Mémoire de synthèse en vue de l'obtention de l'habilitation à diriger des recherches.

"Quelques contributions à l'étude
de
la dépendance
en
Statistique et en Économétrie"

Fabienne Comte*

Juin 2000

*University Paris VI, Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, email: comte@ccr.jussieu.fr

Remerciements

Je tiens tout d'abord à remercier Eric Renault qui fut mon directeur de thèse et dont l'enthousiasme et la générosité me donnèrent le goût de la recherche. Il a su m'encadrer en me laissant ma liberté puis ne pas perdre le contact. Il revient du bout du monde pour participer à ce jury aujourd'hui, c'est pour moi un honneur et une joie.

J'éprouve aussi une grande reconnaissance à l'égard de Lucien Birgé, qui m'a amenée à découvrir davantage la Statistique et avec qui je me suis initiée à la Sélection de Modèles. Qu'il soit remercié de sa patience, son indéfectible disponibilité pour écouter les questions et aider à trouver les réponses. Je ne serai jamais assez perfectionniste à son goût, je n'en ai que trop conscience, mais il me le pardonnera encore, je l'espère.

Je lui sais gré par ailleurs de m'avoir poussée, en même temps que Jean Jacod, à déposer le dossier qui m'amène à soutenir aujourd'hui cette habilitation : je mentirais si je niais en être fière et heureuse.

Je remercie les rapporteurs Rainer Dahlhaus, Jean-Marie Dufour et Oleg Lepski d'avoir eu le courage de se livrer à l'exercice quelque peu ingrat de référer l'ensemble de mes travaux : je suis heureuse que leurs conclusions m'aient été favorables. Je ne sais comment leur dire à quel point j'apprécie leur effort.

Je souhaite aussi exprimer ma gratitude à l'égard de Pascal Massart, Dominique Picard et Emmanuel Rio, qui me font l'honneur de participer également à ce jury. Pascal Massart m'a tout à la fois initiée, lors d'un séjour à Schmerwitz, à l'estimation par contraste pénalisé et à la cueillette de champignons non identifiés; merci à lui, même si tous les surnoms dont il affuble ceux qu'il côtoie m'ont coûté la mise au rencart d'un tailleur jaune. Mon féminisme evanescent s'enorgueillit irrésistiblement de la présence ici de Dominique Picard. J'apprécie sa confiance pour avoir partagé avec moi un cours de DEA; nos interrogations pédagogiques sur l'impassibilité de nos élèves certains jours ne finiront sans doute qu'avec nos carrières ! Merci à elle. Emmanuel Rio m'a permis, un Schmerwitz plus loin, d'y voir plus clair dans le mélange, j'espère avoir fait bon usage de ses lumières, qu'il en soit remercié.

Je voudrais dire aussi un mot à tous ceux avec qui j'ai ou j'ai eu le plaisir de collaborer : Offer Lieberman, Gabrielle Viennet, Yannick Baraud, Yves Rozenholc, Marie-Luce Taupin, Florence Merlevède, par ordre d'entrée en scène. Ce mot, quels qu'aient été les moments de tension, et pour tous les instants de jubilation, de joie ou toutes les disgressions : merci.

Je remercie également la vieille URA 1321 de m'avoir accueillie jadis dans l'UFR 920, et en particulier Denis Bosq. Je remercie le nouveau Laboratoire de Probabilités et Modèles Aléatoires dans son ensemble de m'accueillir aujourd'hui. Le professeur Huyen Pham en particulier, faute de partager son bureau avec un autre professeur, sera désormais en compagnie d'une habilitée, j'espère qu'il apprécie la promotion.

A l'origine de toute cette aventure, mais qui a préféré m'encadrer autrement qu'en thèse, il y a Gilles Pagès; alors, même si je ne suis pas du tout insupportable, merci à lui et à sa force de conviction. Merci à Julie et Romain pour tout le temps que je ne leur donne pas pour cause de travail, merci à Radojka Stépanovic, leur nourrice en qui j'ai une telle confiance que je ne me sens qu'à peine coupable, merci à mes parents qui ne comprendront jamais pourquoi je n'ai jamais fini mes devoirs mais qui se feront bien une raison un jour!

Présentation Générale

Les travaux présentés dans ce mémoire se décomposent en 7 travaux publiés et 7 travaux soumis pour publication ou en révision. En voici la liste :

Articles publiés :

1. "Simulation and estimation of long memory continuous time models", *Journal of Time Series Analysis*, 1, 19-36, 1996.

2. "Non Causality in Continuous Time Models", en collaboration avec Eric Renault, *Econometric Theory*, 12, 215-256, 1996.

3. "Long Memory Continuous Time Models", en collaboration avec Eric Renault, *Journal of Econometrics*, 76, 101-149, 1996.

4. "Regularized log-periodogram regression for Gaussian processes with bounded spectral density", en collaboration avec C. Hardouin, *Note aux CRAS*, t.325, Série I, p.1203-1206, 1997.

5. "Long memory in continuous time stochastic volatility models" en collaboration avec E. Renault, *Mathematical Finance*, Vol.8, 4, 291-323, 1998.

6. "Discrete and Continuous Time Cointegration", *Journal of Econometrics*, Vol. 88, 2, 207-226, 1999.

7. "Variance Noncausality in Multivariate GARCH Processes", en collaboration avec O. Lieberman, *Journal of Time Series Analysis*, vol. 21, 5, 535-557, 2000.

