH, TARARCH, EGARCH ($f_\theta = 0$, voir le chapitre 2 pour une étude spécifique) et plus généralement les modèles type ARMA-GARCH où $f_\theta \neq 0$ et $\sigma_\theta \neq 1$.

1.2 Le problème de ruptures de stationnarité

Pour traiter de manière réaliste de longues périodes d'observations, nous supposons des ruptures dans les valeurs des coefficients : les observations (X_1, \dots, X_n) proviennent d'un processus $X = (X_t)_{t \in \mathbb{Z}}$ tel que $X \in \mathcal{M}_{T_j^*}(M_{\theta_j^*}, f_{\theta_j^*})$ pour tout $j = 1, \dots, K^*$ avec

- $K^* \geq 0$, $T_j^* = \{t_{j-1}^* + 1, t_{j-1}^* + 2, \dots, t_j^*\}$ avec $0 < t_1^* < \dots < t_{K^*-1}^* < n$, $t_j^* \in \mathbb{N}$ et par convention $t_0^* = -\infty$ and $t_{K^*}^* = \infty$;
- $\theta_j^* = (\theta_{j,1}^*, \dots, \theta_{j,d}^*) \in \Theta \subset \mathbb{R}^d$ pour tout $j = 1, \dots, K^*$.

Nous considérons que non seulement les paramètres θ_j^* sont inconnus mais aussi les instants de ruptures $t_1^*, \dots, t_{K^*-1}^*$ et le nombre $K^* - 1$ de ruptures. Le fait que le nombre de ruptures K^* soit inconnu est particulièrement difficile à traiter en statistique car le nombre de paramètres à estimer dépend de K^* .

1.3 Estimation par Quasi-Maximum de Vraisemblance (QMV)

Par définition, si $X \in \mathcal{M}_T(f_\theta, \sigma_\theta)$ alors l'espérance de X_{t+1} conditionnellement à son propre passé X_t, X_{t-1}, \dots est égale à $f_\theta(X_{t-1}, X_{t-2}, \dots) =: f_t(\theta)$ et sa variance

conditionnelle à $\sigma_\theta(X_{t-1}, X_{t-2}, \dots)^2 =: \sigma_t(\theta)$. La méthode de Quasi Vraisemblance (QV) consiste à considérer la vraisemblance dans le cas où les innovations Z_t sont gaussiennes :

$$L_n(T, \theta) := -\frac{1}{2} \sum_{t \in T} \ell_t(\theta) \quad \text{avec} \quad \ell_t(\theta) := \frac{(X_t - f_t(\theta))^2}{\sigma_t^2(\theta)} + \log(\sigma_t^2(\theta)).$$

Comme seules X_1, \dots, X_n sont observées, la fonction L_n ne peut pas être calculée car elle dépend des valeurs passées inconnues X_{-j} pour $j \geq 0$. On approche L_n par la QV \widehat{L}_n définie comme

$$\widehat{L}_n(T, \theta) := -\frac{1}{2} \sum_{t \in T} \widehat{\ell}_t(\theta) \quad \text{où} \quad \widehat{\ell}_t(\theta) := \frac{(X_t - \widehat{f}_t(\theta))^2}{\widehat{\sigma}_t^2(\theta)} + \log(\widehat{\sigma}_t^2(\theta))$$

avec $\widehat{f}_t(\theta) = f_\theta(X_{t-1}, \dots, X_1, u)$ et $\widehat{\sigma}_t^2(\theta) = \sigma_\theta^2(X_{t-1}, \dots, X_1, u)$ pour toute suite de valeurs initiales $u = (u_n)$ avec un nombre fini de valeurs non nulles.

Remarque 1.2. Sous des conditions de dépendance faible, les valeurs non nulles u_i peuvent être choisies arbitrairement en fonction du coefficient θ , d'un nombre fixe de premières observations, ... sans changer les résultats asymptotiques qui sont donnés dans ce chapitre. La méthode est dite stable par rapport aux conditions initiales choisies arbitrairement. Le chapitre 2 traite des hypothèses minimales assurant que la méthode par QMV soit stable.

La QV se généralise au cas de $K - 1 \geq 0$ ruptures avec $\underline{t} = (t_1, \dots, t_K)$ le vecteur des instants de ruptures et $\underline{\theta} = (\theta_1, \dots, \theta_K)$ la matrice des coefficients. Pour $T_j = \{t_{j-1} + 1, t_{j-1} + 2, \dots, t_j\}$, le critère à minimiser est le suivant

$$\widehat{J}_n(K, \underline{t}, \underline{\theta}) := -2 \sum_{k=1}^K \widehat{L}_n(T_k, \theta_k).$$

Pour estimer K^* nous pénalisons ce critère par le nombre de ruptures K :

$$\widetilde{J}_n(K, \underline{t}, \underline{\theta}) := \widehat{J}_n(K, \underline{t}, \underline{\theta}) + \kappa_n K \tag{2}$$

où $0 < \kappa_n \leq n$ est le paramètre de régularisation. En considérant un nombre maximum $K_{max} > 0$ de ruptures nous obtenons un estimateur par QV pénalisée en minimisant :

$$(\widehat{K}_n, \widehat{\underline{t}}_n, \widehat{\underline{\theta}}_n) \in \underset{1 \leq K \leq K_{max}}{\text{Argmin}} \underset{\underline{t}, \underline{\theta}}{\text{Argmin}} (\widetilde{J}_n(K, \underline{t}, \underline{\theta})). \tag{3}$$

Remarque 1.3. Le choix d'une pénalité additive nous permet d'utiliser des algorithmes dynamiques optimisant en K avec un coût algorithmique linéaire $O(K_{max})$. Le calcul de la QV pour toutes les périodes de stationnarité possibles est plus long car quadratique en le nombre d'observations $O(n^2)$.

1.4 Hypothèses de Lipschitz uniforme et de dépendance faible

Afin d'étudier les propriétés de cet estimateur, nous spécifions un compact Θ sur lequel nous minimisons le critère pour chaque paramètre θ_j , $1 \leq j \leq K$. Le modèle affine

dépendant des fonctions Ψ_θ (écriture commune pour $\Psi_\theta = \sigma_\theta$ et $\Psi_\theta = f_\theta$), nous considérons les espaces de fonctions régulières (continues ou deux fois continûment dérivables) sur le compact Θ . Deux types d'hypothèses sont faites. Afin de les traiter simultanément, notons \mathcal{K} un compact qui est soit un singleton $\{\theta\}$ pour $\theta \in \Theta$, soit Θ lui-même. Munissons les espaces de fonctions continues de la norme uniforme notée $\|\cdot\|_{\mathcal{K}}$. Nous supposons que les fonctions Ψ_θ sont Lipschitz uniformes [W4] :

Définition 1.4. Pour $i = 0, 1, 2$, supposant $\|\partial^i \Psi_\theta(0)/\partial \theta^i\|_{\mathcal{K}} < \infty$, nous définissons les coefficients de Lipschitz uniforme $(\alpha_k^{(i)}(\Psi_\theta, \mathcal{K}))_{k \geq 1}$ tels que $\sum_{k=1}^{\infty} \alpha_k^{(i)}(\Psi_\theta, \mathcal{K}) < \infty$ et

$$\left\| \frac{\partial^i \Psi_\theta(x)}{\partial \theta^i} - \frac{\partial^i \Psi_\theta(y)}{\partial \theta^i} \right\|_{\mathcal{K}} \leq \sum_{k=1}^{\infty} \alpha_k^{(i)}(\Psi_\theta, \mathcal{K}) |x_k - y_k| \quad \text{pour tout } x = (x_i), y = (y_i) \in \mathbb{R}^{\mathbb{N}}.$$

La première hypothèse consiste à supposer l'existence de moments d'ordre $r \geq 1$ pour le critère de QV pénalisé \tilde{J}_n pour chaque $\theta \in \Theta$ grâce à un théorème de point fixe dans l'espace des fonctions r -intégrables. Notons

$$\beta^{(i)}(\mathcal{K}) := \sum_{k \geq 1} \beta_k^{(i)}(\mathcal{K}) \quad \text{avec} \quad \beta_k^{(i)}(\mathcal{K}) := \alpha_k^{(i)}(f_\theta, \mathcal{K}) + (\mathbb{E}|Z_0|^r)^{1/r} \alpha_k^{(i)}(\sigma_\theta, \mathcal{K}),$$

et supposons le modèle contractant

(A) : Pour tout $\theta \in \Theta$ nous avons $\beta^{(0)}(\{\theta\}) < 1$.

La seconde hypothèse consiste à contrôler la décroissance des coefficients de Lipschitz uniforme sur Θ afin que le critère \tilde{J}_n dépende faiblement des valeurs observées au début de la période de stationnarité T_j^* . Comme notre critère contient un terme de pénalisation, cette condition de dépendance faible s'écrit en fonction du paramètre de régularisation κ_n :

(H_{*i*}), $i = 0, 1, 2$: Le paramètre de régularisation est tel que $\kappa_n \wedge n \kappa_n^{-1} \rightarrow \infty$ lorsque $n \rightarrow \infty$ et pour tout $j = 1, \dots, i$:

$$\sum_{k \geq 2} \kappa_k^{-(r/4 \wedge 1)} \left(\sum_{\ell \geq u_k} \beta_\ell^{(j)}(\Theta) \right)^{(r/4 \wedge 1)} < \infty$$

où $u_k = k/(c \log k)$ avec $c > 0$.

Remarque 1.5. Nous verrons dans le chapitre 2 comment unifier efficacement ces conditions de moments et de dépendance faible dans le cadre des modèles à volatilité où $f_\theta = 0$.

1.5 Résultats asymptotiques

Nous prouvons la consistance de notre procédure généralisant [37].

Théorème 1.6 ([W9]). Supposons le modèle affine général identifiable tel que $h_\theta^t \geq m > 0$ pour tout $\theta \in \Theta$. Si les instants de ruptures vérifient la relation $t_i^* = \tau_i^* n$ pour $0 < \tau_i^* < 1$ pour tout $1 \leq i \leq K^*$ et (A) et (H₀) sont satisfaites avec $r = 2$ alors

$$(\hat{K}_n, \hat{t}_n/n, \hat{\theta}_n) \rightarrow (K^*, \underline{\tau}^*, \underline{\theta}^*) \quad \text{en probabilité.}$$

Nous obtenons sous des hypothèses plus fortes (moments d'ordre $r = 4$, fonctions deux fois continûment dérivables, ...) la normalité asymptotique des estimateurs $\hat{\theta}_j$.

Théorème 1.7 ([W9]). *Sous les conditions du Théorème 1.6 avec de plus (\mathbf{A}) et (\mathbf{H}_2) satisfaites pour $r \geq 4$, si $\kappa_n = O(\sqrt{n})$ et si $\theta_j^* \in \overset{\circ}{\Theta}$ nous obtenons*

$$\sqrt{(t_j^* \wedge n) - (t_{j-1}^* \vee 0)} (\hat{\theta}_n(\hat{T}_j) - \theta_j^*) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}_d(0, \Sigma), \quad (4)$$

où la matrice de variance-covariance asymptotique Σ est inversible.

Le paramètre de régularisation κ_n dans le Théorème 1.7 est contraint à être plus petit que \sqrt{n} . En pratique nous calibrons κ_n par l'heuristique de pente.

1.6 Calibration pratique par l'heuristique de pente

Les résultats précédents sont vérifiés sous l'hypothèse que les périodes de stationnarité T_j^* sont proportionnelles à n afin de négliger le cas de périodes qui seraient trop courtes pour pouvoir faire une estimation précise des coefficients θ_j . La suite (u_n) apparaissant dans les conditions (\mathbf{H}_i) indique la taille de ces périodes trop courtes. Ne traitant que de périodes de longueurs plus grandes que u_n (proportionnel à $n/\log n$ comme dans les hypothèses (\mathbf{H}_i)), nous prenons K_{max} égal à $[n/u_n]$ dans la procédure décrite ci-dessous.

L'heuristique de pente utilise la linéarité (locale) du critère de QV en fonction du nombre de ruptures et fournit une calibration $\hat{\kappa}_n$ du paramètre de régularisation :

1. Pour tout $1 \leq K \leq K_{max}$, nous traçons $(K, -\min_{\underline{t}, \underline{\theta}} \hat{J}_n)_{1 \leq K \leq K_{max}}$ et estimons la pente de la partie linéaire qui vaut $\hat{\kappa}_n/2$.
2. Fixant $\kappa_n = \hat{\kappa}_n$, nous calculons le minimum du critère pénalisé \tilde{J}_n qui fournit une estimation du nombre de ruptures \hat{K}_n , des instants de ruptures \hat{t}_n et des paramètres du modèle $\hat{\theta}_n$.

1.7 Application sur données réelles

Nous appliquons la procédure décrite ci-dessus de détection de ruptures avec critère de QV pénalisée et calibration par l'heuristique de pente sur l'indice boursier anglais du Footsee. Nous analysons une trajectoire de $n = 1428$ données journalières du 27 Juillet 2005 au 18 Mars 2011. Nous étudions les log-rendements des prix à la clôture modélisés par des modèles GARCH(1, 1) [10] :

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0^*(j) + \alpha_1^*(j) X_{t-1}^2 + \beta_1^*(j) \sigma_{t-1}^2 \quad \forall t \in T_j^*, \quad \forall j \in \{1, \dots, K^*\}. \quad (5)$$

Ici $\theta_j^* = (\alpha_0^*(j), \alpha_1^*(j), \beta_1^*(j))$ et $d = 3$. Nous fixons à priori $u_n = [n/(4 * \log(n))] = 49$ d'où le nombre de ruptures maximum $K_{max} = 25$. L'heuristique de pente calibre le paramètre de régularisation à $\hat{\kappa}_n \approx 15$ et la QV pénalisée fournit $\hat{K} = 4$ soit trois ruptures $\hat{t}_1 = 499$, $\hat{t}_2 = 792$ et $\hat{t}_3 = 853$. Les ruptures semblent provenir d'événements exogènes identifiables :

$\hat{t}_1 = 499$, correspond au 16 juillet 2007, semaine durant laquelle la banque d'investissement Bear Stearns a révélé que 2 fonds sur les subprimes perdaient toute valeur.

$\hat{t}_2 = 792$, correspond au 11 septembre 2008 alors que le 15 septembre Lehman Brothers se plaçait en faillite.

$\hat{t}_3 = 853$, correspond au 5 décembre 2008, après que le G-20 ait annoncé qu'il prenait des mesures pour faire face à une crise sans précédent.

Ces événements significatifs ont déjà été identifiés dans [29] par une méthode de détection de ruptures complètement différente et moins précise que la notre.

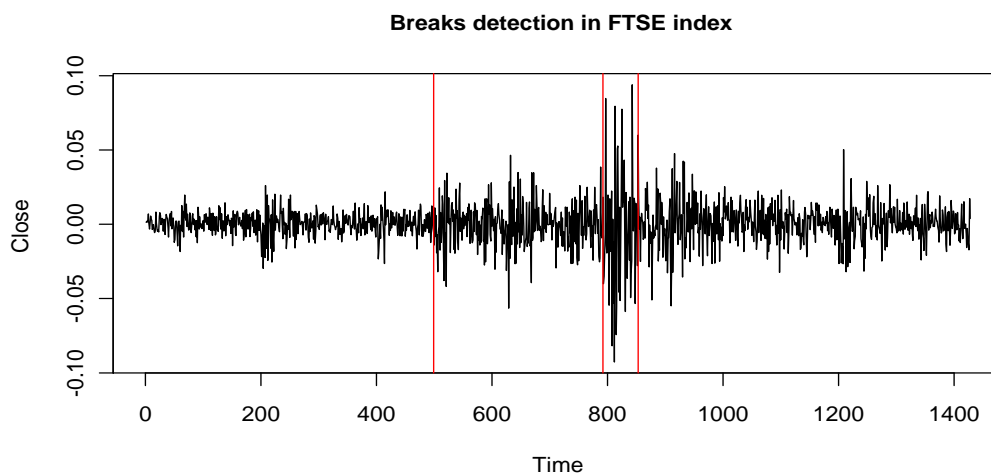


FIGURE 1 – L'indice Footsee avec les estimations des 3 instants de ruptures représentées par les lignes verticales.

2 L'estimateur du QMV dans le modèle EGARCH

L'approche d'inversibilité continue de [W11] est le sujet de la thèse de S. Cai que je co-encadre avec P. Doukhan et J.-L. Prigent. J'ai introduit cette notion pour étudier la stabilité aux conditions initiales du QMV dans le modèle EGARCH. Avec C. Francq et J.-M. Zakoïan, j'étudie dans [W16] le modèle log-GARCH comme alternative au modèle EGARCH pour lequel le QMV est stable.

2.1 Les modèles à volatilité stochastique

Nous considérons les modèles affines (1) tels que $f_\theta = 0$. Désormais $X_t = \sigma_t Z_t$ où seule la volatilité σ_t^2 dépend d'un coefficient $\theta_0 \in \mathbb{R}^d$ fixe au cours du temps (les ruptures de stationnarité ne sont plus envisagées). La dynamique du modèle porte sur une fonction h de cette volatilité : il existe des fonctions aléatoires ψ_t dépendantes des innovations iid Z_t telles que

$$h(\sigma_{t+1}^2) = \psi_t(h(\sigma_t^2), \theta_0), \quad \forall t \in \mathbb{Z}. \quad (6)$$

Le but est d'étendre dans un cadre le plus général possible les résultats asymptotiques pour l'estimateur du QMV du modèle GARCH univarié obtenus par Berkes *et al.* [8]. Une attention particulière est portée au modèle EGARCH :

Exemple 2.1 ([48]). *Le modèle EGARCH(1,1) est défini par la relation*

$$\text{EGARCH}(1,1) : \quad \log(\sigma_{t+1}^2) = \alpha_0 + \beta_0 \log(\sigma_t^2) + (\gamma_0 Z_t + \delta_0 |Z_t|). \quad (7)$$

On reconnaît $h = \log$ et les paramètres $\theta_0 = (\alpha_0, \beta_0, \gamma_0, \delta_0) \in \mathbb{R}^d$ avec $d = 4$.