Travaux soumis pour publication ou en révision:

8. "Adaptive estimation in an autoregression and a geometrical β-mixing regression framework", en collaboration avec Y. Baraud et G. Viennet, Prépublication Crest-Insee 9807 en révision pour les *Annals of Statistics*.

9. "Model selection for (auto)-regression with dependent data", en collaboration avec Y. Baraud et G. Viennet, Prépublication ENS 99-12 et soumis pour publication à *ESAIM*.

10. "Adaptive Estimation of the spectrum of a stationary Gaussian sequence", Prépublication Paris 6 n^o 525, accepté pour publication à *Bernoulli*.

11. "Kernel Estimation of a Stochastic Volatility Model", Prépublication Paris 6 n^o 560, en révision pour *Econometric Theory*.

12. "Asymptotic Theory for Multivariate GARCH Processes", en collaboration avec O. Lieberman, Prépublication Paris 6 n^o 593, en révision pour *Journal of Multivariate Analysis*.

13. "Adaptive estimation of mean and volatility functions in (auto-)regressive models", en collaboration avec Y. Rozenholc, soumis pour publication à *Stochastic Processes and Their Applications*.

14 "Semi-parametric estimation in the (auto)-regressive β-mixing model with errors-in-variables", en collaboration avec M.-L. Taupin, Prépublication Paris 6 n^o 606, soumis pour publication à *Mathematical Methods of Statistics*.

1 Introduction

J'ai abordé au cours de ces dernières années différents sujets qui pourraient donner lieu a différents regroupements : par discipline (Statistique ou Econométrie), par méthodologie (Estimation paramétrique ou Sélection de modèles) ou par type de représentation (spectrale ou temporelle). J'opte ici pour un plan visant plutôt à séparer les types de dépendance entre les variables. Pendant longtemps en effet, on a fait de la statistique des variables indépendantes et identiquement distribuées (i.i.d.). Mais l'Économie, en se formalisant, a voulu insérer ses variables dans des modèles. Ainsi est née cette discipline que l'on appelle Économétrie et qui vise en grande partie à mettre en place les méthodes mathématiques et avant tout statistiques capables de gérer rigoureusement les variables, les aléas et les paramètres. On ne peut plus désormais placer d'hypothèses de type i.i.d. que sur les termes d'erreurs, les variables en sortie des modèles étant dès lors dépendantes; les idées d'autorégression (i.e. de dépendance de la valeur d'aujourd'hui vis à vis de celle du passé) ou de stationarité (i.e. de variabilité globalemant identique au cours du temps) ont fait leur chemin.

Un seconde évolution, due celle-là davantage à l'émergence – puis l'hégémonie – de la Finance, a été le passage de modèles en temps discret à des modèles en temps continu plus à même de rendre compte de l'évolution des processus de prix. Il faut alors re-définir un grand nombre de notions en temps continu, vérifier leur intérêt et étudier leurs caractérisations dans des cadres bien définis. Les échantillons d'observations restant en revanche à temps discret, ces nouveaux modèles à temps continu ont donné lieu, en retour, à l'étude de leurs discrétisations dans le but de mettre en place des méthodes d'estimation et de test adaptées. De plus, pour expérimenter ces méthodes on souhaite simuler ces modèles; on s'intéresse ainsi aux vitesses de convergence des discrétisés.

Mon travail est axé sur ces deux orientations : re-définir en temps continu des notions telles que la causalité (dépendance "orientée" dans le temps, la cause survenant communément avant l'effet), la mémoire longue, la cointégration, en s'interrogeant sur les méthodes statistiques à utiliser pour estimer les paramètres du temps continu sous-jacents à des échantillons à temps discret; mettre au point des méthodes statistiques élaborées pour traiter les modèles les plus simples tels que l'autorégression vue comme exemple archétypique de la régression en contexte dépendant.

Entrons un peu dans le détail. Mon tout premier travail [2], en collaboration avec E. Renault, s'attachait à modéliser cette dépendance particulière qu'est la causalité. Préexistait une définition de la causalité en moyenne pour des modèles à temps discret, généralement appliquée à des séries économiques : cette définition formalise l'idée de l'utilité, mesurée en moyenne, du passé d'une variable pour en prévoir une autre. Les développements récents de la Finance nécessitaient alors deux types de généralisations. Tout d'abord, puisque beaucoup de modèles sont représentés en temps continu, peut-on définir, caractériser et tester une causalité plus instantanée et plus structurelle en temps continu ? Ensuite, les applications financières se préoccupant tout particulièrement des oscillations de la volatilité, peut-on étudier, caractériser et tester de la causalité en variance, en temps continu, mais aussi et d'abord en temps discret, dans des modèles ARCH