Supposons désormais les conditions d'existence d'une solution stationnaire satisfaites. Pour le modèle EGARCH(1,1), cette condition s'écrit simplement $|\beta_0| < 1$.

2.2 Inversibilité du modèle

Soit, comme précédemment, l'estimateur du QMV qui minimise le critère

$$2\widehat{L}_n(\theta) = \sum_{t=1}^n \frac{X_t^2}{\widehat{\sigma}_t^2(\theta)} + \log(\widehat{\sigma}_t^2(\theta))$$

où $\widehat{\sigma}_t^2(\theta)$ est une approximation de σ_t^2 (non observée). Pour construire une telle approximation, on remplace dans le modèle (6) les innovations Z_t non observées par X_t/σ_t . On obtient à partir de (6) une équation récurrente stochastique sur $h(\sigma_t^2)$ notée

$$h(\sigma_{t+1}^2) = \phi_t(h(\sigma_t^2)), \quad \forall t \geq 0. \quad (8)$$

Désormais ϕ_t dépend des observations X_t . Supposons que h soit une bijection et notons $\ell = h^{-1}$. Une approximation de la volatilité σ_t^2 est obtenue par $\ell(\widehat{g}_t)$ où \widehat{g}_t est issue de la récurrence $\widehat{g}_{i+1} = \phi_i(\widehat{g}_i)$ pour $1 \leq i \leq t-1$ à partir d'une valeur initiale \widehat{g}_1 .

Définition 2.2. *Le modèle est inversible s'il existe un choix de \widehat{g}_1 tel que $\|\widehat{g}_t - h(\sigma_t^2)\| \rightarrow 0$ p.s. lorsque t tend vers l'infini.*

Les propriétés asymptotiques du QMV dans le modèle EGARCH(1,1) sont établies lorsque $\beta_0 = 0$ dans [53]. Cette condition est restrictive comparée aux résultats obtenus dans les autres modèles GARCH univariés [28]. La stabilité du critère QV est liée au problème d'inversibilité du modèle EGARCH(1,1) pour lequel on écrit

$$\phi_t(x) = \alpha_0 + \beta_0 x + (\gamma_0 X_t + \delta_0 |X_t|) \exp(-x/2). \quad (9)$$

La présence de l'exponentielle dans l'équation récurrente (9) rend cette équation récurrente stochastique instable. Le choix de la valeur initiale \widehat{g}_1 peut poser problème. L'inversibilité du modèle est assurée lorsque la loi asymptotique de \widehat{g}_t ne dépend pas de la condition initiale \widehat{g}_1 . Cette propriété de stabilité est satisfaite sous les conditions générales sur l'équation récurrente d'inversibilité (8) que nous rappelons dans un cadre général (nous avons besoin des versions fonctionnelles de ce résultat dans la suite).

Proposition 2.3. *Supposons une équation récurrente $X_t = \phi_t(X_{t-1})$ où ϕ_t est une suite stationnaire ergodique de fonctions $\Lambda(\phi_t)$ -Lipschitz sur un espace métrique complet (E, d) . Supposons qu'il existe $r \geq 1$ tel que $\mathbb{E}[\log(\Lambda(\phi_0 \circ \dots \circ \phi_{-r}))] < 0$, $\mathbb{E}[\log^+(\Lambda(\phi_0))] < \infty$ et $\mathbb{E}[\log^+(d(\phi_0(x_0, x_0)))] < \infty$ pour $x_0 \in E$. Il existe alors une unique solution stationnaire (X_t) ergodique non anticipative. De plus, pour toute solution $\widehat{X}_t = \phi_t(\widehat{X}_{t-1})$ avec une valeur initiale \widehat{X}_0 quelconque, il existe $c < 0$ tel que $\overline{\lim}_{t \rightarrow \infty} t^{-1} \log(d(X_t, \widehat{X}_t)) \leq c$ p.s.*

L'existence de la solution a été prouvée par Elton [23]. Le second résultat, dû à Bougerol [13], assure la stabilité de l'équation récurrente stochastique vis à vis du choix de la condition initiale. Straumann et Mikosch applique ce résultat au modèle EGARCH(1,1) dans [53]. Du fait de l'exponentielle dans la dynamique (9), ce modèle est inversible localement seulement lorsque le paramètre θ_0 vérifie la condition

$$\text{INV}(\theta_0) : \delta_0 \geq |\gamma_0| \text{ et}$$

$$\mathbb{E}[\log(\max\{\beta, 2^{-1}(\gamma_0 X_0 + \delta_0 |X_0|) \exp(-2^{-1}\alpha_0/(1-\beta_0)) - \beta_0\})] < 0. \quad (10)$$

2.3 Modèles continûment inversibles

L'inversibilité du modèle est utile pour approcher la volatilité (non observée) dans le critère QMV. Toutefois, l'équation récurrente d'invertibilité (8) nécessite la connaissance du paramètre θ_0 . Pour θ quelconque, nous calculons $\hat{\sigma}_t^2(\theta) = \ell(\hat{g}_t(\theta))$ par récurrence

$$(\hat{g}_{t+1}(\theta)) = \phi_t(\hat{g}_t(\theta), \theta), \quad \forall t \geq 1, \quad (11)$$

avec $\hat{g}_0(\theta)$ choisie arbitrairement. Notons $(g_t(\theta))$ la solution stationnaire de cette équation lorsqu'elle existe pour tout $\theta \in \Theta$.

Considérons l'équation récurrente stochastique (11) sous sa forme fonctionnelle et supposons les fonctions ϕ_t et \hat{g}_0 continues sur le compact Θ . Par récurrence, les fonctions $\theta \rightarrow \hat{g}_t(\theta)$ sont continues pour tout $t > 0$. Soit $\|\cdot\|_{\Theta}$ la norme uniforme, nous définissons la condition d'inversibilité uniforme :

Définition 2.4. *Le modèle est continûment inversible sur Θ si $(g_t(\theta))$ existe en tout point $\theta \in \Theta$ et si $\|\hat{g}_t - g_t\|_{\Theta} \rightarrow 0$ p.s. lorsque t tend vers l'infini.*

La condition d'invertibilité uniforme n'est pas explicite. Des conditions suffisantes sont les conditions de Lipschitz uniforme (**A**) et de dépendance faible (**H_i**) du chapitre précédent. Dans le cadre d'une équation récurrente stochastique $\phi_t(\cdot, \theta)$ uniformément Lipschitz sur Θ , Straumann et Mikosch ont obtenu dans [53] la condition suffisante d'uniforme inversibilité $\mathbb{E}[\log(\|\Lambda(\phi_0 \circ \dots \circ \phi_{-r})\|_{\Theta})] < 0$ par application de la proposition 2.3 dans l'espace des fonctions continues sur Θ . Cette condition permet de traiter le modèle EGARCH(1,1) sous la contrainte $\beta = 0$ pour tout $\theta \in \Theta$.

Nous introduisons la condition suffisante d'inversibilité continue plus faible suivante :

(CI) Les coefficients de Lipschitz $\Lambda(f(\cdot, \theta))$ sont continus pour toute fonction f de (x, θ) à valeur dans \mathbb{R} , $\mathbb{E}[\log(\Lambda(\phi_0 \circ \dots \circ \phi_{-r})(\theta))] < 0$ pour $r > 0$ et tout $\theta \in \Theta$. De plus, $\mathbb{E}[\log^+ \|\Lambda_0\|_{\Theta}] < \infty$ et $\mathbb{E}[\log^+(\|\phi_0(0, \theta)\|_{\Theta})] < \infty$.

En utilisant la convergence uniforme locale au voisinage de tout point $\theta \in \Theta$ de l'équation récurrente stochastique (11), nous obtenons

Théorème 2.5 ([W11]). *La condition (CI) est suffisante pour que le modèle soit continûment inversible sur Θ . De plus, (CI) est satisfaite pour le modèle EGARCH(1,1) sur tout compact Θ vérifiant la condition d'inversibilité INV(θ) en tout point $\theta \in \Theta$.*

Grâce à la condition d'inversibilité continue (**CI**), nous traitons l'estimation du QMV dans le modèle EGARCH pour des cas où $\beta_0 \neq 0$.

2.4 Résultats asymptotiques pour des modèles continûment inversibles

Nous prouvons la convergence forte de l'estimateur du QMV :

Théorème 2.6 ([W11]). *Supposons que le modèle soit identifiable tel que $\sigma_0^2(\theta) > m$ p.s. pour $m > 0$ et tout $\theta \in \Theta$, continûment inversible en Θ et admette une solution stationnaire et ergodique en $\theta_0 \in \Theta$. Alors $\hat{\theta}_n$ converge p.s. vers θ_0 .*

Considérons désormais la prévision de la volatilité σ_{n+1}^2 issue de l'équation récurrente d'inversibilité (11) : $\hat{\sigma}_n^2 = \ell(\hat{g}_t(\hat{\theta}_n))$ pour une condition initiale \hat{g}_0 choisie arbitrairement. Afin que cette procédure soit fortement consistante, nous supposons que la condition **(CI)** est satisfaite. On a alors

Théorème 2.7 ([W11]). *Sous les conditions du Théorème 2.6 et si de plus la condition **(CI)** est satisfaite alors $|\hat{\sigma}_n^2 - \sigma_n^2| \rightarrow 0$ p.s. lorsque n tend vers l'infini.*

Sans inversibilité du modèle, les propriétés de la prévision $\hat{\sigma}_n^2$ sont radicalement différentes. Comme l'équation de récurrence (11) n'est plus stable, $\hat{\sigma}_n^2$ dépend asymptotiquement et de manière chaotique de la valeur initiale \hat{g}_0 [54].

2.5 Normalité asymptotique dans le modèle EGARCH

La normalité asymptotique découle d'une propriété de continuité sur les dérivées secondes de la solution stationnaire (g_t) de l'équation récurrente stochastique (11). On considère désormais que les fonctions ϕ_t sont deux fois continûment dérivables sur un voisinage de $\theta_0 \in \overset{\circ}{\Theta}$, l'intérieur de Θ . Grâce à l'application successive de la proposition 2.3, nous prouvons que les dérivées premières et secondes de (g_t) sont continues. Supposons $\mathbb{E}[Z_0^4] < \infty$ et soit $\Sigma = (\mathbb{E}[Z_0^4] - 1)\mathbf{I}^{-1}$ où \mathbf{I} est la matrice de variance-covariance du vecteur $\nabla \log \sigma_0^2(\theta_0)$:

$$\mathbf{I} = \mathbb{E}[\nabla \log \sigma_0^2(\theta_0) \nabla \log \sigma_0^2(\theta_0)'].$$

Nous obtenons la normalité asymptotique de l'estimateur du QMV pour le modèle EGARCH(1,1) dans des cas où $\beta_0 \neq 0$.

Théorème 2.8 ([W11]). *Pour tout compact Θ tel que $INV(\theta)$ soit satisfaite pour $\theta \in \Theta$ et si $\theta_0 \in \Theta$ alors l'estimateur du QMV est fortement convergent. Si de plus $\mathbb{E}[Z_0^4] < \infty$, $\mathbb{E}[(\beta_0 - 2^{-1}(\gamma_0 Z_0 + \delta_0 |Z_0|))^2] < 1$ et Z_0 est non singulière par rapport à la mesure de Lebesgue alors il est asymptotiquement normal en $\theta_0 \in \overset{\circ}{\Theta}$.*

L'hypothèse supplémentaire que la densité de Z_0 soit non singulière par rapport à la mesure de Lebesgue induit des propriétés de mélange fort à taux géométriques sur le processus (X_t). L'estimateur du QMV est utilisé en pratique dans le modèle EGARCH(1,1) sans se restreindre à des paramètres θ dans le domaine d'inversibilité $INV(\theta)$. La procédure n'est alors pas stable et dépend asymptotiquement du choix arbitraire de la valeur initiale $\hat{g}_1(\theta) = \log \hat{\sigma}_1^2(\theta)$ dans la procédure d'estimation. Nous comparons les performances de l'estimateur du QMV sur simulations dans un cas inversible INV , un cas non

TABLE 1 – Performances du QMV sur simulations avec innovations gaussiennes dans le cas continûment inversible INV $\theta = (-0.399, 0.7, -0.4, 0.5)$ et non inversibles NI1 $\theta = (-0.99, 0.978, -0.45, 0.4)$ et NI2 $\theta = (-0.5, 0.5, -0.5, 10)$. La variance asymptotique $\Sigma_{i,i}$ est fournie explicitement via le théorème 2.8 si possible. Les résultats conformes à la théorie sont en gras et les résultats non conformes sont soulignés.

	$n =$	512		1024		2048	
INV	$\Sigma_{i,i}$	Mean	Var*512	Mean	Var*1024	Mean	Var*2048
-0.399	2.735	-.388	3.068	-.394	3.410	-0.399	2.714
.7	1.572	.683	3.258	.692	1.829	.699	1.576
-.4	1.603	-.405	1.784	-.404	2.022	-.400	1.504
.5	4.138	.482	4.196	.495	5.044	.500	4.097
NI1	$\Sigma_{i,i}$	Mean	Var*512	Mean	Var*1024	Mean	Var*2048
<u>-.099</u>	<u>1.771</u>	-.576	571.183	-.426	865.570	<u>-.153</u>	<u>139.037</u>
<u>.978</u>	<u>.119</u>	.817	63.947	.870	93.417	.961	<u>14.647</u>
<u>-.045</u>	.212	-.0643	1.90	<u>-.050</u>	.742	<u>-.051</u>	.375
<u>.040</u>	.459	.035	2.381	.044	1.810	.041	.861
NI2	$\Sigma_{i,i}$	Mean	Var*512	Mean	Var*1024	Mean	Var*2048
<u>-.5</u>	NA	10.661	1.670*10 ⁵	11.706	3.965*10 ⁵	<u>11.399</u>	<u>7.809*10⁵</u>
<u>.5</u>	NA	.530	8.463*10 ²	.534	1.812*10 ²	<u>.567</u>	<u>3.479*10²</u>
<u>-.5</u>	NA	-2.934	3.998*10 ⁵	-2.379	4.779*10 ⁵	<u>1.524</u>	<u>3.974*10⁶</u>
<u>10</u>	NA	3.543	3.736*10 ⁵	19.177	5.038*10 ⁷	<u>6.647</u>	<u>3.941*10⁶</u>

inversible traité dans [14] comme si la procédure était stable et un cas non inversible complètement instable NI2 [54].

Les deux cas non inversibles NI1 et NI2 sont distincts, le premier vérifiant la contrainte d'inversibilité (10) mais pas $\delta \geq |\gamma|$ alors que le second ne vérifie pas (10). Seuls les résultats du cas inversible INV sont conformes aux résultats de normalité asymptotique obtenus dans le théorème 2.8.

Afin de rendre la procédure d'estimation stable, nous préconisons de minimiser le critère de QV \hat{L}_n sur un ensemble Θ vérifiant $\delta \geq |\gamma|$ et la contrainte empirique d'inversibilité

$$\sum_{t=1}^n \log(\beta, 2^{-1}(\gamma X_t + \delta |X_t|) \exp(-2^{-1}\alpha/(1-\beta)) - \beta) < 0.$$

La procédure est stable car elle ne dépend plus asymptotiquement du choix de la valeur initiale $\hat{\sigma}_1^2(\theta)$. De plus, l'estimateur obtenu a les mêmes propriétés asymptotiques que celles décrites dans le théorème 2.8 lorsque θ vérifie INV(θ).

2.6 Le modèle log-GARCH

Le modèle EGARCH est très populaire du fait de son asymétrie : une valeur négative de l'innovation Z_t n'a pas le même impact sur la volatilité qu'une valeur positive lorsque $\gamma \neq 0$. Toutefois, nous avons vu que l'inférence statistique du modèle EGARCH par

QMV n'est pas toujours stable. Avec C. Francq et J.-M. Zakoian nous étudions comme alternative au modèle EGARCH(1,1) le modèle log-GARCH(1,1)

$$\log \sigma_{t+1}^2 = \omega_0 + \beta_0 \log \sigma_t^2 + (\alpha_{0+} 1_{\{X_t > 0\}} + \alpha_{0-} 1_{\{X_t < 0\}}) \log X_t^2.$$

Ce modèle a de nombreuses propriétés similaires à celles du modèle EGARCH(1,1) comme l'asymétrie quand $\alpha_{0+} \neq \alpha_{0-}$. Comme $\log X_t^2 = \log \sigma_t^2 + \log Z_t^2$, le modèle admet une solution stationnaire dès que $|\beta_0 + \alpha_{0+}| \vee |\beta_0 + \alpha_{0-}| < 1$ et $\mathbb{E} \log^+ |\log \eta_0^2| < \infty$. Par construction, le modèle est inversible dès que $|\beta_0| < 1$.