par exemple ? [2] est une réponse à la première question, [6] étudie des cas de singularités des modèles détaillés dans [2] et [7], en collaboration avec O. Lieberman, correspond à un traitement de la seconde question. Les outils théoriques qui légitiment certains tests proposés dans [7] sont rigoureusement mis en place dans [12], toujours en collaboration avec O. Lieberman.
Les travaux [1], [3], [4], [5], [10] étudient ensuite des modèles à dépendance "stable" via des propriétés de stationarité et éventuellement très importante via la mémoire longue. Le point commun de ces contributions réside aussi dans un passage systématique par le domaine spectral. Le point de vue est plus modélisateur dans [3], [1] et [5] où il s'agit de définir des modèles stationnaires et à mémoire longue en temps continu. L'étude de ces modèles, représentés sous forme d'intégrales par rapport à des Browniens Fractionnaires, a connu depuis d'importants développements tant théoriques qu'appliqués. [3] et [5] sont des travaux en collaboration avec E. Renault. Le cadre est là résolument paramétrique. On se base par ailleurs sur des observations à temps discret pour l'estimation : cela permet d'utiliser des méthodes statistiques discrètes modulo l'étude des discrétisations et du transfert des propriétés de mémoire longue du temps continu au temps discret. L'ensemble de ces contributions aboutit dans [5] à des modèles à volatilité stochastique et à mémoire longue pour lesquels on peut proposer tout à la fois des justifications, des caractérisations de la mémoire, des méthodes de simulation et d'estimation. [4], en collaboration avec C. Hardouin, synthétise un gros travail sur les méthodes d'estimation à temps discret par régression sur le log-périodogramme régularisé. Ces travaux n'ont pas donné lieu à publication mais m'ont permis d'affiner grandement mes connaissances en estimation dans le domaine spectral. Enfin [10] propose une méthode d'estimation non-paramétrique et adaptative de la densité spectrale d'un processus stationnaire par sélection de modèles, tant que cette densité reste de carré intégrable, c'est-à-dire essentiellement en courte mémoire. Néanmoins, la méthode est robuste à des formes limitées de longue mémoire.
Enfin, le thème récurrent aux travaux [8], [9], [11] [13], [14] réside dans la gestion d'une forme de dépendance appelée β-mélange. Cette dépendance résulte dans tous les cas de modèles de type autorégressifs avec ([11], [14]) ou sans ([8], [9], [13]) erreurs de mesure. [8] et [9], en collaboration avec Y. Baraud et G. Viennet, considèrent des modèles de régression mélangeants pour des erreurs à variance constante. [13], en collaboration avec Y. Rozenholc, étend les résultats ainsi obtenus à des autorégressions en présence de fonctions de volatilité c'est-à-dire d'hétéroscédasticité. Une étude empirique conséquente, à base de simulations, est de plus menée afin de valider la méthode. Par ailleurs, les économistes mettent souvent en concurrence, lorsqu'ils veulent rendre compte de variations de la volatilité, la fonction de volatilité déterministe des modèles régressifs hétéroscédastiques et les modèles à volatilité stochastique pour lesquels la volatilité est non seulement variable mais aussi aléatoire, et associée à un aléa qui lui est spécifique. Le problème majeur alors est que cette volatilité n'est pas observée, ce qui dégrade les vitesses de convergence des estimateurs. Ainsi, [11] relie les modèles à volatilité stochastique à des modèles régressifs avec erreurs-dans-les-variables (i.e. erreur de mesure) et étend au cadre mélangeant les propriétés d'estimateurs à noyaux, bien connus dans le cadre indépendant.

Cependant, les vitesses optimales de convergence vers la vraie fonction de ces estimateurs à noyaux sont souvent aussi déplorables que des puissances de $\ln(n)$. C'est la raison pour laquelle on cherche dans [14], en collaboration avec M.L. Taupin, à mieux spécifier la fonction de régression en menant une étude dans un cadre semi-paramétrique : on est conduit là à généraliser en terme de type d'erreurs et de forme de dépendance un estimateur mis en place par Taupin (2000).

Il est à signaler que j'ai, dans mon travail, réalisé de constants allers et retours de la théorie à l'applicabilité des méthodes proposées, qui ont toutes ou presque été testées sur des données simulées, parfois aussi sur des données réelles (tests de causalité ou de longue mémoire sur données financières). Aller ainsi jusqu'au bout du processus de mise au point était indispensable puisqu'il s'agissait souvent de valider des procédés nouveaux : soit la mise en place de modélisations nouvelles associées à l'adaptation de méthodes plus anciennes, soit l'utilisation de méthodes nouvelles pour des modèles très anciens.

2 Causalité

2.1 Causalité en temps continu : [2], [6]

Le premier problème, abordé dans [2], en collaboration avec E. Renault, est celui de la modélisation de la notion de causalité dont la définition la plus usitée se fonde sur des comparaisons d'espérances conditionnelles d'un processus donné par rapport à des tribus plus ou moins riches en information.

Plus précisément, on considère $X(t) = (X_1(t)', X_2(t)', X_3(t)')',\ t \in \mathbb{R}^+$, où les $X_i(t)$, $i = 1, 2, 3$, sont des vecteurs lignes de taille $n_i \times 1$, $n_1 + n_2 + n_3 = n$ et $X_t = (X'_{1,t}, X'_{2,t}, X'_{3,t})',\ t \in \mathbb{N}$, son homologue en temps discret. Si $n_3 = 0$, i.e. s'il n'y a pas de troisième variable X_3, on dit qu'il n'y a pas d'environnement. Pour étudier la causalité de X_1 sur X_2, on associe à ces vecteurs, à chaque instant entier $t \in \mathbb{N}$, les tribus :

$$\mathcal{I}_D(t) = \sigma\{X_\tau,\ \ \tau \leq t,\ \ \tau, t \in \mathbf{Z}\} \text{ et } \mathcal{I}_D^{(1,3)}(t) = \sigma\{X_{1,\tau}, X_{3,\tau},\ \ \tau \leq t,\ \ \tau, t \in \mathbf{Z}\}.$$

Par analogie, on définit en temps continu, pour une date réelle t,

$$\mathcal{I}(t) = \sigma\{X(\tau),\ \ \tau \leq t,\ \ \tau, t \in \mathbb{R}\} \text{ et } \mathcal{I}^{(1,3)}(t) = \sigma\{X_1(\tau), X_3(\tau),\ \ \tau \leq t,\ \ \tau, t \in \mathbb{R}\}.$$

La définition usuelle en temps discret :

$$X_2 \text{ ne cause pas au sens de Granger} X_1 \Leftrightarrow \mathbb{E}[X_{1,t+1} | \mathcal{I}_D(t)] = \mathbb{E}[X_{1,t+1} | \mathcal{I}_D^{(1,3)}(t)] \qquad (1)$$

rend compte de l'idée que l'on n'a besoin de rien savoir sur le passé de X_2 pour prévoir X_1. Néanmoins, cette définition se révèle instable par agrégation temporelle, c'est-à-dire qu'elle dépend du pas de temps choisi, ici noté de façon unitaire, mais arbitraire. De ce point de vue, cette notion n'est pas structurelle et, pour des processus à évolution rapide, modélisés en temps continu, elle paraît inadaptée.

On s'aperçoit alors grâce à Florens et Fougère (1990), que la définition de la non causalité au sens de Granger en temps discret peut se caractériser dans le cadre de processus sous-martingales en termes de la décomposition de Doob associée à X_1. Par analogie avec cette caractérisation, on se place en temps continu dans le cadre de processus semi-martingales càdlàg ($X(t) = X(0)+H(t)+M(t), H(0) = M(0) = 0$, M martingale locale, H prévisible), et on considère la définition :

Définition 2.1 *: X_2 ne cause pas X_1 immédiatement au sens de Granger si les deux décompositions canoniques de $X_1(t)$ par rapport à $\mathcal{I}(t)$ et à $\mathcal{I}^{(1,3)}(t)$ sont identiques (à un processus evanescent près).*

On prouve alors (section 2.2) que cette définition peut se caractériser en terme d'une autre généralisation envisageable, plus en rapport avec la définition discrète :

$$\operatorname*{limq}_{h\to 0} \mathbb{E}\left[\left.\frac{X_1(t+h)-X_1(t)}{h}\right|\mathcal{I}(t)\right] = \operatorname*{limq}_{h\to 0} \mathbb{E}\left[\left.\frac{X_1(t+h)-X_1(t)}{h}\right|\mathcal{I}^{(1,3)}(t)\right]$$

où limq signifie "limite en moyenne quadratique", c'est-à-dire dans $\mathbb{L}_2$. Bien sûr, il faut d'abord que ces deux limites existent : on étudie sous quelles conditions de régularité on trouve une équivalence. Notons que l'on travaille dans un cadre de processus continus dans $\mathbb{L}_2$ et de variance finie, de sorte qu'une définition en accroissements seuls serait trivialement vérifiée.

On traite de façon similaire le passage en temps continu de la non causalité instantanée discrète ([1]). En effet, la non causalité instantanée discrète se pose la question de savoir si la connaissance de $X_{2,t+1}$, en plus de $\mathcal{I}_D(t)$ améliore ou non la prévision de $X_{1,t+1}$; en d'autres termes, on compare :

$$\mathbb{E}\left[X_{1,t+1}|\,\mathcal{I}_D(t)\right] \quad \text{et } \mathbb{E}\left[X_{1,t+1}|\,\mathcal{I}_D(t)\vee X_{2,t+1}\right]$$

Cette notion se caractérise par ailleurs en terme de covariance, ce qui fait apparaître une symétrie qui n'était pas véritablement souhaitée au départ et qui a constitué une des principales critiques émises à l'encontre de ce concept. On aboutit alors, en étudiant les décompositions de Doob comme précédemment, à la définition pour les semi-martingales (spéciales) en temps continu :

Définition 2.2 *: Soit X une semi-martingale càdlàg, $X(t) = X(0) + M(t) + H(t)$, et de façon similaire, $X_i(t) = X_i(0) + M_i(t) + H_i(t), i = 1, 2, 3,$ leurs décompositions canoniques; on suppose que $\mathcal{I}(t)$ et $\mathcal{I}^{(1,3)}(t)$ vérifient les "conditions habituelles". Il n'y a pas de causalité immédiate instantanée entre X_1 et X_2 si $M_1(t)\,{}^tM_2(t)$ est une martingale locale par rapport à $\mathcal{I}(t)$.*

On prouve alors (section 2.3) une caractérisation en termes de covariances conditionnelles, peut-être plus proche de la définition discrète. On présente (section 2.4) un modèle tout

[1]C'est le terme "instantané" employé ici en temps discret qui explique que nous ayons appelé la causalité en temps continu "immédiate"; le terme instantané aurait sinon paru tout à fait adéquat.

particulièrement adapté à l'etude des décompositions semi-martingales : le modèle à volatilité stochastique de Hull et White (1987) :

$$\begin{cases} dS_t/S_t = \mu_1(\sigma_t)dt + \sigma_t dw_1(t) \\ d\sigma_t = \mu_2(\sigma_t)dt + f_2(\sigma_t)dw_2(t) \end{cases}$$

où $(w_1, w_2)'$ est un mouvement Brownien standard. Ce modèle contient plusieurs hypothèses de non causalité, par rapport à la forme générale d'une Equation Différentielle Stochastique en dimension 2 qui ferait intervenir des coefficients fonctions quelconques des deux processus S_t et σ_t. On est même amené à définir une notion de non causalité en variance relativement nouvelle, même en temps discret.
On choisit ensuite (section 3.1.1) d'appliquer ces définitions dans un cadre plus simple de processus :