Les deux modèles ont toutefois une différence notable : contrairement au modèle EGARCH(1,1) continûment inversible, la volatilité (σ_t^2) du modèle log-GARCH(1,1) n'est pas bornée inférieurement. Les petites valeurs de la volatilité persistent dans le temps. Nous ne pouvons pas appliquer directement le théorème de consistance 2.6 nécessitant une borne inférieure sur la volatilité. Toutefois, sous des conditions de type Cramer sur le logarithme des innovations $|\log Z_0^2|$, les petites valeurs n'interfèrent pas dans l'estimation par QMV : nous prouvons que la procédure statistique est fiable et que l'estimateur est asymptotiquement normal.

Deuxième partie

La prédiction de séries Γ -faiblement dépendantes

Sommaire

3	Apprentissage statistique et transport de lois faiblement dépendantes	23
3.1	Le cadre d'apprentissage statistique	23
3.2	Inégalités de transport faible	25
3.3	La Γ -faible dépendance	26
3.4	Transport faible pour les lois Γ -faiblement dépendantes	28
3.5	Inégalités exponentielles	28
4	Inégalités d'oracle	29
4.1	Transport faible conditionnel	29
4.2	Inégalités d'oracle exactes à vitesses rapides	31
4.3	Inégalités d'oracle non exactes à vitesses rapides	32
4.4	Inégalités d'oracle exactes à vitesses lentes	32
4.5	Etude de Monte-Carlo dans le cas de prévisions linéaires	33

Cette partie regroupe les principaux résultats obtenus dans les travaux personnels [W6, W14] et en collaboration avec P. Alquier [W8, W13] sur la prédiction de processus faiblement dépendants du point de vue de l'apprentissage statistique. L'approche classique, présentée en iid par Massart dans [41] et étendue par Baraud *et al.* au cas dépendant [4], repose sur les propriétés gaussiennes de concentration de la loi des observations. Ayant suivi le GT EVOL organisé par F. Bolley, I. Gentil et C. Léonard sur le transport de lois, j'utilise une approche nouvelle de transport faible conditionnel [W14].

3 Apprentissage statistique et transport de lois faiblement dépendantes

3.1 Le cadre d'apprentissage statistique

Soient $(X_t)_{1 \leq t \leq n}$ les observations appartenant à $E = \mathbb{R}^k$, $k \geq 1$. Le but est de construire une prévision $\hat{X}_{n+1}^\theta = f_\theta(X_{n-1}, \dots, X_{n-p})$ de la valeur future X_{n+1} aussi bonne que possible. Nous utilisons comme modèle de prévisions $(f_\theta)_{\theta \in \Theta}$ un modèle linéaire généralisé. Cette terminologie nous permet de considérer aussi bien le cadre dit paramétrique (un modèle est fixé) que non paramétrique (aussi appelé de sélection de modèles). Une attention particulière est portée aux modèles linéaires simples

Exemple 3.1 (prévisions linéaires, $k = 1$). Ici $\theta = (\theta_0, \theta_1, \dots, \theta_p) \in \Theta \subset \mathbb{R}^{p+1}$ et

$$f_\theta(x_1, \dots, x_k) = \theta_0 + \sum_{j=1}^p \theta_j x_j.$$

Dans un cadre non paramétrique de sélection de modèles, l'ensemble des paramètres Θ peut être vu comme une réunion de modèles : $\Theta = \cup_{j=1}^m \Theta_j$. Les dimensions $d(\Theta_j)$ des modèles Θ_j peuvent être distinctes :

Exemple 3.2 (prévisions fonctionnelles, $k = 1$). Soient des fonctions notée φ_i de $\mathbb{R}^p \rightarrow \mathbb{R}$ convenablement choisies (sinusoïdes, noyaux, ondelettes,..). Alors $\theta = (\theta_1, \dots, \theta_\ell) \in \Theta_j \subset \mathbb{R}^j$ et

$$f_\theta(x_1, \dots, x_p) = \sum_{i=1}^j \theta_i \varphi_i(x_1, \dots, x_p).$$

J'ai utilisé cette approche non-paramétrique pour estimer la densité marginale de la loi stationnaire avec des estimateurs à noyaux dans [W1] et des estimateurs par ondelettes dans [W5].

Revenons au cadre de la prédiction. Afin de quantifier la prédiction de \hat{X}_{n+1}^θ , nous choisissons une fonction de perte ℓ comme par exemple la perte absolue $\ell(x, y) = |x - y|$ ou la perte quadratique $\ell(x, y) = (x - y)^2$. Le critère à minimiser est celui du risque $R(\theta) = \mathbb{E}[\ell(\hat{X}_n^\theta, X_n)]$ où \mathbb{E} est l'espérance sous la loi \mathbb{P} des observations $(X_t)_{1 \leq t \leq n}$ issues d'un processus stationnaire (X_t) . L'oracle $\bar{\theta}$ est le paramètre fournissant la meilleure prédiction possible dans le modèle : $R(\bar{\theta}) \leq \inf_\theta R(\theta) + c n^{-1}$ pour $c > 0$ petit. On

souhaite construire une procédure d'estimation $\hat{\theta}_n$ à partir des observations $(X_t)_{1 \leq t \leq n}$ (et non de R , \mathbb{E} ou \mathbb{P} non observés) qui vérifie une inégalité d'oracle

$$R(\hat{\theta}_n) \leq (1 + \eta)R(\bar{\theta}) + \Delta_n \eta^{-1} \eta \neq 0.$$

Cette inégalité d'oracle est vérifiée avec grande probabilité $1 - \varepsilon$ ($\hat{\theta}_n$ étant aléatoire, $\varepsilon > 0$ étant aussi petit que l'on veut). L'inégalité d'oracle est dite exacte lorsque $\eta = 0$ et non-exacte lorsque $\eta > 0$ [39]. La vitesse dans l'inégalité d'oracle est quantifiée par la suite de réels positifs (Δ_n) tendant vers 0. On parle de vitesse rapide lorsque $\Delta_n \propto n^{-1}$, lente lorsque $\Delta_n \propto n^{-1/2}$.

Décrivons succinctement comment on obtient des inégalités d'oracle à partir des propriétés gaussiennes de concentration dans le cadre paramétrique d'un modèle Θ fixé. L'estimateur est construit à partir du risque empirique

$$r_n(\theta) = \frac{1}{n} \sum_{t=p+1}^n \ell(\hat{X}_t^\theta, X_t).$$

Un estimateur naturel est le minimiseur du risque empirique $\hat{\theta}_{ERM}$ tel que $r_n(\hat{\theta}_{ERM}) \leq \inf_{\Theta} r_n(\theta) + cn^{-1}$ pour $c > 0$ petit. Sous des conditions de moments de \mathbb{P} , le théorème de la limite centrale gaussienne nous assure que r_n se concentre autour du risque R

$$\sqrt{n}(r_n(\theta) - r_n(\bar{\theta}) - (R(\theta) - R(\bar{\theta}))) \rightarrow \mathcal{N}(0, \text{var}(r_n(\theta) - r_n(\bar{\theta}))).$$

Pour une constante $C > 0$ à déterminer, on obtient avec grande probabilité l'inégalité

$$|r_n(\theta) - r_n(\bar{\theta}) - (R(\theta) - R(\bar{\theta}))| \leq C \sqrt{\text{var}(r_n(\theta) - r_n(\bar{\theta}))}.$$

Ceci implique que

$$\begin{aligned} R(\hat{\theta}_{ERM}) - R(\bar{\theta}) &\leq r_n(\hat{\theta}_{ERM}) - r_n(\bar{\theta}) + C \sqrt{\text{var}(r_n(\hat{\theta}_{ERM}) - r_n(\bar{\theta}))} \\ &\leq C \sqrt{\text{var}(r_n(\hat{\theta}_{ERM}) - r_n(\bar{\theta}))} \end{aligned}$$

car $r_n(\hat{\theta}_{ERM}) \leq r_n(\bar{\theta})$ par définition. En iid, comme

$$\text{var}(r_n(\theta) - r_n(\bar{\theta})) \leq \frac{\mathbb{E}(\ell(\hat{X}_t^\theta, X_t) - \ell(\hat{X}_t^{\bar{\theta}}, X_t))^2}{n},$$

dès que la perte ℓ est bornée on obtient une inégalité d'oracle exacte à vitesse lente. Sous une hypothèse de type Bernstein $\mathbb{E}(\ell(\hat{X}_t^\theta, X_t) - \ell(\hat{X}_t^{\bar{\theta}}, X_t))^2 \leq K(R(\theta) - R(\bar{\theta}))$ pour tout $\theta \in \Theta$, on obtient une inégalité de la forme

$$R(\hat{\theta}_{ERM}) - R(\bar{\theta}) \leq C \sqrt{K(R(\hat{\theta}_{ERM}) - R(\bar{\theta}))/n}.$$

On en déduit une inégalité d'oracle exacte à vitesse rapide $R(\hat{\theta}_{ERM}) - R(\bar{\theta}) \leq C^2 K/n$.

Les inégalités d'oracles découlent donc des propriétés gaussiennes de concentration du risque empirique $r_n(\theta)$ autour de sa moyenne $R(\theta)$ sous la loi \mathbb{P} . Dans le cas de données

faiblement dépendantes, la difficulté technique provient du contrôle du terme de variance qui entraîne d'ordinaire une perte logarithmique en vitesse [43, W6]. Nous traitons dans ce mémoire d'inégalités d'oracle dans un cadre faiblement dépendant obtenues grâce aux inégalités de transport faible [W14]. Le transport faible fournit pour des lois faiblement dépendantes des propriétés de concentration gaussiennes aussi précises qu'en iid. Les inégalités d'oracles obtenues ont des vitesses similaires au cas iid.

3.2 Inégalités de transport faible

Les observations (X_1, \dots, X_n) sont des vecteurs aléatoires de $E = \mathbb{R}^k$ muni d'une distance d , typiquement la distance euclidienne $\|x - y\|$ ou la distance de Hamming $1_{x \neq y}$. On note $\mathbb{E}_Q[h] = \int h dQ$ pour toute loi Q et toute fonction mesurable positive h . L'inégalité de transport classique est notée $T_{p,d}(C)$ pour tout $1 \leq p \leq 2$:

Définition 3.3 ([56, 32]). *La loi \mathbb{P} satisfait l'inégalité de transport $T_{p,d}(C)$ lorsque pour toute loi Q*

$$W_{p,d}(\mathbb{P}, Q) =: \inf_{\pi} \mathbb{E}_{\pi}[d^p(X, Y)]^{1/p} \leq \sqrt{2C\mathcal{K}(Q|\mathbb{P})}.$$

Ici π est un couplage de \mathbb{P} et Q , i.e. une loi sur le couple (X, Y) de marginales (\mathbb{P}, Q) . $\mathcal{K}(Q|\mathbb{P})$ est l'entropie relative aussi appelée divergence de Kullback-Leibler $\mathbb{E}_Q[\log(dQ/d\mathbb{P})]$. La distance entre lois de probabilité $W_{p,d}$ est appelée la distance de Wasserstein d'ordre p .

Remarque 3.4. Sur \mathbb{R} , la loi normale $\mathbb{P} = \mathcal{N}(\mu, \sigma^2)$ vérifie $T_{2,|\cdot|}(\sigma^2)$. L'égalité dans $T_{2,|\cdot|}(\sigma^2)$ est atteinte pour $Q = \mathcal{N}(\mu', \sigma^2)$. Autrement dit, le transport le plus difficile pour la gaussienne est atteint lorsqu'on change la moyenne de la loi. La gaussienne concentre bien autour de sa moyenne. Nous souhaitons reproduire ce type de propriétés gaussiennes à d'autres lois \mathbb{P} .

Les inégalités de transport $T_{p,d}(C)$ se généralisent sur des espaces produits E^n , c'est à dire pour la loi \mathbb{P} des observations $(X_t)_{1 \leq t \leq n}$, lorsque les X_t sont iid ou géométriquement faiblement dépendantes (voir Djellout *et al.* [21]). Nous affaiblissons la notion de transport pour pouvoir étendre ces inégalités à des processus sous-géométriquement faiblement dépendants. La première notion de transport faible a été introduite par Marton [40] et s'écrit

$$\inf_{\pi} \sup_{\alpha} \frac{\mathbb{E}_{\pi}[\alpha(Y)1_{X \neq Y}]}{\mathbb{E}_Q[\alpha(Y)^2]^{1/2}} \leq \sqrt{2C\mathcal{K}(Q|\mathbb{P})}.$$

où α est une fonction mesurable positive. C'est une notion affaiblie de $T_{2,1_{x \neq y}}(C)$ car

$$\sup_{\alpha} \frac{\mathbb{E}_{\pi}[\alpha(Y)d(X, Y)]}{\mathbb{E}_Q[\alpha(Y)^2]^{1/2}} = \mathbb{E}_Q[\pi(X \neq Y|Y)^2]^{1/2} \leq \sqrt{\pi(X \neq Y)}.$$

Ce transport faible a deux avantages : l'inégalité avec $C = 1$ est vérifiée pour toute loi \mathbb{P} [40] et son extension sur l'espace produit E^n est satisfaite pour des observations $(X_t)_{1 \leq t \leq n}$ sous-géométriquement mélangeantes [51]. Le choix de la distance de Hamming $d(x, y) = 1_{x \neq y}$ permet de traiter les processus faiblement dépendants uniformément mélangeants. Toutefois, nous verrons dans la section suivante que les processus classiques se traitent plus facilement en choisissant la distance issue de la norme euclidienne.

Par une application du théorème min-max de Sion on obtient

$$\inf_{\pi} \sup_{\alpha} \frac{\mathbb{E}_{\pi}[\alpha(Y)1_{X \neq Y}]}{\mathbb{E}_Q[\alpha(Y)^2]^{1/2}} = \sup_{\alpha} \inf_{\pi} \frac{\mathbb{E}_{\pi}[\alpha(Y)1_{X \neq Y}]}{\mathbb{E}_Q[\alpha(Y)^2]^{1/2}}.$$

D'où la définition du transport faible suivante, valable pour toute distance d :

Définition 3.5. La loi \mathbb{P} satisfait l'inégalité de transport faible $\tilde{T}_{p,d}(C)$ si pour tout Q

$$\sup_{\alpha} \inf_{\pi} \frac{\mathbb{E}_{\pi}[\alpha(Y)d(X, Y)]}{(\mathbb{E}_Q[\alpha(Y)^q])^{1/q}} \leq \sqrt{2C\mathcal{K}(Q|\mathbb{P})}$$

où $q \geq 2$ est tel que $1/p + 1/q = 1$.

Comme les rôles de \mathbb{P} et Q ne sont pas les mêmes, nous définissons de même l'inégalité $\tilde{T}_{p,d}^{(i)}(C)$ pour laquelle les rôles de \mathbb{P} et Q sont intervertis dans le membre de gauche.

Remarque 3.6. Dans le cas $p = 1$, alors $q = \infty$ et le dénominateur se lit comme étant le supremum essentiel de $\alpha(Y)$ et $\tilde{T}_{1,d}(C) = \tilde{T}_{1,d}^{(i)}(C)$ coïncide avec le transport classique $T_{1,d}(C)$. Pour $p = 2$, le transport faible est plus général que le transport classique. En effet, toute loi à support compact vérifie $\tilde{T}_{2,d}(C)$ alors qu'il existe des lois à support compact ne vérifiant pas le transport classique $T_{2,d}(C)$. D'autres exemples de lois satisfaisant le transport faible $\tilde{T}_{1,d}^{(i)}(C)$ sont fournies par Gozlan *et al.* dans [33].

Soit désormais \mathbb{P} la loi des observations $(X_t)_{1 \leq t \leq n}$ sur l'espace produit E^n . La loi \mathbb{P} vérifie l'inégalité de transport faible, notée aussi $T_{p,d}(C)$, lorsque pour toute loi Q

$$\sup_{\alpha} \inf_{\pi} \frac{\sum_{j=1}^n \mathbb{E}_{\pi}[\alpha_j(Y)d(X_j, Y_j)]}{(\sum_{j=1}^n \mathbb{E}_Q[\alpha_j(Y)^q])^{1/q}} \leq \sqrt{2C\mathcal{K}(Q|\mathbb{P})}.$$

3.3 La Γ -faible dépendance

Rajoutons des valeurs initiales artificielles $X_0 = Y_0 = x_0 = y_0 \in E$ et notons $\mathbb{P}_{|x^{(i)}}$ la loi conditionnelle de (X_{i+1}, \dots, X_n) sachant que $(X_i, \dots, X_0) = x^{(i)} := (x_i, \dots, x_0)$. La notion de Γ -faible dépendance impose des conditions sur un couplage trajectorien des X_{i+1}, \dots, X_n sachant que seule leur première valeur passée X_i est distincte :

Définition 3.7. Soient d_1 et d_2 deux distances vérifiant $d_1 \leq Md_2$ pour $M > 0$ et $(\gamma_{k,i}(p))_{1 \leq i < k \leq n}$ des coefficients positifs. La loi \mathbb{P} est $\Gamma_{d_1, d_2}(p)$ -faiblement dépendante si pour tout $1 \leq i \leq n$, pour tous $(x^{(i)}, y_i) \in E^{i+2}$, il existe un couplage π_i des lois $(\mathbb{P}_{|x^{(i)}}, \mathbb{P}_{|y_i, x^{(i-1)}})$ vérifiant

$$\mathbb{E}_{\pi_i}(d_1^p(X_k, Y_k))^{1/p} \leq \gamma_{k,i}(p) d_2(x_i, y_i), \quad \text{pour tout } i+1 \leq k \leq n.$$

Par extension, le processus (X_t) est dit Γ -faiblement dépendant lorsque la loi \mathbb{P} de $(X_t)_{1 \leq t \leq n}$ l'est quelque soit $n \geq 1$. Si le processus est stationnaire alors $\gamma_{k,i} = \gamma_{k',i'}$ dès que $k - i = k' - i'$.