$$X(t) = \int_0^t A(t,s)dW(s)$$

où W est un mouvement Brownien n-dimensionnel et A une fonction matricielle déterministe. Ce cadre est suggéré par la décomposition de Wold continue (cf. Rozanov (1967)) des processus stationnaires gaussiens sans composante déterministe, qui se représentent sous la forme : $\int_{-\infty}^t A(t-s)dW(s)$. On travaille alors sur ces analogues à temps continu des moyennes mobiles du temps discret. On donne la décomposition semi-martingale associée à ces processus ainsi que des conditions sous lesquelles on peut obtenir une représentation inverse :

$$W(t) = \int_0^t B(t,s)dX(s).$$

On peut alors caractériser la non causalité par des conditions naturelles de nullité de sous matrices correspondant au découpage des matrices $n \times n$ en 9 blocs de taille $n_i \times n_j$, $i, j = 1, 2, 3$ (découpage associé de façon naturelle à la partition du vecteur $X = (X_1', X_2', X_3')$). On démontre ainsi (section 3.2) que la non causalité immédiate au sens de Granger de X_2 sur X_1 est équivalente à :

$$\left[B(t,t)^{-1}\frac{\partial B}{\partial t}(t,s)\right]_{1,2} = 0, \ \forall (s,t) \in \mathbf{T},$$

où la fonction matricielle désignée est donc $n_1 \times n_2$. On s'attache ensuite à prouver diverses propriétés (de type invariance fonctionnelle par exemple), puis on mène une étude approfondie du cas $n_3 = 0$ (pas de variable d'environnement) où l'on obtient une caractérisation de la non causalité au sens de Granger en termes de la fonction matricielle A associée à la représentation directe.

On peut alors appliquer ces résultats au cas particulier de solutions d'Equations Différentielles Stochastiques (EDS désormais) à coefficients constants (section 3.4), dont le cadre théorique est présenté par Bergstrom (1984) :

$$dD^{p-1}X(t) = (C_0X(t) + C_1DX(t) + \ldots + C_{p-1}D^{p-1}X(t))dt + \Sigma dW(t) \tag{2}$$

où D désigne l'opérateur de dérivation dans $\mathbb{L}_2$. Comme on connaît dans ce cas la forme exacte du VARMA(p, p-1) réultant de l'observation à temps discret de ces processus, on peut étudier les liens entre la non causalité discrète et la non causalité immédiate, ainsi qu'illustrer l'éventuel problème d'identification du modèle continu d'après données des discrètes.

L'étude d'une classe linéaire régulière qui a été menée au cours du travail sur la causalité amène assez logiquement à étudier plus précisément le cas d'une singularité linéaire. Ce travail existe en temps discret $-$ dans le cadre des processus ARMA $-$, et utilise de nombreuses propriétés particulières des marches aléatoires; il est regroupé, dans le cadre de ses applications à l'économie, sous le terme de "Problèmes de racines unités et co-intégration". La cointégration est l'élargissement de la notion de racine unité au cadre multivarié où des singularités "intermédiaires" peuvent apparaître : c'est mathématiquement un problème d'algèbre linéaire s'appuyant sur les polynômes à coefficients matriciels. Plus précisément, des non stationarités peuvent apparaître dans le cas de polynômes à coefficients matriciels : $\Pi(L) = Id - A_1 L - \ldots - A_p L^p$ pour des modèles de type $\Pi(L)X_t = \varepsilon_t$ où $LX_t = X_{t-1}$, si $\det(\Pi(1)) = 0$. Si $\Pi(1)$ est nul, alors on peut factoriser $(1-L)$ exactement dans $\Pi(L)$ et on parle de "racine unité". Mais il se peut très bien que 1 soit racine du polynôme à coefficients scalaires $\det(\Pi(x))$ sans que $\Pi(1)$ ne soit nul. Il suffit bien sûr pour cela que $\Pi(1)$ soit une matrice singulière. Ainsi, le vecteur de variables n'est pas stationnaire, alors que des combinaisons de ces variables suivant ce que l'on appelle des "vecteurs de cointégration" (i.e. des combinaisons de type $alpha' X_t$, avec α vecteur du noyau de la transposée de $\Pi(1)$) le sont. Économiquement, le concept de cointégration est puissant parce qu'il permet de décrire une relation stationnaire (donc un équilibre) entre des variables qui, individuellement, ne le sont pas. Il est donc important de détecter de telles propriétés, et pour ce faire, de bien les modéliser et d'étudier leurs conséquences.

[6] fait apparaître l'existence de problèmes similaires en temps continu. On y définit et on étudie une notion de cointégration à temps continu, son impact sur les discrétisés et on cherche à la caractériser, notamment dans le domaine spectral.

2.2 Causalité au second ordre : [7], [12]

Lorsqu'on analyse des données très volatiles, c'est souvent à la variance conditionnelle que l'on est amené à s'intéresser, comme en témoignent l'émergence des modèles ARCH et de leurs variantes. Dans de tels cas, on souhaiterait tester la causalité en variance ou au second ordre plutôt que la causalité en moyenne. Néanmoins, celle-ci a été fort peu étudiée, hormis, dans le cas bivarié, quelques tentatives de Granger, Robins et Engle (1986) et de Cheung et Ng (1996). C'est l'objectif de [7], en collaboration avec O. Lieberman, que de mettre en place, d'étudier et de caractériser des définitions de ces notions.
La plupart des résultats vaudraient dans un cadre incluant des variables d'environnement, mais nous les avons supprimées pour alléger les notations. Toutes les variables sont supposées de carré intégrable. On note X_t, $t \in \mathbf{Z}$, un processus n-dimensionnel de composantes $(X_{i,t})_{1 \le i \le 2}$, chacune de dimension n_i, $i = 1, 2$, $n_1 + n_2 = n$ et $X_{i,t} = (X_{ij,t})_{1 \le j \le n_i}$.

On note ici $\mathcal{I}_D^{(1)}(t)$ la σ-algèbre engendrée par X_1 et $\mathbb{E}(.|\mathcal{I}(t))$ l'espérance conditionnelle sachant $\mathcal{I}(t)$. On considère $\overline{X(t)}$, l'espace de Hilbert engendré par les variables $X_{i,s}$, $s \leq t$, $i = 1, 2$ et $\overline{X_1(t)}$ l'espace de Hilbert engendré par les variables dans $X_{1,s}$, $s \leq t$. On note $\Pi(X_{ij,t+1}|\overline{X(t)})$ la projection orthogonale de $X_{ij,t+1}$ sur $\overline{X(t)}$ et $\Pi(X_{i,t+1}|\overline{X(t)})$ le vecteur $(\Pi(X_{i1,t+1}|\overline{X(t)}), \ldots, \Pi(X_{in_i,t+1}|\overline{X(t)}))'$. $\Pi(X_{i,t+1}|\overline{X_1(t)})$ est définie de façon similaire. Notons que $\mathbb{E}(X_{t+1}|\mathcal{I}(t))$ est le meilleur prédicteur de X_{t+1}, alors que $\Pi(X_{t+1}|\overline{X(t)})$ est son meilleur prédicteur linéaire. Rappelons avec nos notations simplifiées :

(i) X_2 ne cause pas au sens de Granger $X_1 (X_2 \overset{G}{\not\to} X_1)$ si $\mathbb{E}(X_{1,t+1}|\mathcal{I}_D(t)) = \mathbb{E}(X_{1,t+1}|\mathcal{I}_D^{(1)}(t))$, $\forall t \in \mathbf{Z}$.

(ii) X_2 ne Granger-cause pas linéairement $X_1 (X_2 \overset{LG}{\not\to} X_1)$ si $\Pi(X_{1,t+1}|\overline{X(t)}) = \Pi(X_{1,t+1}|\overline{X_1(t)})$, $\forall t \in \mathbf{Z}$.

Si X_t est un processus gaussien, les définitions (i) et (ii) sont bien sûr équivalentes. Quelques difficultés apparaissent lorsqu'on veut généraliser ces définitions au second ordre. En effet, les élévations au carré font apparaître des non linéarités et cela accroît la différence entre projection et espérance conditionnelle. Ainsi, alors que les σ-algèbres n'ont pas lieu d'être modifiées, les espaces de projection doivent être adaptés. Soient $\overline{X^2(t)}$ l'espace de Hilbert engendré par les variables $X_{ij,s}X_{kl,s}$, $s \leq t$, $1 \leq i, k \leq 2, 1 \leq j \leq n_i, 1 \leq l \leq n_k$, et $\overline{X_1^2(t)}$ l'espace de Hilbert engendré par $X_{1j,s}X_{1l,s}$, $s \leq t$, $1 \leq j, l \leq n_1$. On peut alors naturellement considérer les définitions suivantes :

Définition 2.3 *(i) X_2 ne cause pas X_1 en variance ($X_2 \overset{V}{\not\to} X_1$), si $V(X_{1,t+1}|\mathcal{I}_D(t)) = V(X_{1,t+1}|\mathcal{I}_D^{(1)}(t))$, $\forall t \in \mathbf{Z}$.*
(ii) X_2 ne cause pas X_1 linéairement en variance ($X_2 \overset{LV}{\not\to} X_1$), si $\forall t \in \mathbf{Z}$,

$$\Pi\left[\left(X_{1,t+1} - \Pi(X_{1,t+1}|\overline{X(t)})\right)^2 |\overline{X^2(t)}\right] = \Pi\left[\left(X_{1,t+1} - \Pi(X_{1,t+1}|\overline{X_1(t)})\right)^2 |\overline{X_1^2(t)}\right].$$

Dans la définition 2.3 (ii) le carré d'un vecteur est mis pour le produit de ce vecteur par sa transposée. On note X' la transposée de X.

Granger, Robins et Engle (1986) ont donné une définition de la non-causalité en variance qui paraît davantage correspondre à la notion de non-causalité au second ordre. On peut l'exprimer de la façon suivante :

Définition 2.4
(i) Il n'y a pas de causalité au second ordre de X_2 sur X_1 ($X_2 \overset{GRE}{\not\to} X_1$) si $\forall t \in \mathbf{Z}$,

$$\mathbb{E}\left[(X_{1,t+1} - \mathbb{E}(X_{1,t+1}|\mathcal{I}(t))^2 |\mathcal{I}_D(t)\right] = \mathbb{E}\left[(X_{1,t+1} - \mathbb{E}(X_{1,t+1}|\mathcal{I}_D(t))^2 |\mathcal{I}_D^{(1)}(t)\right].$$

(ii) Il n'y a pas de causalité linéaire au second ordre de X_2 sur X_1 ($X_2 \overset{LGRE}{\not\to} X_1$) si $\forall t \in \mathbf{Z}$,

$$\Pi\left[\left(X_{1,t+1} - \Pi(X_{1,t+1}|\overline{X(t)})\right)^2 |\overline{X^2(t)}\right] = \Pi\left[\left(X_{1,t+1} - \Pi(X_{1,t+1}|\overline{X(t)})\right)^2 |\overline{X_1^2(t)}\right].$$

On donne ci-dessous les relations d'équivalence entre non-causalités en moyenne, en variance et au second ordre.