Remarque 3.8. La dualité de Kantorovitch-Rubinstein s'applique dans le cas $p = 1$. En notant $\text{Lip}_1(i)$ l'ensemble des fonctions f vérifiant $|f(x) - f(y)| \leq \sum_{k=i+1}^n d_1(x_k, y_k)$ pour tout $x = (x_{i+1}, \dots, x_n)$ et tout $y = (y_{i+1}, \dots, y_n)$, on obtient

$$\sum_{k=i+1}^n \gamma_{k,i}(1) = \sup_{f \in \text{Lip}_1(i)} \sup_{x^{(i)}, y_i} \frac{|\mathbb{E}[f(X_{i+1}, \dots, X_n)|x^{(i)}] - \mathbb{E}[f(X_{i+1}, \dots, X_n)|y_i, x^{(i-1)}]|}{d_2(x_i, y_i)}.$$

Dans le cas d_1 quelconque et d_2 la distance de Hamming on retrouve les coefficients de Γ -dépendance faible introduits par Rio [50]. Dans le cas $d_1 = d_2$ quelconque on retrouve la condition (C1') de [21].

Remarque 3.9. Le cas où $d_1 = d_2$ est la distance de Hamming et $p = 2$ est traité par Samson dans [51] sous le nom de $\tilde{\Gamma}$ -faible dépendance. Nous gardons cette notation pour un tel choix de distance. Le couplage $\pi_{|i}$ est alors explicite : c'est le couplage maximal de Goldstein [31] En particulier, on a

$$\tilde{\gamma}_{k,i}(p) = \sup_{x^{(i)}, y_i} \|\mathbb{P}_{|x^{(i)}} - \mathbb{P}_{|y_i, x^{(i-1)}}\|_{TV}^{1/p}$$

où $\|\mathbb{P} - Q\|_{TV} = \sup_A |\mathbb{P}(A) - Q(A)|$ pour toutes lois \mathbb{P} et Q .

Même s'ils ne sont pas toujours explicites, les coefficients $\gamma_{k,i}(p)$ sont faciles à majorer dans de nombreux exemples. Si \mathbb{P} est Γ -faiblement dépendante, définissons $\Gamma(p)$ la matrice $n \times n$

$$\Gamma(p) = \begin{pmatrix} M & 0 & 0 & \dots & 0 \\ \gamma_{2,1}(p) & M & 0 & \dots & 0 \\ \gamma_{3,1}(p) & \gamma_{3,2}(p) & M & \ddots & \vdots \\ \vdots & \vdots & \ddots & M & 0 \\ \gamma_{n,1}(p) & \gamma_{n,2}(p) & \dots & \gamma_{n,n-1}(p) & M \end{pmatrix}.$$

Soit \mathbb{R}^n muni de la norme ℓ^p , $1 \leq p \leq 2$ et soit $\|\cdot\|_p$ la norme subordonnée sur les matrices de taille $n \times n$. Pour contrôler la quantité $\|\Gamma(p)\|_p$ qui apparaît dans les résultats de la section suivant, il suffit de contrôler les moments d'un couplage naturel, par exemple :

Exemple 3.10. *Considérons de nouveau l'équation récurrente stochastique de la forme*

$$X_0(x) := x \in E, \quad X_{t+1}(x) = \psi_{t+1}(X_t(x)), \quad t \geq 0, \quad (12)$$

où (ψ_t) est une suite de fonctions iid. Supposons l'existence d'une solution stationnaire (X_t) vérifiant (12) pour tout $t \in \mathbb{Z}$ et non anticipative (X_t indépendant de $(\psi_{t+1}, \psi_{t+2}, \dots)$ pour tout t). Pour tout $1 \leq p \leq 2$, s'il existe $S > 0$ tel que

$$\sum_{t=1}^{\infty} \mathbb{E}[d_1(X_t(x_0), X_t(y_0))^p]^{1/p} \leq S d_2(x_0, y_0) \quad \forall x_0, y_0 \in E. \quad (13)$$

alors (X_t) est $\Gamma_{d_1, d_2}(p)$ -faiblement dépendant et $\|\Gamma(p)\|_p \leq (M + S)$.

Cet exemple, appelé aussi fonctions itératives aléatoires, a déjà été traité par Dedecker et Priour dans [20] lorsque $d_2(x, y) = 1_{x \neq y}$. Les modèles ARMA et les modèles affines (1) (lorsqu'ils sont markoviens et à volatilité bornée) satisfont (13) pour la distance d_1 issue de la norme euclidienne. La condition (13) avec $d_1(x, y) = 1_{x \neq y}$ est équivalente à l'ergodicité uniforme de la chaîne de Markov. Cette condition est restrictive et exclut les cas gaussiens, les cas non-irréductibles, les cas linéaires non bornés,...

3.4 Transport faible pour les lois Γ -faiblement dépendantes

L'inégalité de transport faible s'étend aux lois Γ -faiblement dépendantes :

Théorème 3.11 ([W14]). *Si \mathbb{P} est $\Gamma_{d_1, d_2}(p)$ -faiblement dépendant et si $P_{x_j | x^{(j-1)}}$ satisfait $\tilde{T}_{p, d_2}(C)$ pour tout $1 \leq j \leq n$ alors \mathbb{P} vérifie $\tilde{T}_{p, d_1}(C \|\Gamma(p)\|_p^2 n^{2/p-1})$.*

Sous les mêmes conditions de dépendance faible une inégalité similaire est vérifiée pour le transport faible $\tilde{T}^{(i)}$ où les rôles de \mathbb{P} et Q sont inversés.

Remarque 3.12. Lorsque $d_1 = d_2$ est la distance de Hamming et $p = 2$ on retrouve l'inégalité obtenue par Samson [51]. Lorsque $d_1 = d_2$ est une distance quelconque et $p = 1$ on retrouve l'inégalité de Djellout *et al.* [21].

Exemple 3.13. *Pour le modèle (12) vérifiant (13), si la loi de $\psi_1(x)$ vérifie $\tilde{T}_{p, d_2}(C)$ pour tout $x \in E$ alors $P \in \tilde{T}_{p, d_1}(C(M + S)^2 n^{2/p-1})$.*

Les cas $p = 1$ et $p = 2$ sont très différents. Dans le premier cas, la constante dans l'inégalité de transport faible est multipliée par le nombre d'observations n alors que dans le second cas elle est indépendante de n . D'un point de vue statistique, le cas $p = 2$ permet d'atteindre des vitesses rapides dans les inégalités d'oracle. Le cas $p = 1$ est plus général mais ne traite que des inégalités d'oracle à vitesses lentes.

3.5 Inégalités exponentielles

Lorsque la loi \mathbb{P} vérifie $\tilde{T}_{p, d}(C)$ sur l'espace produit E^n alors elle vérifie des inégalités exponentielles et se concentre bien autour de sa moyenne. Considérons une fonction $f : E^n \mapsto \mathbb{R}$ telle qu'il existe des fonctions $L_j : E^n \mapsto \mathbb{R}_+$, $1 \leq j \leq n$ vérifiant

$$f(y) - f(x) \leq \sum_{j=1}^n L_j(y) d(x_j, y_j) \quad \forall x, y \in E^n. \quad (14)$$

Théorème 3.14 ([W14]). *Si \mathbb{P} vérifie $\tilde{T}_{p, d}(C)$ et f vérifie (14) alors pour tout $\lambda > 0$*

$$\mathbb{E} \left[\exp \left(\lambda (f - \mathbb{E}[f]) - \frac{C\lambda^2}{2} \left(\left(2 - \frac{2}{p}\right) \sum_{j=1}^n L_j^{\frac{p}{p-1}} + \left(\frac{2}{p} - 1\right) \right) \right) \right] \leq 1. \quad (15)$$

Des inégalités exponentielles similaires sont prouvées sous la condition de transport faible inverse $\tilde{T}_{p, d}^{(i)}(C)$. Dans le cas $p = 1$, d'après Bobkov et Gotze [9], le transport $T_{1, d}(C)$ est équivalent à l'inégalité (15) dite de Hoeffding. On retrouve ainsi l'inégalité de Hoeffding obtenue par Rio [50] en $\Gamma_{d_1, 1, x \neq y}(1)$ -dépendance faible. En effet, comme toute loi satisfait l'inégalité de Pinsker $T_{1, 1, x \neq y}(1/4)$, on obtient :

$$\mathbb{P}(f \geq \mathbb{E}[f] + t) \leq \exp \left(- \frac{2t^2}{n \|\Gamma(1)\|_1^2} \right), \quad \forall t > 0.$$

On obtient aussi l'extension au cas $\Gamma_{\|\cdot\|, 1, x \neq y}(2)$ -faiblement dépendant d'un résultat de Talagrand [55] que nous rappelons : soit \mathbb{P} iid, f convexe coordonnée par coordonnée

et L -Lipschitz sur $[0, 1]^n$ alors

$$\mathbb{P}(|f - \mathbb{E}[f]| \geq t) \leq 2 \exp\left(-\frac{t^2}{2L^2}\right), \quad \forall t > 0. \quad (16)$$

Ce résultat a été généralisé au cas des chaînes de Markov uniformément ergodiques par Marton [40] et à la $\tilde{\Gamma}(2)$ -dépendance faible par Samson [51] lorsque la fonction f est globalement convexe (non plus coordonnée par coordonnée). On obtient à partir de (15)

Corollaire 3.15. *Pour toute loi $\mathbb{P} \Gamma_{\|\cdot\|, 1_{x \neq y}}(2)$ -faiblement dépendante sur $[0, 1]^n$, pour toute fonction f convexe et L -Lipschitz on a*

$$\mathbb{P}(|f - \mathbb{E}[f]| \geq t) \leq 2 \exp\left(-\frac{t^2}{2\|\Gamma(2)\|_2^2 L^2}\right), \quad \forall t > 0.$$

Il est aussi possible d'étendre des inégalités exponentielles de Boucheron *et al.* [12].

Corollaire 3.16. *Soient un ensemble \mathcal{G} de fonctions g L -lipschitziennes et $f = \sup_{\mathcal{G}} \sum_{i=1}^n g^2(X_i)$. Si \mathbb{P} satisfait $\tilde{T}_{2, \|\cdot\|}(C)$ et $\tilde{T}_{2, \|\cdot\|}^{(i)}(C)$ alors pour tout $t \geq 0$ on a*

$$\begin{aligned} \mathbb{P}(f \geq \mathbb{E}[f] + t) &\leq \exp\left(-\frac{t^2}{8CL^2(\mathbb{E}[f] + t)}\right), \\ \mathbb{P}(f \leq \mathbb{E}[f] - t) &\leq \exp\left(-\frac{t^2}{8CL^2\mathbb{E}[f]}\right). \end{aligned}$$

Le cas de la distance de Hamming est particulier car, pour toute fonction h positive, on a $h(x) - h(y) \leq h(x)1_{x \neq y}$. Considérons $f(x) = |\sup_{\mathcal{G}} \sum_{i=1}^n g(X_i)|$ où $0 \leq g \leq M$ pour tout $g \in \mathcal{G}$. Alors f vérifie (14) avec $\sum_{j=1}^n L_j^2 \leq Mf$ lorsque $d_1 = d_2$ est la distance de Hamming. On retrouve le Théorème 2 de [51] comme corollaire du Théorème 3.14 :

Corollaire 3.17 ([51]). *Si \mathbb{P} est $\tilde{\Gamma}(2)$ -dépendante alors pour tout $t \geq 0$ on a*

$$\begin{aligned} \mathbb{P}(f \geq \mathbb{E}[f] + t) &\leq \exp\left(-\frac{t^2}{2M\|\tilde{\Gamma}(2)\|_2^2(\mathbb{E}[f] + t)}\right), \\ \mathbb{P}(f \leq \mathbb{E}[f] - t) &\leq \exp\left(-\frac{t^2}{2M\|\tilde{\Gamma}(2)\|_2^2\mathbb{E}[f]}\right). \end{aligned}$$

Ici les constantes sont optimales car dans le cas iid $\|\tilde{\Gamma}(2)\|_2^2 = 1$ [11].

4 Inégalités d'oracle

4.1 Transport faible conditionnel

Pour obtenir les inégalités d'oracle, j'utilise une nouvelle méthode inspirée de la méthode PAC-bayésienne [42, 15] dite de transport faible conditionnel. Supposons que l'excès de risque empirique $f_\theta = r_n(\theta) - r_n(\bar{\theta})$ vérifie une version affaiblie de la condition (14) : il existe des fonctions $L_j, L_j^{(i)} : E^n \mapsto \mathbb{R}_+$, $1 \leq j \leq n$ telles que

$$f_\theta(y) - f_\theta(x) \leq \sum_{j=1}^n (L_j(y, \theta) + L_j(x, \theta)^{(i)}) d(x_j, y_j) \quad \forall x, y \in E^n.$$

Si \mathbb{P} vérifie $T_{2,d}(C)$ et $T_{2,d}^{(i)}(C)$ alors pour toute loi Q et tout $\lambda > 0$ on a

$$\begin{aligned} \mathbb{E}_Q[\mathbb{E}_\mathbb{P}[f_\theta] - f_\theta] &\leq \inf_\pi \mathbb{E}_\pi \left[\sum_{j=1}^n (L_j(y, \theta) + L_j^{(i)}(x, \theta)) d(x_j, y_j) \right] \\ &\leq \left(\sqrt{\sum_{j=1}^n \mathbb{E}_Q[L_j(\theta)^2]} + \sqrt{\sum_{j=1}^n \mathbb{E}_\mathbb{P}[(L_j^{(i)}(\theta))^2]} \right) \sqrt{2C\mathcal{K}(Q|\mathbb{P})} \\ &\leq 2 \sqrt{\sum_{j=1}^n \mathbb{E}_Q[L_j(\theta)^2] + \sum_{j=1}^n \mathbb{E}_\mathbb{P}[(L_j^{(i)}(\theta))^2]} \sqrt{C\mathcal{K}(Q|\mathbb{P})} \\ &\leq \frac{C \left(\sum_{j=1}^n \mathbb{E}_Q[L_j(\theta)^2] + \sum_{j=1}^n \mathbb{E}_\mathbb{P}[(L_j^{(i)}(\theta))^2] \right)}{\lambda} + \lambda \mathcal{K}(Q|\mathbb{P}). \end{aligned}$$

Pour obtenir les inégalités d'oracle, il suffirait d'utiliser cette inégalité pour $\theta = \hat{\theta}$. C'est impossible car $\hat{\theta}$ dépend des observations et donc des intégrations en \mathbb{P} et Q . Considérons plutôt une version conditionnelle de cette inégalité de transport faible en appliquant l'inégalité précédente à $Q = Q_\theta$, une loi définie conditionnellement à θ . Intégrons cette inégalité conditionnelle par rapport à une loi ν indépendante de \mathbb{P} et telle que $\nu Q_\theta = Q\mu$, μ étant une loi dépendante de $\hat{\theta}$. En appliquant le théorème de Fubini sur \mathbb{P} et ν on obtient

$$\mathbb{E}_Q \mathbb{E}_\mu [\mathbb{E}_\mathbb{P}[f_\theta] - f_\theta] \leq \frac{C \left(\sum_{j=1}^n \mathbb{E}_Q \mathbb{E}_\mu [L_j(\theta)^2] + \sum_{j=1}^n \mathbb{E}_\mathbb{P} \mathbb{E}_\nu [(L_j^{(i)}(\theta))^2] \right)}{\lambda} + \lambda \mathbb{E}_\nu [\mathcal{K}(Q_\theta|\mathbb{P})].$$

On remarque que

$$\begin{aligned} \mathbb{E}_\nu [\mathcal{K}(Q_\theta|\mathbb{P})] &= \mathbb{E}_\nu \mathbb{E}_{Q_\theta} [\log(d\nu dQ_\theta / (d\nu d\mathbb{P}))] \\ &= \mathbb{E}_Q \mathbb{E}_\mu [\log(d\mu/d\nu) + \log(dQ/d\mathbb{P})] = \mathbb{E}_Q [\mathcal{K}(\mu|\nu)] + \mathcal{K}(Q|\mathbb{P}). \end{aligned}$$

On obtient finalement

$$\begin{aligned} \mathbb{E}_Q [\mathbb{E}_\mu [\mathbb{E}_\mathbb{P}[f_\theta] - f_\theta] - \lambda \mathcal{K}(\mu|\nu)] - \frac{C \left(\sum_{j=1}^n \mathbb{E}_Q \mathbb{E}_\mu [L_j(\theta)^2] + \sum_{j=1}^n \mathbb{E}_\mathbb{P} \mathbb{E}_\nu [(L_j^{(i)}(\theta))^2] \right)}{\lambda} \\ \leq \lambda \mathcal{K}(Q|\mathbb{P}). \end{aligned}$$

Dans cette inégalité, on choisit la loi Q égale à la loi \mathbb{P}_A , la loi \mathbb{P} restreinte à l'événement A "le terme de gauche soit supérieur à $\lambda \log(\varepsilon^{-1})$ ". On obtient

$$\lambda \log(\varepsilon^{-1}) \leq \lambda \mathcal{K}(\mathbb{P}_A|\mathbb{P}) = \lambda \log(\mathbb{P}(A)^{-1}).$$

Ainsi, avec grande probabilité $1 - \varepsilon$, le complémentaire de A est vérifié :

$$\begin{aligned} \mathbb{E}_\mu [R(\theta)] &\leq R(\bar{\theta}) + \mathbb{E}_{\mathbb{P}_A} [\mathbb{E}_\mu [f_\theta] + \lambda \mathcal{K}(\mu|\nu)] \\ &\quad + \frac{C \left(\sum_{j=1}^n \mathbb{E}_{\mathbb{P}_A} \mathbb{E}_\mu [L_j(\theta)^2] + \sum_{j=1}^n \mathbb{E}_\mathbb{P} \mathbb{E}_\nu [(L_j^{(i)}(\theta))^2] \right)}{\lambda} + \lambda \log(\varepsilon^{-1}) \quad (17) \end{aligned}$$

en utilisant l'identité $\mathbb{E}_\mathbb{P}[f_\theta] = R(\theta) - R(\bar{\theta})$. Finalement, on obtient des inégalités d'oracle en spécifiant convenablement f_θ , μ , ν et λ . Les fonctions L_j et $L_j^{(i)}$ déterminent le terme équivalent à la variance du cas gaussien. Elles dépendent de la distance d considérée.

4.2 Inégalités d'oracle exactes à vitesses rapides

Pour obtenir des inégalités d'oracle à vitesses rapides nous choisissons $\lambda \propto 1/n$. Comme dans le cas iid, nous allons travailler sous l'hypothèse de Bernstein $\mathbb{E}(\ell(\hat{X}_t^\theta, X_t) - \ell(\hat{X}_t^{\bar{\theta}}, X_t))^2 \leq K(R(\theta) - R(\bar{\theta}))$ pour tout $\theta \in \Theta$.

Remarque 4.1. Pour le risque quadratique $\ell(x, y) = (x - y)^2$, l'hypothèse de Bernstein est vérifiée dans le modèle linéaire (f_θ, Θ) s'il existe $K > 0$ telle que

$$K = \sup_{\theta \in \Theta} \frac{\|f_\theta(X_{t-1}, \dots, X_{t-k})\|_\infty}{\mathbb{E}[f_\theta^2(X_{t-1}, \dots, X_{t-k})]}, \quad p.s. \quad (18)$$

Afin de faire apparaître le terme de variance $\mathbb{E}(\ell(\hat{X}_t^\theta, X_t) - \ell(\hat{X}_t^{\bar{\theta}}, X_t))^2$ via le moment d'ordre 2 de f_θ , nous choisissons d comme étant la distance de Hamming telle que $\sum_{j=1}^n L_j(\theta)^2 \vee (L_j^{(i)}(\theta))^2 \leq (r_n(\theta) - r_n(\bar{\theta}))^2/n$. Ce choix nous restreint au cadre de la $\tilde{\Gamma}(2)$ -faible dépendance. Pour les lois μ et ν , deux choix sont possibles :

1. Soit nous considérons l'estimateur classique du risque empirique $\hat{\theta}_{ERM}$ pour un modèle Θ donné. Alors ν et μ sont fixés de telle sorte que $\mathcal{K}(\mu|\nu) \propto \ell(\hat{X}_t^{\hat{\theta}}, X_t) - \ell(X_t^{\bar{\theta}}, X_t)$. Pour le risque quadratique $\ell(x, y) = (x - y)^2$ on choisit μ et ν gaussiennes de même variance et centrées respectivement en $\hat{\theta}$ et $\bar{\theta}$. On obtient

Théorème 4.2 ([W14]). *Soit (X_t) un processus $\tilde{\Gamma}(2)$ -faiblement dépendant. Si le modèle (f_θ, Θ) vérifie (18) alors, avec probabilité $1 - \varepsilon$, on a pour tout $T > 0$*

$$R(\hat{\theta}_{ERM}) \leq R(\bar{\theta}) + C(K, B, T, \|\tilde{\Gamma}(2)\|, R(\bar{\theta})) \frac{d(\Theta) + \log(\varepsilon^{-1}) - \log \mathbb{P}(r_n(\bar{\theta}) > T)}{n}.$$

De telles inégalités ont été obtenues par Audibert et Catoni [2] dans le cas iid. Le terme explicite $-\log \mathbb{P}(r_n(\bar{\theta}) > T)$ dans la borne est obtenu grâce à la souplesse de l'approche par transport faible conditionnel.

2. Soit nous n'avons pas a priori d'estimateur convenable (par exemple nous avons plusieurs modèles disponibles $\Theta = \cup_{j=1}^m \Theta_j$). Choisissons alors ν comme étant le mélange de lois π_j sur chaque modèle Θ_j soit $\nu \propto \sum_{j=1}^m 2^{-d(\Theta_j)} \pi_j$. Puis minimisons le second terme dans la somme de l'inégalité (17) grâce à la dualité de l'entropie :

$$d\mu = \arg \min \mathbb{E}_\mu[f_\theta] + \lambda \mathcal{K}(\mu|\nu) = \frac{\exp(-\lambda^{-1} r_n(\theta))}{\mathbb{E}_\nu[\exp(-\lambda^{-1} r_n(\theta))]} d\nu.$$

Si la fonction de perte ℓ est convexe, alors $\mathbb{E}_\mu[R(\theta)] \geq R[\hat{\theta}_\lambda]$ où $\hat{\theta}_\lambda$ est l'estimateur de Gibbs de température λ défini par $\hat{\theta}_\lambda = \mathbb{E}_\mu[\theta]$. Définissons $\bar{\theta}_j$ comme étant l'oracle dans le modèle Θ_j et \bar{j} tel que $\bar{\theta} = \bar{\theta}_{\bar{j}}$. Nous obtenons

Théorème 4.3 ([W8]). *Soit (X_t) un processus $\tilde{\Gamma}(2)$ -faiblement dépendant bornés par T . Si les modèles (f_θ, Θ_j) sont tels que les diamètres des Θ_j soient bornés par $B > 0$ et les fonctions f_θ soient L -lipschitziennes pour tout $\theta \in \Theta$ et si la perte ℓ vérifie l'hypothèse de Bernstein alors, avec probabilité $1 - \varepsilon$, on a*

$$R(\hat{\theta}_{\lambda^*}) \leq R(\bar{\theta}) + C(K, B, T, L, \|\tilde{\Gamma}(2)\|) \frac{d(\Theta_{\bar{j}}) \log(n/d(\Theta_{\bar{j}})) + \log(\varepsilon^{-1})}{n}.$$

La valeur du paramètre de température λ^* proportionnelle à $1/n$ est connue en théorie mais dépend des valeurs T et $\|\tilde{\Gamma}(2)\|$ inconnues en pratique.

4.3 Inégalités d'oracle non exactes à vitesses rapides

Comme nous souhaitons obtenir des inégalités d'oracle à vitesses rapides, nous choisissons comme précédemment $\lambda \propto 1/n$. Désormais, l'hypothèse de Bernstein n'est plus nécessaire et nous pouvons choisir la distance issue de la norme euclidienne. Dans le cadre de la perte ℓ quadratique, comme $x^2 - y^2 \leq 2x(x - y) \leq 2|x||x - y|$ on obtient la relation $\sum_{j=1}^n L_j(\theta)^2 \vee (L_j^{(i)}(\theta))^2 \leq C r_n(\theta)/n$ pour $C > 0$ et tout $\theta \in \Theta$. En employant la même méthodologie que précédemment, on obtient une inégalité d'oracle non exacte valable pour des processus $\Gamma_{\|\cdot\|, \|\cdot\|}(2)$ -faiblement dépendants et non bornés (incluant par exemple les modèles ARMA gaussiens) :

Théorème 4.4 ([W14]). *Si \mathbb{P} vérifie $\tilde{T}_{2, \|\cdot\|}$ alors il existe des constantes B_1, B_2 et $B_3 > 0$ telles qu'avec probabilité $1 - \varepsilon$ on ait, pour tout $B_1 d(\Theta)/n < \eta < 1$,*

$$R(\hat{\theta}_{ERM}) \leq (1 + \eta)R(\bar{\theta}) + B_2 \frac{d + \log(\varepsilon^{-1})}{n\eta} + \frac{B_3}{(n\eta)^2}.$$

4.4 Inégalités d'oracle exactes à vitesses lentes

Toute inégalité d'oracle non exacte à vitesse rapide fournit une inégalité d'oracle exacte à vitesse lente en choisissant $\eta = 1/\sqrt{n}$. Toutefois, le cadre où les vitesses lentes sont atteignables est plus général : les propriétés gaussiennes de concentration avec un terme de variance borné suffisent. Pour toute fonction de perte ℓ 1-lipschitzienne on obtient, par le même raisonnement que précédemment,

$$R(\hat{\theta}_\lambda) \leq R(\bar{\theta}) + C(k, B, T, L, \|\Gamma(2)\|) \frac{d(\Theta_{\bar{j}}) \log(n/d(\Theta_{\bar{j}})) + \log(\varepsilon^{-1})}{\sqrt{n}}.$$

Toutefois, la vitesse en d/\sqrt{n} de cet estimateur de Gibbs, dont la température λ est indépendante de la dimension des modèles $d(\Theta_j)$, n'est pas optimale pour la perte absolue $\ell(x, y) = |x - y|$. Nous considérons dans [W8] l'estimateur de Gibbs par sélection de modèles $\tilde{\theta}_{\hat{p}, \hat{l}}^\lambda$ où

$$(\hat{p}, \hat{l}, \hat{\lambda}) = \arg \min_{p, l, \lambda} -\frac{1}{\lambda} \log \mathbb{E}_{\pi_{p, l}} \exp(-\lambda r_n(\theta)) + \frac{1}{\lambda} \log(n) + \frac{\lambda(1 + L)^2 \log^3(n)}{n},$$

$\pi_{p, l}$ étant la loi de Gibbs sur le modèle $\Theta_{p, l}$ et $\tilde{\theta}_{p, l}$ étant l'estimateur de Gibbs associé. Cette approche, plus difficile à mettre en oeuvre en pratique, nécessite moins de propriétés de transport faible ($p = 1$ et non $p = 2$) :

Théorème 4.5 ([W8]). *Si \mathbb{P} vérifie $\tilde{T}_{1, \|\cdot\|}$, si les modèles (f_θ, Θ_j) sont de diamètres bornés par $B > 0$ et si les fonctions f_θ et ℓ sont L -lipschitziennes alors, avec probabilité $1 - \varepsilon$, on a*

$$R(\hat{\theta}_\lambda) \leq R(\bar{\theta}) + C(k, B, T, L, \|\Gamma(1)\|) \frac{\sqrt{d(\Theta_{\bar{j}}) \log^{5/2}(n) + \log(\varepsilon^{-1})}}{\sqrt{n}}.$$

4.5 Etude de Monte-Carlo dans le cas de prévisions linéaires

L'estimateur de Gibbs $\hat{\theta}_\lambda = \mathbb{E}_\mu[\theta]$ est une intégrale sur un espace de grande dimension. Nous l'approchons dans [W13] par la méthode de Reversible Jump Markov Chain Monte Carlo développée dans [34]. On compare sur simulations le risque quadratique de l'estimateur de Gibbs $\hat{\theta}_\lambda$ avec l'approche classique de sélection de modèles AIC [1] lorsque $\Theta = \cup_{j=1}^{20} \Theta_j$ où les Θ_j sont des modèles AR(j), $1 \leq j \leq 20$. L'approche AIC consiste à calculer l'ERM $\hat{\theta}_{ERM,j}$ dans chaque sous modèle Θ_j puis de sélectionner j par le critère pénalisé en la dimension $j + 1$.

TABLE 2 – Risques estimés par Monte Carlo (et les écarts types estimés) pour les estimateurs de Gibbs, AIC et $\hat{\theta}_{ERM,20}$. En gras sont indiquées les meilleures performances.

n	Processus	Innovations	Gibbs	AIC	ERM
100	AR(2)	unif.	0.165 (0.022)	0.165 (0.023)	0.182 (0.029)
	Eqn. (19)	Gaussian	0.167 (0.023)	0.161 (0.023)	0.173 (0.027)
	AR(8)	unif.	0.163 (0.020)	0.169 (0.022)	0.178 (0.022)
	Eqn. (20)	Gaussian	0.172 (0.033)	0.179 (0.040)	0.201 (0.049)
	Non linéaire	unif.	0.174 (0.022)	0.179 (0.028)	0.201 (0.040)
	Eqn. (21)	Gaussian	0.179 (0.025)	0.182 (0.025)	0.202 (0.031)
1000	AR(2)	unif.	0.163 (0.005)	0.163 (0.005)	0.166 (0.005)
	Eqn. (19)	Gaussian	0.160 (0.005)	0.160 (0.005)	0.162 (0.005)
	AR(8)	unif.	0.164 (0.004)	0.166 (0.004)	0.167 (0.004)
	Eqn. (20)	Gaussian	0.160 (0.008)	0.161 (0.008)	0.163 (0.008)
	Non linéaire	unif.	0.171 (0.005)	0.172 (0.006)	0.175 (0.006)
	Eqn. (20)	Gaussian	0.173 (0.009)	0.173 (0.009)	0.176 (0.010)

Les observations sont générées selon les processus vérifiant, pour tout $t \in \mathbb{Z}$,

$$X_t = 0.5X_{t-1} + 0.1X_{t-2} + Z_t \quad (19)$$

$$X_t = 0.6X_{t-4} + 0.1X_{t-8} + Z_t \quad (20)$$

$$X_t = \cos(X_{t-1}) \sin(X_{t-2}) + Z_t \quad (21)$$

où les innovations Z_t suivent une loi uniforme $\mathcal{U}[-a, a]$ ou gaussienne $\mathcal{N}(0, \sigma^2)$. Les hypothèses des Théorèmes 4.2 et 4.3 sont satisfaites pour les simulations (19), (20) et (21) dans le cas uniforme mais pas dans le cas gaussien. Dans le cas gaussien, les données ne sont pas $\tilde{\Gamma}(2)$ mais $\Gamma(2)$ -faiblement dépendantes et seuls les résultats des théorèmes 4.4 et 4.5 s'appliquent. Bien que ce ne soit pas significatif, l'estimateur de Gibbs semble donner des résultats meilleurs que ceux donner par AIC lorsque le nombre d'observations n est petit ($= 100$), en particulier dans les modèles (20) et (21).

Troisième partie

Théorie asymptotique des agrégats de valeurs extrêmes

Sommaire

5	L'indice de <i>cluster</i>	35
5.1	Grandes déviations dans le cas iid	35
5.2	Les processus à variation régulière	35
5.3	Condition de dérive des chaînes de Markov	36
5.4	Définition de l'indice de <i>cluster</i>	37
5.5	Lien avec l'indice extrémal	38
5.6	Calcul de l'indice de <i>cluster</i> sur des exemples	39
6	Théorie asymptotique des sommes partielles dépendantes et à variation régulière	40
6.1	Théorème de la limite centrale gaussienne	40
6.2	Théorème de la limite centrale α -stable	40
6.3	Principe uniforme de grandes déviations	42
6.4	Exemples de processus à queues épaisses	42
6.5	Comportement asymptotique des auto-covariances empiriques	43
	Conclusion et perspectives	46
	Références	48

Lors de mon séjour post doctoral à l'Université de Copenhague avec T. Mikosch j'ai étudié la dépendance des valeurs extrêmes de processus faiblement dépendants. Notre collaboration a été soutenue financièrement durant 2 années consécutives par les services "Recherche" de la coopération Franco-Danoise. T. Mikosch a été professeur invité deux années consécutives (2011-2012) à l'université Paris Dauphine. Cette partie décrit quelques résultats des articles [W7, W10, W15], le premier étant écrit en collaboration avec A. Jakubowski et son étudiante en thèse K. Bartkiewicz.

5 L'indice de *cluster*

5.1 Grandes déviations dans le cas iid

Les distributions à queues épaisses modélisent des situations où les queues de distribution des sommes partielles sont caractérisées par les queues de distribution de quelques observations seulement. Dans le langage mathématique, cela correspond aux lois \mathbb{P} des observations (X_1, \dots, X_n) (supposées centrées lorsque l'espérance existe) telles que les grandes déviations des sommes partielles $S_n = \sum_{i=1}^n X_i$ de variables aléatoires (v.a.) sont proportionnelles à celles du maximum des observations :

$$\exists c > 0, \quad \mathbb{P}(S_n > x_n) \sim_{n \rightarrow \infty} c \mathbb{P}(\max_{1 \leq i \leq n} X_i > x_n). \quad (22)$$

Ici, (x_n) est une suite tendant vers l'infini suffisamment rapidement ; on étudie donc les grandes déviations des sommes partielles [25]. Dans le cadre classique iid, les distributions \mathbb{P} vérifiant (22) sont les lois à queues épaisses appelées lois sous exponentielles et la constante c vaut nécessairement 1 [26]. Autrement dit, les grandes déviations d'une somme de variables iid sont caractérisées par celles d'une seule observation extrême (le maximum des variables). Dans ce cadre iid, nous avons de plus l'équivalence $\mathbb{P}(\max_{1 \leq i \leq n} X_i > x_n) \sim n \mathbb{P}(X_0 > x_n)$. Un principe uniforme de grandes déviations s'en déduit, à savoir l'existence d'une suite $b_n \rightarrow \infty$ telle que

$$\lim_{n \rightarrow \infty} \sup_{x \geq b_n} \left| \frac{\mathbb{P}(S_n > x)}{n \mathbb{P}(X_0 > x)} - 1 \right| = 0. \quad (23)$$

La loi \mathbb{P} doit vérifier des conditions supplémentaires de régularité pour que la suite (b_n) soit explicite [17].

5.2 Les processus à variation régulière

Nous supposons dans la suite que la loi marginale de \mathbb{P} est à variation régulière.