Proposition 2.1
$$X_2 \overset{V}{\not\to} X_1 \quad \Leftrightarrow \quad (X_2 \overset{G}{\not\to} X_1 \text{ et } X_2 \overset{GRE}{\not\to} X_1).$$
$$X_2 \overset{LV}{\not\to} X_1 \quad \Leftrightarrow \quad (X_2 \overset{LG}{\not\to} X_1 \text{ et } X_2 \overset{LGRE}{\not\to} X_1).$$

Il résulte de la Proposition 2.1 que si la causalité en moyenne est établie, alors il en est de même de la causalité en variance, alors qu'il peut encore y avoir ou non causalité au second ordre. C'est la raison pour laquelle on se concentre plutôt sur cette dernière notion.
On constate à titre d'exemple que dans les modèles ARMA vectoriels à innovations de type différence de martingales et de variance conditionnelle constante, il y a toujours non-causalité au second ordre. Ainsi, les tests développés par Boudjellaba, Dufour and Roy (1992, Section 4) pour la non-causalité en moyenne vaudraient aussi comme tests de non-causalité linéaire en variance, dans ce cadre particulier.
De plus, on peut donner des conditions suffisantes de non-causalité au second ordre pour des processus à variance conditionnelle dépendant du temps :

Proposition 2.2 *Soit X_t une suite de variables aléatoires avec $V(X_t|\mathcal{I}_D(t-1)) = H_t$, où H_t est un processus matriciel $\mathcal{I}_D(t-1)$-mesurable. Alors*

$$X_2 \overset{GRE}{\not\to} X_1 \Leftrightarrow \left(\mathbb{E}([H_{t+1}]_{1,1}|\mathcal{I}_D^{(1)}(t)) = [H_{t+1}]_{1,1}, \forall t \in \mathbb{N}\right) .$$

On illustre avec des exemples l'utilité aussi bien que les limites de ce résultat. En particulier, on donne des conditions suffisantes de non-causalité dans les modèles GARCH(1,1) bivariés ainsi que dans le modèle ARCH à facteurs proposé par Engle, Ng and Rothschild (1990).
Si le modèle est plus précisément défini, en moyenne et en variance, on est de plus en mesure (section 3 et Appendice B) de fournir des relations d'équivalence pour la non-causalité en moyenne et au second ordre. Ainsi, dans le cas des modèles GARCH(p,q), on obtient une caractérisation de la non causalité linéaire au second ordre en termes des coefficients (matriciels) des polynômes intervenant dans la représentation. On peut aussi ré-exprimer (on le fait dans deux corollaires de notre théorème central) la condition de non causalité sous la forme de condition de nullité de déterminants. L'intérêt de ces résultats est que, si des estimateurs de ces coefficients sont disponibles, on pourra mettre en œuvre des tests de nullité et donc de causalité.
Nous discutons donc en section 4 des tests possibles pour ces contraintes de nullité. Des tests de rapport de vraisemblance sont envisageables à condition que la normalité asymptotique des estimateurs de maximum de vraisemblance pour des GARCH(p,q) multivariés soit prouvée. Celle-ci n'étant établie que dans des cas simples ou univariés, cela explique que le travail [12], en collaboration avec O. Lieberman, soit venu compléter notre étude.
Notons enfin que nous développons (section 5) un exemple empirique en dimension 3 en

considérant des rendements quotidiens d'actions.

On vient de le signaler, les estimateurs de maximum de vraisemblance pour des modèles GARCH(p,q) étaient considérés en pratique, comme consistants et asymptotiquement normaux; mais la théorie avait quelque retard puisque, tant du point de vue des conditions de stationarité que du point de vue des propriétés théoriques des estimateurs, seuls des cas simples avaient été étudiés : Nelson (1990) a donné des conditions de stationarité du GARCH(1,1), Lumsdaine (1996) a prouvé la consistance et la normalité asymptotique de l'estimateur de quasi-Maximum de Vraisemblance dans les GARCH(1,1) univariés sous conditions de moments d'ordre 32, ainsi que Lee and Hansen (1994), sous des conditions légèrement moins fortes; Bougerol and Picard (1992) ont donné une CNS de stationarité du GARCH(p,q) univarié, Boussama (1998) s'est intéressé à la consistance et à la normalité asymptotique de ce même modèle. Jeantheau (1998) s'est intéressé du point de vue statistique au cas multivarié, mais pour donner des critères de consistance et les appliquer à des cas un peu particuliers. [12] s'intéresse à la théorie asymptotique des GARCH(p,q) multivariés.

Soit $(X_t)_{t\in\mathbf{Z}}$ une suite de vecteurs aléatoires de $\mathbb{R}^d$ et soit $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$. On suppose que X_t est de carré intégrable et vérifie :

$$\mathbb{E}(X_t|\mathcal{F}_{t-1}) = 0 \ \text{ a.}s. \tag{3}$$

Soit

$$H_t = \mathbb{E}(X_t X_t'|\mathcal{F}_{t-1}), \tag{4}$$

où X' est la transposée de X. La représentation d'Engle and Kroner (1993), dite repésentation BEKK, du modèle GARCH(p,q) multivarié est donnée par

$$H_t = C + \sum_{i=1}^{q}\left(\sum_{j=1}^{k} A_{ij}X_{t-i}X_{t-i}'A_{ij}'\right) + \sum_{i=1}^{p}\left(\sum_{j=1}^{k} B_{ij}H_{t-i}B_{ij}'\right), \tag{5}$$

où C est définie positive et les matrices A_{ij} et B_{ij} sont des matrices carrées $d \times d$. k est un entier inférieur à $\frac{d(d+1)}{2}$. Le principal avantage de ce modèle est qu'il garantit que H_t est définie positive. On note *vech* l'opérateur qui empile les morceaux de colonnes correspondant à la partie triangulaire inférieure d'une matrice. Alors le modèle (5) peut s'écrire

$$vech(H_t) = vech(C) + \sum_{i=1}^{q}\tilde{A}_i vech(X_{t-i}X_{t-i}') + \sum_{i=1}^{p}\tilde{B}_i vech(H_{t-i}), \tag{6}$$

pour des $\tilde{A}_i$, $\tilde{B}_i$ déduites des A_{ij} et des B_{ij}. Sans perte de généralité, nous prenons $k = 1$.