Définition 5.1. *La loi marginale de \mathbb{P} est à variation régulière ssi il existe une fonction à variation lente L , telle que $L(ax) \sim_{x \rightarrow \infty} L(x)$ pour tout $a > 0$ fixé, un indice $\alpha > 0$ et deux réels p et q tels que $p + q = 1$ et*

$$\mathbb{P}(X_0 > x) \sim p x^{-\alpha} L(x) \quad \text{et} \quad \mathbb{P}(X_0 \leq -x) \sim q x^{-\alpha} L(x), \quad x \rightarrow \infty.$$

Les lois à variation régulière ont de nombreuses propriétés remarquables. En particulier, Feller [27] montre dans le cadre iid que seuls les v.a. suivant une loi à variation régulière d'indice $0 < \alpha < 2$ satisfont des théorèmes de la limite centrale α -stable. Les v.a. iid à variation régulière satisfont aussi un principe uniforme de grandes déviations :

Théorème 5.2 ([47]). *Si (X_t) est une suite de v.a. iid centrées (si elles sont intégrables) à variation régulière alors il existe une suite (b_n) telle que*

$$\lim_{n \rightarrow \infty} \sup_{x \geq b_n} \left| \frac{\mathbb{P}(S_n > x)}{n \mathbb{P}(|X_0| > x)} - p \right| = 0 \quad \text{et} \quad \lim_{n \rightarrow \infty} \sup_{x \geq b_n} \left| \frac{\mathbb{P}(S_n \leq -x)}{n \mathbb{P}(|X_0| > x)} - q \right| = 0. \quad (24)$$

Si $\alpha > 2$ on peut choisir $b_n = \sqrt{an \log n}$ où $a > \alpha - 2$. Si $\alpha \in (0, 2]$, $b_n = n^{\delta+1/\alpha}$ convient pour tout $\delta > 0$.

Ce résultat explicite la suite (b_n) dans le principe uniforme de grandes déviations (23) pour des lois à variation régulière, car $\mathbb{P}(|X_0| > x) \sim p\mathbb{P}(X_0 > x) \sim q\mathbb{P}(X_0 \leq -x)$ quand $x \rightarrow \infty$. Remarquons que ce principe de grandes déviations est très différent de celui étudié dans [W12] pour des queues de distribution légères.

La classe des v.a. X à variation régulière s'étend au cadre des processus stationnaires (X_t) . Remarquons que la définition 5.1 est équivalente à l'existence d'une variable $\Theta_0 \in \{\pm 1\}$ telle que $\mathbb{P}(\Theta_0 = 1) = p = 1 - q = 1 - \mathbb{P}(\Theta_0 = -1)$ et

$$\frac{\mathbb{P}(X_0 > ux, X_0/|X_0| = \pm 1)}{\mathbb{P}(|X_0| > x)} \xrightarrow{x \rightarrow \infty} u^{-\alpha} \mathbb{P}(\Theta_0 = \pm 1), \quad u > 0.$$

En généralisant cette remarque, Basrak et Segers [6] caractérisent la notion de variation régulière des processus stationnaires introduite par Davis et Hsing [18] de la manière suivante :

Définition 5.3. *Le processus (X_t) est à variation régulière d'indice $\alpha > 0$ si il existe un processus spectral de queue (Θ_t) vérifiant, pour tout $k \geq 1$, lorsque $x \rightarrow \infty$:*

$$\frac{\mathbb{P}(|X_0| > ux, |X_0|^{-1}(X_0, \dots, X_k) \in \cdot)}{\mathbb{P}(|X_0| > x)} \xrightarrow{d} u^{-\alpha} \mathbb{P}((\Theta_0, \dots, \Theta_k) \in \cdot). \quad (25)$$

Par définition $\Theta_0 \in \{\pm 1\}$. Remarquons que le processus (Θ_t) n'est pas forcément lui même stationnaire. Dans le cas indépendant, $\Theta_t = 0$ pour tout $t > 0$. Dans toute la suite nous considérons la condition de variation régulière :

Condition (\mathbf{RV}_α) : Le processus stationnaire (X_t) est à variation régulière d'indice $\alpha > 0$ et de processus spectral de queue (Θ_t) .

5.3 Condition de dérive des chaînes de Markov

Représentons désormais (X_t) comme étant une fonction f d'une chaîne de Markov $(\Phi_t) : X_t = f(\Phi_t)$. Cette représentation n'est pas restrictive car $\Phi_t = (X_t, X_{t-1}, X_{t-2}, \dots)$ est une chaîne de Markov. Nous supposons que la chaîne de Markov sous-jacente vérifie la condition de dérive suivante :

Condition (\mathbf{DC}_p) : Il existe des constantes $\beta \in (0, 1)$, $b > 0$, et une fonction $V : \mathbb{R}^d \rightarrow (0, \infty)$ telle que $c_1|x|^p \leq V(x) \leq c_2|x|^p$, $c_1, c_2 > 0$ et pour tout y ,

$$\mathbb{E}(V(f(\Phi_1)) \mid \Phi_0 = y) \leq \beta V(f(y)) + b.$$

De telles conditions de dérive sont classiques et très utiles, voir le livre référence de Meyn et Tweedie [45]. Lorsque la chaîne (Φ_t) est irréductible, nous supposons de plus qu'elle admet un petit ensemble C de la forme $C = \{x \mid V(f(x)) \leq M\}$ avec $M > 0$ (voir le livre de Meyn et Tweedie [45] pour des détails sur ces notions d'irréductibilité, petit ensemble, ...). Dans ce cadre assez général de chaîne (Φ_t) irréductible nous pouvons toujours supposer l'hypothèse plus forte que (Φ_t) possède un atome noté A . Pour cela, il suffit d'adjoindre à Φ_t une suite de variables de Bernoulli (B_t) en suivant le schéma d'extension des chaînes irréductibles dû à Nummelin [49]. Les observations X_t s'écrivent alors comme $g(\Phi_t, B_t)$ pour une fonction g issue de f où la chaîne élargie (Φ_t, B_t) possède un atome A . Nous simplifierons les notations en considérant toujours que Φ_t possède un atome A dans les cas d'irréductibilité. La condition (\mathbf{DC}_p) est alors équivalente à l'existence d'une constante $\beta \in (0, 1)$ telle que pour tout y ,

$$\mathbb{E}(V(f(\Phi_1)) \mid \Phi_0 = y) \leq \beta V(f(y)) + b\mathbb{1}_A(y).$$

Soit la suite de temps d'atteinte de l'atome A , i.e. $\tau_A(1) = \tau_A = \min\{k > 0 : \Phi_k \in A\}$ et $\tau_A(j+1) = \min\{k > \tau_A(j) : \Phi_k \in A\}$, $j \geq 1$. Le schéma de régénération de (Φ_t) consiste à scinder la chaîne en trajectoires d'excursions successives de l'atome A $(\Phi_{\tau_A(j)}, \dots, \Phi_{\tau_A(j+1)})$. Ces trajectoires, de longueurs variables aléatoires, sont iid. Soit les cycles de sommes partielles de taille aléatoire :

$$S(0) = \sum_{t=1}^{\tau_A} X_t \quad \text{and} \quad S(i) = \sum_{t=\tau_A(i)+1}^{\tau_A(i+1)} X_t, \quad i \geq 1. \quad (26)$$

Le schéma de régénération permet de représenter S_n comme une somme $\sum_{i=1}^{N_A(n)} S(i)$ de cycles iid et $N_A(n) = \max\{j : \tau_A(j) \leq n\}$ à deux termes près : le premier cycle $S(0)$ qui ne commence pas forcément dans l'atome et le dernier cycle qui ne finit pas forcément dans l'atome. Ce cadre de régénération est très utile mais restrictif : l'irréductibilité nécessite des propriétés topologiques sur l'espace d'état de Φ_t qui ne sont pas satisfaites dans le cadre général $\Phi_t = (X_t, X_{t-1}, X_{t-2}, \dots)$.

5.4 Définition de l'indice de cluster

Il est désormais temps de définir l'indice de *cluster*. Supposons que (X_t) vérifie (\mathbf{RV}_α) avec $\alpha > 0$ de sorte que la limite $b_+(k) = \lim_{n \rightarrow \infty} n \mathbb{P}(S_k > a_n)$ soit bien définie pour tout $k \geq 1$ et pour (a_n) telle que $\lim_{n \rightarrow \infty} n \mathbb{P}(|X| > a_n) = 1$. L'expression de la différence $b_+(k+1) - b_+(k)$ en fonction du processus spectral de queue est très simple :

Lemme 5.4. *Supposons que (X_t) vérifie (\mathbf{RV}_α) avec $\alpha > 0$ alors*

$$b_+(k+1) - b_+(k) = \mathbb{E} \left[\left(\sum_{t=0}^k \Theta_t \right)_+^\alpha - \left(\sum_{t=1}^k \Theta_t \right)_+^\alpha \right].$$

Nous définissons l'indice de *cluster* comme

Définition 5.5. *L'indice de cluster b_+ est défini comme étant la limite, si elle existe,*

$$b_+ = \lim_{k \rightarrow \infty} b_+(k+1) - b_+(k).$$

Quand la limite ci-dessus existe, le lemme de Césaro fournit une autre caractérisation de l'indice de *cluster* : $b_+ = \lim_{k \rightarrow \infty} k^{-1}b_+(k)$. Cette caractérisation montre que l'indice de *cluster* est forcément un réel positif. Nous fournissons un cadre général d'existence de cet indice et sa caractérisation en fonction du processus spectral de queue :

Théorème 5.6 ([W15]). *Supposons que $X_t = f(\Phi_t)$ vérifie (\mathbf{RV}_α) avec $\alpha > 0$ et que la chaîne de Markov (Φ_t) vérifie (\mathbf{DC}_p) pour un réel positif $p \in]\alpha - 1, \alpha[$. Alors l'indice de *cluster* existe et vaut*

$$b_+ = \mathbb{E} \left[\left(\sum_{t \geq 0} \Theta_t \right)_+^\alpha - \left(\sum_{t \geq 1} \Theta_t \right)_+^\alpha \right].$$

A. Jakubowski dans [35] a prouvé l'existence de cet indice dans les cas de mélange uniforme (où alors $b_+ = \mathbb{P}(\Theta_0 = 1)$ comme dans le cas iid) ou de corrélations maximales (sous des conditions de ρ -faible dépendance). Le rôle de la condition de dérive (\mathbf{DC}_p) pour un réel positif $p \in]\alpha - 1, \alpha[$ apparaît clairement dans la preuve : elle implique que la série des (Θ_t) est normalement convergente dans l'espace $\mathbb{L}^{\alpha-1}$ ce qui entraîne l'existence de b_+ grâce au théorème des valeurs intermédiaires. L'avantage de notre approche par rapport aux précédentes est que, pour la première fois, le cas de séries (Θ_t) divergentes dans \mathbb{L}^α peut être traité.

Nous allons voir que l'indice de *cluster* b_+ caractérise les grandes déviations positives. Par symétrie, on définit b_- sous les mêmes conditions que celles du théorème 5.6 par la relation

$$b_- = \mathbb{E} \left[\left(\sum_{t \geq 0} \Theta_t \right)_-^\alpha - \left(\sum_{t \geq 1} \Theta_t \right)_-^\alpha \right].$$

5.5 Lien avec l'indice extrémal

La valeur maximale d'un processus faiblement dépendantes, est souvent précédée et suivie de valeurs elles aussi très grandes. On parle alors de phénomène de *cluster* des valeurs extrêmes. Ce phénomène a été observé sur de très nombreuses données en finance, en sismologie, en météorologie, ... Des modèles ont été proposés pour reproduire ce phénomène de *cluster*. Toutefois, d'un point de vue théorique, ce phénomène de *cluster* est encore mal quantifié. Une première étape a été réalisée par Leadbetter [38] définissant l'indice extrémal. Sous des conditions générales Leadbetter obtient l'existence d'un coefficient $0 \leq \theta_+ \leq p$ et $\theta_+ \neq 0$ dès que $p \neq 0$ appelé indice extrémal et tel que

$$\mathbb{P}(\max_{1 \leq i \leq n} X_i > x_n) \sim \theta_+ n \mathbb{P}(|X| > x_n), \quad n \rightarrow \infty \text{ et la suite } x_n \uparrow \infty \text{ bien choisie.}$$

L'indice θ_+ quantifie donc la perte d'information fournit par le seul maximum pour les grandes déviations dans le cas faiblement dépendant. On définit de la même manière θ_- l'indice extrémal du processus $(-X_t)$. Leadbetter [38] a montré que la longueur du

cluster, i.e. le nombre de données consécutives au-dessus d'un niveau u fixé très haut, vaut en moyenne θ_+^{-1} . Basrak *et al.* [7] donnent une représentation des indices extrémaux θ_\pm en fonction du processus spectral de queue (Θ_t) , très proche de la représentation que nous avons obtenue pour l'indice de *cluster* :

$$\theta_\pm = \mathbb{E} \left[\left(\sup_{t \geq 0} \Theta_t \right)_\pm^\alpha - \left(\sup_{t \geq 1} \Theta_t \right)_\pm^\alpha \right]. \quad (27)$$

Remarque 5.7. Dans le cas iid ou d'indépendance des extrêmes, nous avons $\Theta_t = 0$ pour $t > 0$. Les indices θ_+ et b_+ valent nécessairement $p = \mathbb{P}(\Theta_0 = 1) = \mathbb{E}[(\Theta_0)_+^\alpha]$. Ces cas d'indépendance des extrêmes sont des cas limites pour l'indice extrémal car $\theta_+ \leq p$. C'est aussi le cas pour l'indice de *cluster* lorsque $0 < \alpha \leq 1$. Par sous-additivité de la fonction $x \rightarrow x_+^\alpha$, nous avons

$$\left(\sum_{t=0}^k \Theta_t \right)_+^\alpha - \left(\sum_{t=1}^k \Theta_t \right)_+^\alpha \leq (\Theta_0)_+^\alpha.$$

On en déduit que $b_+ \leq p$ et donc que les cas d'indépendance des extrêmes sont des cas limites pour l'indice de *cluster* lorsque $0 < \alpha \leq 1$. Ce n'est plus vrai lorsque $\alpha > 1$.

5.6 Calcul de l'indice de *cluster* sur des exemples

Le premier exemple est celui des solutions d'équations récurrentes stochastiques linéaires traité par Kesten [36]

$$X_t = A_t X_{t-1} + B_t, \quad t \in \mathbb{Z}, \quad (28)$$

où (A_t, B_t) sont des suites de vecteurs iid de marginales indépendantes entre elles. Si l'équation $E|A|^\kappa = 1$, $\kappa > 0$, a une unique solution $\kappa = \alpha$ et que $\mathbb{E}|B|^{\alpha+\varepsilon} < \infty$ pour $\varepsilon > 0$, alors le modèle (28) a une unique solution stationnaire à variation régulière d'indice α . Un calcul par récurrence donne

$$X_t = \Pi_t X_0 + R_t, \quad \text{avec } \Pi_t = A_t \cdots A_1, \quad t \geq 1, \quad (29)$$

et R_t une v.a. telle que $\mathbb{E}|R_t|^{\alpha+\varepsilon} < \infty$. On obtient

$$\mathbb{P}(|X_0| > ux, |X_0|^{-1}(X_0, \dots, X_t) \in \cdot \mid |X_0| > x) \xrightarrow{d} u^{-\alpha} \mathbb{P}((1, \Pi_1, \dots, \Pi_t)\Theta_0 \in \cdot)$$

avec, par définition, $\mathbb{P}(X_0/|X_0| = \pm 1 \mid |X_0| > x) \xrightarrow{d} \mathbb{P}(\Theta_0 = \pm 1)$. Le processus spectral de queue s'écrit donc

$$\Theta_t = \Pi_t \Theta_0, \quad \forall t \geq 0.$$

Comme Θ_0 est indépendant des A_t pour $t \geq 1$ on en déduit que $\mathbb{E}(|\Theta_t|^\alpha) = \mathbb{E}(|\Pi_t|^\alpha) \mathbb{E}(|\Theta_0|^\alpha) = 1$ et que la série (Θ_t) n'est pas normalement convergente dans \mathbb{L}^α . Pourtant on vérifie facilement la condition **(DC_p)** pour tout $0 < p < \alpha$:

$$\mathbb{E}(|X_1|^p \mid X_0 = y) = \mathbb{E}(|A_1 y + B_1|^p) \leq (\mathbb{E}(|A|^p)^{1/p} y + \mathbb{E}(|B_1|^p)^{1/p})^p$$

par l'inégalité de Minkowski et on conclut par un argument de croissance comparée car $\mathbb{E}(|A|^p)^{1/p} < 1$. Ainsi, le théorème 5.6 représente l'indice de *cluster* sous la forme

$$b_\pm = \mathbb{E} \left[\left(\sum_{t \geq 0} \Pi_t \Theta_0 \right)_\pm^\alpha - \left(\sum_{t \geq 1} \Pi_t \Theta_0 \right)_\pm^\alpha \right].$$

Le cas du modèle GARCH(1,1) est plus compliqué. On a alors $X_t = \sigma_t Z_t$ avec les innovations Z_t iid telles que $\mathbb{E}[Z_0] = 0$ et $\mathbb{E}[Z_0^2] = 1$. Rappelons que la volatilité (σ_t) est la solution de l'équation (5) qui peut s'écrire sous la forme d'une équation récurrente stochastique linéaire

$$\sigma_t^2 = \alpha_0 + \sigma_{t-1}^2(\alpha_1 Z_{t-1}^2 + \beta_1) =: B_t + \sigma_{t-1}^2 A_t.$$

Afin de vérifier (\mathbf{DC}_p) et d'exprimer les espérances conditionnelles de (X_t) en fonction du passé, nous devons considérer le processus GARCH élargi (X_t, σ_t) . Cette astuce nécessite l'extension des notions de (\mathbf{DC}_p) , (\mathbf{RV}_α) et de l'indice de *cluster* à la dimension deux telle que nous la détaillons dans [W15]. Sous des conditions d'apériodicité, d'irréductibilité, $\mathbb{E}(\alpha_1 Z_0^2 + \beta_1)^{\alpha/2} = 1$, $\alpha > 0$, Z_0 symétrique et $\mathbb{E}|Z_0|^{\alpha+\epsilon} < \infty$ nous obtenons

$$b_\pm = \frac{\mathbb{E} \left[\left| Z_0 + \sum_{t=1}^{\infty} Z_t \prod_{i=1}^t \sqrt{A_i} \right|^\alpha - \left| \sum_{t=1}^{\infty} Z_t \prod_{i=1}^t \sqrt{A_i} \right|^\alpha \right]}{2\mathbb{E}|Z|^\alpha}.$$

6 Théorie asymptotique des sommes partielles dépendantes et à variation régulière

Dans toute ce chapitre, nous supposons que la v.a. X est centrée, $\mathbb{E}(X) = 0$, dès qu'elle est intégrable.

6.1 Théorème de la limite centrale gaussienne

Nous rappelons d'abord un théorème de la limite centrale gaussienne classique

Théorème 6.1 ([45]). *Si $X_t = f(\Phi_t)$ avec (Φ_t) qui est irréductible et vérifie (\mathbf{DC}_1) et $\mathbb{E}[X_0^2] < \infty$, alors $n^{-1/2}S_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2)$ où la variance asymptotique a la représentation*

$$\sigma^2 = \lim_{k \rightarrow \infty} \mathbb{E} \left[\left(\sum_{t=0}^{k+1} X_t \right)^2 - \left(\sum_{t=1}^k X_t \right)^2 \right].$$

De plus, les cycles iid appartiennent au domaine d'attraction de la loi normale, i.e. $\mathbb{E}[S(1)^2] = \sigma^2$.

La limite gaussienne est caractérisée par la variance limite dont l'expression est similaire à celle de l'indice de *cluster* avec $\alpha = 2$ et le processus spectral de queue (Θ_t) remplacé par le processus (X_t) . Dans ce cadre, les propriétés de variation régulière ne sont pas utiles et (\mathbf{RV}_α) n'est pas nécessairement vérifiée. La condition de dérive (\mathbf{DC}_p) est particulièrement utile pour contrôler les moments d'ordre $p+1$ des cycles de la chaîne de Markov sous des conditions de moments du même ordre sur la loi marginale [52].

6.2 Théorème de la limite centrale α -stable

Le processus (X_t) satisfait un théorème de la limite centrale α -stable lorsqu'il existe une suite (a_n) telle que $a_n^{-1}S_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} Y_\alpha$ avec Y_α qui suit une loi α -stable non gaussienne

($0 < \alpha < 2$). Ces lois α -stable (la gaussienne correspondant au cas $\alpha = 2$) sont les seules lois limites possibles dans le théorème de la limite centrale pour des v.a. iid. Leur fonction caractéristique est de la forme $\psi_\alpha(x) = \exp(-|x|^\alpha \chi_\alpha(x, a_1, a_2))$ avec pour $\alpha \neq 1$

$$\chi_\alpha(x, a_1, a_2) = \frac{\Gamma(2 - \alpha)}{1 - \alpha} \left(a_1 \cos(\pi\alpha/2) - i \operatorname{sign}(x) a_2 \sin(\pi\alpha/2) \right),$$

et pour $\alpha = 1$

$$\chi_1(x, a_1, a_2) = 0.5 \pi a_1 + i \operatorname{sign}(x) a_2 \log |x|, \quad x \in \mathbb{R}.$$

Afin d'obtenir une loi limite α -stable, nous devons considérer des processus (X_t) faiblement dépendants. La condition de dépendance faible suivante assure l'approximation en loi de la somme partielle par des blocs iid de taille m_n tendant vers l'infini :

Condition (\mathbf{MX}_m) : Soit $m = (m_n)$ une suite d'entiers telle que $m_n \rightarrow \infty$ et $m_n/n = o(1)$ et soit $k_n = \lfloor n/m_n \rfloor$. Il existe une suite d'entiers (r_n) telle que $r_n = o(m_n)$ et

$$\lim_{n \rightarrow \infty} |\varphi_n^{(\ell)}(x) - \varphi_{n, m_n - \ell}^k(x)| = 0, \quad \forall x \in \mathbb{R},$$

pour $\ell = \ell_n = m_n - r_n$ et $\ell = r_n$, avec

$$\begin{aligned} \varphi_n^{(\ell)}(x) &= \mathbb{E} \left[\exp \left(i x a_n^{-1} \sum_{i=1}^{k_n} \sum_{t=(i-1)m_n+1}^{im_n-\ell} X_t \right) \right], \\ \varphi_{n,j}(x) &= \mathbb{E} \left[\exp \left(i x a_n^{-1} \sum_{t=1}^j X_t \right) \right], \quad j \geq 1, \quad \varphi_n(x) = \varphi_{n,n}(x), \quad x \in \mathbb{R}. \end{aligned}$$

La condition (\mathbf{MX}_m) est satisfaite sous des conditions plus classiques comme la θ -faible dépendance ou la τ -faible dépendance introduites respectivement par Doukhan et Louhichi dans [22] et Dedecker et Prieur dans [19], voir le lemme 2 p. 467 de [44] pour une preuve du second point. Nous avons le théorème de la limite centrale suivant

Théorème 6.2 ([W7]). *Si $X_t = f(\Phi_t)$ vérifie (\mathbf{RV}_α) pour $\alpha \in (0, 2)$, (X symétrique si $\alpha = 1$) et (\mathbf{MX}_m) pour $m_n = o(n^{(p+1)/\alpha-1})$ et si (Φ_t) vérifie (\mathbf{DC}_p) pour p positif dans $]\alpha - 1, \alpha[$ alors le théorème de la limite centrale α -stable $a_n^{-1} S_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} Y_\alpha$ est vérifié pour Y_α de fonction caractéristique*

$$\psi_\alpha(x) = \exp(-|x|^\alpha \chi_\alpha(x, b_+ + b_-, b_+ - b_-))$$

où b_\pm sont les indices de cluster (c.f. théorème 5.6).

Dans les cas où la chaîne sous-jacente (Φ_t) est irréductible, le processus (X_t) a des coefficients de mélange fort qui décroissent exponentiellement vite et la condition (\mathbf{MX}_m) est satisfaite pour $m_n = \log^2 n = o(n^{(p+1)/\alpha-1})$ pour tout $p > \alpha - 1$. Dans ce cadre plus restrictif, nous obtenons aussi la caractérisation des cycles de régénération :

Théorème 6.3 ([W15]). *Sous les conditions du théorème 6.2, si de plus (Φ_t) est irréductible alors (X_t) vérifie alors le théorème de la limite central 6.2 et les cycles $(S(i))_{i=1,2,\dots}$ définis dans (26) sont des v.a. iid à variation régulière telles que*

$$\mathbb{P}(S(1) > x) \sim b_+ \mathbb{E}_A(\tau_A) \mathbb{P}(X > x) \quad \text{et} \quad \mathbb{P}(S(1) \leq -x) \sim b_- \mathbb{E}_A(\tau_A) \mathbb{P}(X \leq -x). \quad (30)$$

La condition \mathbf{DC}_p implique que les cycles iid $(S(i))_{i>0}$ sont à variation régulière d'indice $0 < \alpha < p + 1$ lorsque le processus stationnaire est à variation régulière de même indice α .

6.3 Principe uniforme de grandes déviations

En utilisant le résultat de Nagaev [47] sur les grandes déviations des sommes de v.a. iid nous en déduisons un principe uniforme de grandes déviations des sommes de fonctions de chaînes de Markov lorsque $0 < \alpha < 2$. Dans cette section, la suite (u_n) est telle que $u_n/n^{1/2\vee 1/\alpha+\varepsilon} \rightarrow \infty$ pour $\varepsilon > 0$ quelconque. Nous obtenons

Théorème 6.4 ([W10]). *Sous les conditions du Théorème 6.3 pour $0 < \alpha < 2$ et si (\mathbf{DC}_p) et (\mathbf{RV}_α) sont satisfaites pour $\alpha > 2$ et tout $p < \alpha$, alors le processus (X_t) vérifie le principe de grandes déviations uniforme*

$$\lim_{n \rightarrow \infty} \sup_{x \in \Lambda_n} \left| \frac{\mathbb{P}(S_n > x)}{n \mathbb{P}(|X| > x)} - b_+ \right| = 0 \quad \text{avec} \quad (31)$$

- $\Lambda_n = (u_n, \infty)$ pour $0 < \alpha < 1$,
- $\Lambda_n = (u_n, v_n)$ pour $1 < \alpha$, $\alpha \neq 2$, et toute suite (v_n) telle que $v_n > u_n$ et

$$\mathbb{P}(\tau_A > n) = o(n \mathbb{P}(|X| > v_n)). \quad (32)$$

La preuve dans le cas $\alpha > 2$ est beaucoup plus technique que dans le cas $0 < \alpha < 2$. En effet, le théorème de la limite centrale gaussienne 6.1 ne fournit que les propriétés de moments d'ordre 2 des cycles $S(i)$. Pour démontrer que ces cycles sont à variation régulière d'indice α , les déviations moyennes sont donc inutiles. Dans le cas $\alpha > 2$ nous calculons directement les grandes déviations grâce à l'inégalité de Bonferroni [35].

Remarque 6.5. Le cas $0 < \alpha \leq 1$ est particulier : le cas d'indépendance des extrêmes est alors un cas limite pour l'indice de *cluster* (c.f. la remarque 5.7). Ce n'est plus le cas quand $\alpha > 1$. Pour tout $x_n \geq u_n$ on a alors

$$\frac{\mathbb{P}(S_n > x)}{n \mathbb{P}(|X| > x)} \sim b_+ + \frac{\mathbb{P}(S_n > x, \tau_A > n)}{n \mathbb{P}(|X| > x)}.$$

Le second terme de la somme est borné mais pas toujours négligeable pour des déviations très larges (x_n plus grand que e^{cn} pour $c > 0$). Les très grandes déviations sont perturbées par le premier cycle lorsqu'il n'est pas complet, le retour à l'atome A n'ayant pas lieu avant n .

6.4 Exemples de processus à queues épaisses

Suivant l'approche d'Embrechts *et al.* [25], la loi \mathbb{P} d'un processus stationnaire (X_t) est à queues épaisses si les grandes déviations de la somme partielle sont proportionnelles à celles du maximum des observations comme dans l'équation (22). Sous des conditions de variation régulière, de dérive et d'irréductibilité nous avons

$$\mathbb{P}(S_n > x_n) \sim_{n \rightarrow \infty} \frac{b_+}{\theta_+} \mathbb{P}(\max_{1 \leq i \leq n} X_i > x_n),$$

où (x_n) est une suite tendant vers l'infini dans une région convenablement choisie. La loi \mathbb{P} de tels processus est alors à queues épaisses. Le coefficient de proportionnalité b_+/θ_+ peut être différent de 1 ce qui permet la modélisation de comportements des valeurs

extrêmes plus variés que dans le cas iid ou les cas d'indépendance des extrêmes. L'indice extrémal n'est pas suffisant pour caractériser ces queues épaisses : l'indice de *cluster* est indispensable pour quantifier les queues de distribution des agrégats de valeurs extrêmes.

Dans le cas de la solution d'une équation récurrente stochastique linéaire $X_t = A_t X_{t-1} + B_t$, $t \in \mathbb{Z}$, nous obtenons le facteur multiplicatif

$$\frac{b_+}{\theta_+} = \frac{\mathbb{E} \left[\left(\sum_{t \geq 0} \Pi_t \Theta_0 \right)_+^\alpha - \left(\sum_{t \geq 1} \Pi_t \Theta_0 \right)_+^\alpha \right]}{\mathbb{E} \left[\left(\sup_{t \geq 0} \Pi_t \Theta_0 \right)_+^\alpha - \left(\sup_{t \geq 1} \Pi_t \Theta_0 \right)_+^\alpha \right]}.$$

Lorsque A_t est positif, $\Pi_t = A_t \cdots A_1$ est positif, $\Theta_0 = 1$ et nécessairement $b_+/\theta_+ > 1$ pour tout $\alpha > 0$. Dans le cas de la solution d'un modèle GARCH(1,1) symétrique, nous obtenons la représentation

$$\frac{b_+}{\theta_+} = \frac{\mathbb{E} \left[\left| Z_0 + \sum_{t=1}^{\infty} Z_t \prod_{i=1}^t \sqrt{A_i} \right|^\alpha - \left| \sum_{t=1}^{\infty} Z_t \prod_{i=1}^t \sqrt{A_i} \right|^\alpha \right]}{\mathbb{E} \left[\max \left\{ |Z_0|^\alpha, \sup_{t \geq 1} \left| Z_t \prod_{i=1}^t \sqrt{A_i} \right|^\alpha \right\} - \sup_{t \geq 1} \left| Z_t \prod_{i=1}^t \sqrt{A_i} \right|^\alpha \right]},$$

en fonction de $A_t = \alpha_1 Z_{t-1}^2 + \beta_1$.

Les solutions d'équations récurrentes stochastiques linéaire et le modèle GARCH sont des exemples de processus stationnaires à queues épaisses modélisant des grandes déviations très différentes du cas iid lorsque $b_+/\theta_+ \neq 1$. Ils permettent de tenir compte des phénomènes d'agrégation des valeurs extrêmes observés sur les données réelles.

6.5 Comportement asymptotique des auto-covariances empiriques

Une application de ces résultats en statistique des processus et la détermination du comportement asymptotique des auto-covariances empiriques. Nous généralisons et précisons ainsi des résultats de Basrak et al. [5].

Définition 6.6. *Les auto-covariances empiriques d'ordre $h \geq 0$ sont définies par la relation*

$$\hat{\gamma}_n(h) = \frac{1}{n} \sum_{t=1}^{n-h} X_t X_{t+h} - \left(\frac{1}{n} \sum_{t=1}^n X_t \right)^2.$$

Les $\hat{\gamma}_n(h)$ sont des estimateurs des auto-covariances $\gamma(h) = \mathbb{E}(X_0 X_h) - \mathbb{E}(X_0) \mathbb{E}(X_h)$.

Plaçons nous dans le cadre d'un processus à queues épaisses solution du modèle GARCH(1,1) décrit plus haut. Alors, d'après [36], il existe $c_0 > 0$ telle que $\mathbb{P}(X_0^2 > x) \sim c_0 x^{-2/\alpha}$. Notons

$$a_n = n^{4/\alpha} (c_0 \mathbb{E} |Z_0^2 Z_h^2 \Pi_h|)^{4/\alpha}.$$

Considérons le cas des auto-covariances empiriques $\hat{\gamma}_n(h)$ des carrées (X_t^2). Le théorème de la limite centrale α -stable 6.2 fournit les résultats asymptotiques $n a_n^{-1} (\hat{\gamma}_n(h) - \gamma(h)) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} Y_{\alpha/4}(h)$ pour $\alpha \in (4, 8)$ et $n a_n^{-1} \hat{\gamma}_n(h) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} Y_{\alpha/4}(h)$ pour $\alpha \in (0, 4)$. Ici, $Y_{\alpha/4}(h)$ est une loi $\alpha/4$ -stable complètement déterminée par le processus spectral de

queue $(\Theta_t) = (c_0 Z_t^2 Z_{t+h}^2 \Pi_t \Pi_{t+h})$. Par positivité du processus spectral, on en déduit en particulier que $b_- = 0$.

Les auto-covariances empiriques sont utilisées en pratique pour quantifier la dépendance temporelle du processus observé. Reprenons les données réelles de l'indice boursier anglais du Footsee du chapitre 1. Considérons les log-rendements des prix à la clôture durant la dernière période de stationnarité obtenue par le critère de QV pénalisée, à savoir du 5 décembre 2008 au 18 mars 2011 ($n = 575$). Nous traçons les auto-covariances empiriques des carrés des log-rendements dans le premier graphique de la figure 2.

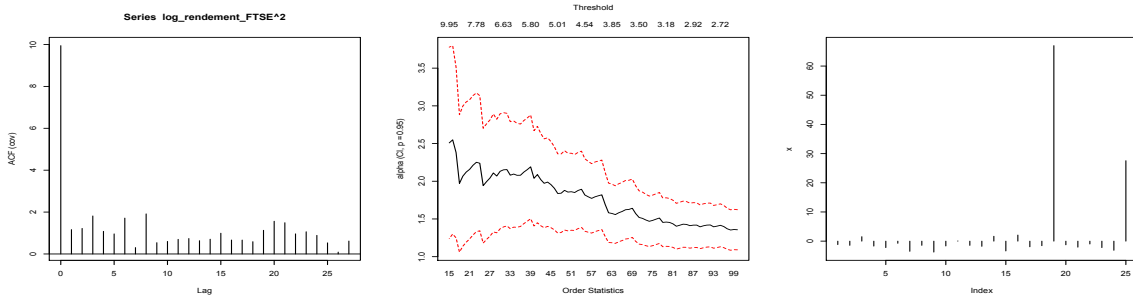


FIGURE 2 – Les auto-covariances empiriques des carrés des log-rendements sont tracées dans les deux premiers graphiques jusqu'à l'ordre $h = 27$. Les estimateurs de Hill et la bande de confiance d'estimation à 95% sont tracés dans le second graphique suivant les valeurs extrêmes retenues parmi les 100 plus grandes. Le troisième graphique représente 25 réalisations d'un loi $\alpha/4$ stable asymétrique.

Il semble que les auto-covariances des carrés des log-rendements ne soient pas négligeables, même pour des ordres h grands. Il y aurait donc de la dépendance forte sur les carrés des log-rendements. Au vu des résultats asymptotiques précédents, il est toutefois légitime de se demander quel crédit doit être donné à cette interprétation graphique. Supposons au contraire que les auto-covariances des carrés des log-rendements $\gamma(h)$ soient négligeables ou n'existent pas car $\mathbb{E}X_0^4 = \infty$. Le second graphique de la figure 2 représente les estimateurs de Hill du paramètre α suivant le nombre de valeurs extrêmes retenues dans l'expression de ces estimateurs. Une estimation raisonnable $\hat{\alpha}/2$ semble être proche de 2, tel que $na_n^{-1} \sim c$ pour une constante $c > 0$. On déduit alors du résultat asymptotique précédent $na_n^{-1}\hat{\gamma}_n(h) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} Y_{\alpha/4}(h)$ que $\hat{\gamma}_n(h) \approx c^{-1} Y_{\alpha/4}(h)$ approximativement. Lorsque $b_- = 0$ la loi asymptotique limite $Y_{\alpha/4}(h)$ est complètement asymétrique ; elle es même bornée inférieurement par $-(\mathbb{E}X_0^2)^2$ lorsque $1/2 < \alpha/4 < 1$. Les réalisations d'une telle loi, comme par exemples celles fournies dans le troisième graphique de la figure 2, sont de nombreuses petites valeurs négatives et des grandes valeurs positives. Ces réalisations correspondent aux erreurs d'estimation des auto-covariances d'un processus à queues épaisses. Les valeurs significativement non nulles des auto-covariances empiriques ne pourraient correspondre qu'à des erreurs d'estimation asymétriques et non aux auto-covariances $\gamma(h)$ qui pourraient être négligeables ou ne pas exister. Comme discuté dans [46], les erreurs d'estimation élevées et asymétriques perturbe les interprétations graphiques fondées sur les auto-covariances empiriques des processus à queues épaisses.

Conclusion et perspectives

Je décris dans ce mémoire quelques contributions personnelles à la statistique des processus faiblement dépendants. J'apprécie particulièrement lorsque les raisonnements mathématiques fournissent des indications sur les méthodes pratiques à employer. Je décris dans la première partie une approche mathématique rigoureuse qui fournit un cadre de validation pour l'estimation par QMV. La rigueur mathématique permet de rendre stable la méthode du QMV en restreignant le critère QV au domaine de continue inversibilité. Dans la seconde partie, l'approche par transport faible conditionnel fournit une nouvelle interprétation mathématique de l'utilisation des estimateurs de Gibbs comme alternative aux estimateurs classiques. Je fournis un nouveau cadre théorique qui explique dans quels cas les procédures classiques de minimisation du risque empirique sont suffisantes et dans quels cas de nouvelles méthodes de prédiction semblent préférables. Je quantifie la dépendance des valeurs extrêmes de processus dans la troisième partie grâce à un nouvel indice dit de *cluster*. Cette approche est très prometteuse et devrait avoir un grand impact sur les méthodes statistiques à employer pour décrire les queues de distribution en présence de valeurs extrêmes dépendantes. Hormis les solutions des équations récurrentes linéaires et le modèle GARCH, d'autres exemples de processus à queues épaisses devront être étudiés. Une nouvelle approche statistique de calibration des modèles aux phénomènes de *cluster* de valeurs extrêmes observés devra être construite, sans utiliser les auto-covariances empiriques. Le cadre markovien décrit dans la dernière partie me semble être particulièrement adapté pour traiter de la dépendance des valeurs extrêmes. Des méthodes, fondées sur les cycles de régénération, permettront vraisemblablement de quantifier ces agrégats en pratique. Des applications en quantification des risques dans un cadre dépendant plus réaliste que le cas d'indépendance des extrêmes sont attendues.

Pour conclure, je souhaite donner quelques exemples d'applications concrètes en quantification des risques entreprises. Depuis trois ans, je suis co-responsable avec R. Elie du Master actuariat de l'université Paris Dauphine. Encadrant des étudiants en stage professionnel de fin d'étude, j'ai un aperçu des problématiques actuelles de la profession. En particulier, les nouvelles réglementations dites de *Solvability II* se fondent sur la prédiction d'une *Value at Risk* (VaR) à horizon un an. D'un point de vue statistique, ce programme est très ambitieux et repose sur une grande confiance au modèle prédictif utilisé. Même en supposant le modèle choisi comme étant satisfait par les données, le calcul des mesures de risques (VaR ou probabilités de ruine) est rarement explicite et nécessite une approximation numérique [3]. Les techniques de Monte Carlo (MC) ne sont pas efficaces pour approcher des valeurs issues d'événements rares. Durant son stage au LFA-CREST sous ma tutelle, G. Terré a développé des algorithmes multi-échelles bien plus efficaces que les méthodes MC dans le cadre classique iid. L'étape suivante est d'étendre ces méthodes dans un contexte markovien. Les résultats théoriques de la partie III sur les cycles de régénération des chaînes de Markov sont utiles à l'étude théorique de la convergence de tels algorithmes.

En pratique, un bon modèle de prédiction n'existe certainement pas et la recherche d'un unique modèle pour la prédiction d'une VaR n'est pas forcément la solution. J'ai encadré le stage de recherche de F. Torron sur l'agrégation de modèles de prévisions de la VaR à horizon un jour sur données financières. La littérature en finance quantitative fournit de nombreux modèles dont l'efficacité est variable. En particulier, Giot

et Laurent [30] comparent deux modèles de prévisions de type GARCH en utilisant la méthode d'estimation par QMV de la partie I. Nous avons développé deux méthodes distinctes d'agrégation de ces modèles de prévisions. La première méthode provient de l'estimation de Gibbs comme décrite dans la partie II. La seconde méthode, dite de suites individuelles [16] et très répandue en théorie des jeux, part du constat qu'un mélange astucieux de toutes les prévisions donne souvent de bien meilleurs résultats que la meilleure des prévisions. En pratique, ces deux méthodes d'agrégation donnent des résultats voisins et semblent d'une efficacité comparable. Il sera intéressant de comparer ces approches en théorie en utilisant le transport faible conditionnel de la partie II.

Références

- [W1] RAGACHE, N. ET WINTENBERGER, O. (2006) Convergence rates for density estimators of weakly dependent time series. *Dependence in Probability and Statistics, Lecture Notes in Statist.* **187**, 349–372.
- [W2] DOUKHAN, P. ET WINTENBERGER, O. (2007) An invariance principle for weakly dependent stationary general models. *Probab. Math. Statist.* **27**, 45–73.
- [W3] DOUKHAN, P. ET WINTENBERGER, O. (2008) Weakly dependent chains with infinite memory. *Stoch. Proc. Appl.* **118**, 1997–2013.
- [W4] BARDET, J.-M. ET WINTENBERGER, O. (2009) Asymptotic normality of the Quasi Maximum Likelihood Estimator for multidimensional causal processes. *Ann. Statist.* **37**, 2730–2759.
- [W5] GANNAZ, I. ET WINTENBERGER, O. (2010) Adaptive density estimation with dependent observations. *ESAIM Probab. Statist.* **14**, 151–172.
- [W6] WINTENBERGER, O. (2010) Deviation inequalities for sums of weakly dependent time series. *Elect. Comm. in Probab.* **15**, 489–503.
- [W7] BARTKIEWICZ, K., JAKUBOWSKI, A., MIKOSCH, T. ET WINTENBERGER, O. (2011) Stable limits for sums of dependent infinite variance random variables. *Probab. Th. Rel. Fields* **150**, 337–372.
- [W8] ALQUIER, P. ET WINTENBERGER, O. (2012) Model selection for weakly dependent time series forecasting. *Bernoulli* **18**, 883–913.
- [W9] BARDET, J.-M., KENGNE, W. C. ET WINTENBERGER, O. (2012) Multiple breaks detection in general causal time series using penalized quasi-likelihood. *Electron. J. Statist.* **6**, 435–477.
- [W10] MIKOSCH, T. ET WINTENBERGER, O. (2012) Precise large deviations for dependent regularly varying sequences. *Probab. Th. Rel. Fields*, à paraître.
- [W11] WINTENBERGER, O. ET CAI, S. (2012) Parametric inference and forecasting in continuously invertible volatility models. Prépublication.
- [W12] TRASHORRAS, J. ET WINTENBERGER, O. (2012) Large deviations for bootstrapped empirical measures. Prépublication.
- [W13] ALQUIER, P. ET WINTENBERGER, O. (2012) Fast rates in learning with dependent observations. Prépublication.
- [W14] WINTENBERGER, O. (2012) Weak transport inequalities and applications to exponential and oracle inequalities. Prépublication.
- [W15] MIKOSCH, T. ET WINTENBERGER, O. (2012) The cluster index of regularly varying sequences with applications to limit theory for functions of multivariate Markov chains. Prépublication.
- [W16] FRANCO, C., WINTENBERGER, O. ET ZAKOÏAN, J.-M. (2012) GARCH models without positivity constraints : Exponential or Log GARCH? Prépublication.

- [1] AKAIKE, H. (1973) Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory*, 267–281.
- [2] AUDIBERT, J.-Y. ET CATONI, O. (2011) Robust linear least squares regression. *Ann. Statist.* **39** (5), 2766–2794.
- [3] ASMUSSEN, S., GLYNN, P.W. (2007) *Stochastic simulation : algorithms and analysis*. Springer, New York.
- [4] BARAUD, Y, COMTE, F., VIENNET, G. (2001) Model selection for (auto-)regression with dependent data. *ESAIM Probab. Statist.*, **5**, 33–49.
- [5] BASRAK, B., DAVIS, R.A. ET MIKOSCH. T. (2002) Regular variation of GARCH processes. *Stoch. Proc. Appl.* **99**, 95–116.
- [6] BASRAK, B. ET SEGERS, J. (2009) Regularly varying multivariate time series. *Stoch. Proc. Appl.* **119**, 1055–1080.
- [7] BASRAK, B. KRIZMANIC, D. ET SEGERS, J. (2011) A functional limit theorem for dependent sequences with infinite variance stable limits. *Ann. Probab.*, à paraître.
- [8] BERKES, I., HORVATH, L. ET KOKOSZKA, P. (2003) GARCH processes : structure and estimation. *Bernoulli*, **9**, 201–227
- [9] BOBKOV, S. G. ET GÖTZE, F. (1999) Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163**, 1–28.
- [10] BOLLERSLEV, T. P. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327.
- [11] BOUCHERON, S., LUGOSI, G. ET MASSART, P. (2000). A sharp concentration inequality with applications. *Random Structures and Algorithms* **16** (3), 277–292.
- [12] BOUCHERON, S., LUGOSI, G. ET MASSART, P. (2009) On concentration of self-bounding functions. *Electron. J. Probab.* **14** (64), 1884–1899.
- [13] BOUGEROL, P. (1993) Kalman filtering with random coefficients and contractions. *SIAM J. Control and Optimization* **31**, 942–959.
- [14] BRANDT, M. W. ET JONES, C. S. (2006) Volatility forecasting with range-based EGARCH models. *Journal of Business & Economic Statistics*, **24**, 470–486.
- [15] CATONI, O. (2004) *Statistical Learning Theory and Stochastic Optimization*. Lecture Notes in Mathematics, Springer, Berlin.
- [16] CESA-BIANCHI, N. ET LUGOSI, G. (2006) *Prediction, learning, and games*. Cambridge University Press, Cambridge.
- [17] CLINE, D.B.H. ET HSING, T. (1998) Large deviation probabilities for sums of random variables with heavy or subexponential tails. *Technical Report*, Texas A&M University.
- [18] DAVIS, R.A. ET HSING, T. (1995) Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Prob.* **23**, 879–917.
- [19] DEDECKER, J. AND PRIEUR, C. (2004) Coupling for τ -dependent sequences and applications. *J. Theor. Probab.*, **17**(4), 861–885.

- [20] DEDECKER, J. AND PRIEUR, C. (2005) New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields*, **132**(2), 203–236.
- [21] DJELLOUT, H., GUILLIN, A. ET WU, L. (2004) Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.* **32**, 2702–2732.
- [22] DOUKHAN, P. AND LOUHICHI, S. (1999) A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* **84**, 313–342.
- [23] ELTON, J. H. (1990) A multiplicative ergodic theorem for Lipschitz maps. *Stoch. Proc. Appl.*, **34**, 39–47.
- [24] ENGLE, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* **50**(4), 987–1007.
- [25] EMBRECHTS, P., KLÜPPELBERG, C. ET MIKOSCH, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.
- [26] EMBRECHTS, P. ET VERAVERBEKE, N. (1982) Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance Math. Econom.* **1**, 55–72.
- [27] FELLER, W. (1971) *An Introduction to Probability Theory and Its Applications*. Vol. II. Second edition. Wiley, New York.
- [28] FRANCO, C. ET ZAKOIAN, J.-M. (2010) *GARCH Models - Structure, Estimation and Finance Applications*. John Wiley.
- [29] FRYZLEWICZ, P. ET SUBBA RAO, S. (2010) BaSTA : consistent multiscale multiple change-point detection for piecewise-stationary ARCH processes. Prépublication.
- [30] GIOT, P. ET LAURENT, P. (2004) Modeling daily Value-at-Risk using realized volatility and ARCH type models. *Journal of Empirical Finance*, **11**, 379–398.
- [31] GOLDSTEIN, S. (1979) Maximal coupling. *Z. Wahrsch. Verw. Gebiete* **46**, 193–204.
- [32] GOZLAN, N. ET LÉONARD, C. (2010) Transport inequalities. A survey. *Markov Processes and Related Fields*, **16**, 635–736.
- [33] GOZLAN, N., ROBERTO, C., SAMSON, P.-M. ET TETALI, P. (2012) Displacement convexity of entropy and related inequalities on graphs. Prépublication.
- [34] GREEN, P. J. (1995) Reversible Jump Markov Chain Monte Carlo computation and Bayesian Model Determination. *Biometrika* **82**, 711–732.
- [35] JAKUBOWSKI, A. (1997) Minimal conditions in p -stable limit theorems - II. *Stoch. Proc. Appl.* **68**, 1–20.
- [36] KESTEN, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**, 207–248.
- [37] LAVIELLE, M. ET MOULINES, E. Least squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis* **21**, (2000) 33–59.
- [38] LEADBETTER, M.R. (1983) Extremes and Local Dependence in Stationary Sequences. *Z. Wahrsch. Verw. Gebiete* **65**, 291–306.

- [39] LECUÉ, G. ET MENDELSON, S. (2012) General nonexact oracle inequalities for classes with a subexponential envelope. *Ann. Statist.* **40** (2), 832–860.
- [40] MARTON, K. (1996) A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.* **6** (3), 556–571.
- [41] MASSART, P. (2007), *Concentration inequalities and model selection*. Springer, Berlin.
- [42] MCALLESTER, D. A. (1999) PAC-Bayesian model averaging. *Proceedings of the Twelfth Annual Conference on Computational Learning Theory*, 164–170.
- [43] MERLEVÈDE, F., PELIGRAD, M. ET RIO, E. (2009) Bernstein inequality and moderate deviations under strong mixing conditions. *IMS Collections, High Dimensional Probability V*, 273–292.
- [44] MERLEVÈDE, F., PELIGRAD, M. ET RIO, E. (2011) A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Th. Rel. Fields*, **151**, 435–474.
- [45] MEYN S.P. ET TWEEDIE R.L. (1993), *Markov Chains and Stochastic Stability*. Springer, London.
- [46] MIKOSCH, T. AND STARICA, C. (2003) Long-range dependence effects and ARCH modeling. In : DOUKHAN, P., OPPENHEIM, G. AND TAQQU, M.S. (Eds.) *Long-Range Dependence*. Birkhäuser, Boston, pp.439–459.
- [47] NAGAEV, A.V. (1969) Integral limit theorems for large deviations when Cramér’s condition is not fulfilled I,II. *Theory Probab. Appl.* **14**, 51–64 et 193–208.
- [48] NELSON, D. B. (1991) Conditional heteroskedasticity in asset returns : a new approach. *Econometrica* **59**, 347–370.
- [49] NUMMELIN, E. (1984) *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge University Press, Cambridge.
- [50] RIO, E. (2000) Inégalités de Hoeffding pour les fonctions lipschitziennes de suites dépendantes. (French) *C. R. Acad. Sci. Paris Sér. I Math.* **330** (10), 905–908
- [51] SAMSON, P.-M. (2000) Concentration of measure inequalities for Markov chains and Φ -mixing processes. *Ann. Probab.* **28** (1), 416–461.
- [52] SAMUR, J.D. (2004) A regularity condition and a limit theorem for Harris ergodic Markov chains. *Stoch. Proc. Appl.*, **111**, 207–235.
- [53] STRAUMANN, D., ET MIKOSCH, T. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series : A stochastic recurrence equations approach. *Ann. Statist.* **34** (2006), 2449–2495.
- [54] SOROKIN, A. (2011) Non-invertibility in some heteroscedastic models. Prépublication.
- [55] TALAGRAND, M. (1995) Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l’I.H.E.S.*, **81**, 73–205.
- [56] VILLANI, C. (2009) *Optimal transport, old and new*. Springer-Verlag, Berlin, 2009.