On note θ le vecteur des paramètres du processus, constitué des coefficients des matrices C, $\tilde{A}_i$ and $\tilde{B}_i$. Le modèle n'est pas supposé Gaussien, mais nous travaillons avec la log-vraisemblance gaussienne. Ainsi, on définit l'estimateur du quasi-maximum de vraisemblance, $\hat{\theta}_n$, comme celui qui maximise :

$$L_n(\theta) = -\frac{1}{2n}\sum_{t=1}^{n}\ell_t(\theta) \text{ avec } \ell_t(\theta) = \ln[\det(H_{t,\theta})] + X_t'H_{t,\theta}^{-1}X_t$$

où $\det(A)$ désigne le déterminant de la matrice A. On notera de plus $\rho(A)$ le rayon spectral de A. Tout d'abord on établit la consistance forte de notre estimateur en utilisant des conditions données par Jeantheau (1998). Soit

$$\varepsilon_t = H_t^{-1/2} X_t, \tag{7}$$

et supposons les ε_t i.i.d. Sans perte de généralité, on choisira $H_t^{1/2}$ symétrique. Soient Θ l'espace de paramètre et $\theta_0 \in \Theta$ la vraie valeur du paramètre. On a alors

Théorème 2.1 *Pour le processus GARCH(p,q) défini par (4)–(7) et pour $\hat{\theta}_n$ défini cidessus, on suppose que :*

1. Θ est compact,

2. le modèle est identifiable au sens de Engle et Kroner (1995),

3. les erreurs ε_t sont i.i.d. et admettent une densité absolument continue par rapport à la mesure de Lebesgue, strictement positive dans un voisinage de l'origine,

4. $\rho(\sum_{i=1}^{q} \tilde{A}_i + \sum_{i=1}^{p} \tilde{B}_i) < 1$.

Alors $\hat{\theta}_n$ est fortement consistant, c'est-à-dire : $\hat{\theta}_n \longrightarrow_{n\to+\infty} \theta_0, \ I\!P_{\theta_0} - p.s.$

La condition $\rho(\sum_{i=1}^{q} \tilde{A}_i + \sum_{i=1}^{p} \tilde{B}_i) < 1$ découle d'un résultat de Boussama (1998) qui établit que cette condition est une CNS d'existence de solution stationnaire ergodique pour le modèle, pour des erreurs vérifiant *2.*
On établit ensuite la normalité asymptotique de $\hat{\theta}_n$ en supposant d'abord les conditions initiales pour H_t dans la loi stationnaire. Basawa, Feigin and Heyde (1976) ont donné des conditions pour obtenir la normalité asymptotique du maximum de vraisemblance pour des processus stochastiques généraux. En utilisant l'ergodicité et la stationarité stricte de notre processus (d'où l'intérêt de notre condition initiale particulière) ainsi qu'un theorème central limite de Billingsley (1968), on se ramène à vérifier des conditions plus simples pour notre contraste. Ainsi, avec des techniques assez standard en statistique paramétrique, on obtient :

Théorème 2.2 *Sous les hypothèses du Théorème 2.1, en supposant de plus que les composantes de ε_t sont indépendantes, que X_t admet des moment d'ordre 8 et que la condition initiale (en H) suit la loi stationnaire,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}}_{n\to+\infty} \mathcal{N}(0, C_1^{-1} C_0 C_1^{-1}), \ \text{sous } I\!P_{\theta_0}.$$

On étend ensuite le résultat au cas d'une condition initiale déterministe quelconque pour H (le lien (7) entre X et H permet d'en déduire la valeur initiale aléatoire en X).

3 Longue mémoire et stationarité

3.1 Modèles à mémoire longue en temps continu : [1], [3], [5]

Le point de départ de [3], en collaboration avec E. Renault, est l'étude d'une singularité fonctionnelle dans les représentation moyenne mobile en temps continu; cela a débouché

sur des travaux assez complets, sur un thème qui a connu depuis des développements très fructueux, en calcul stochastique (Coutin-Decreusefond (1999), Carmona-Coutin (1999)) aussi bien qu'en Finance (Hu, Oksendal et Sulem (1999)) où les données hautes fréquences se prêtent parfaitement à la fois aux modélisations (comportement longue mémoire, modèles à temps continu) et aux procédures d'estimation nécessitant des jeux de données de grande longueur.
Il s'agissait d'envisager des processus $X(t) = \int_0^t A(t-s)dW(s)$ de même type que dans [2], avec à présent :

$$A(x) = \frac{x^\alpha}{\Gamma(1+\alpha)}\tilde{A}(x),\ \tilde{A}\ \ C^1 \text{ sur } [0,T] \text{ ou } \mathbb{R}^+,\ |\alpha| < \frac{1}{2}. \tag{8}$$

Dans de tels cas, soit $A(0)$ n'existe pas $(-1/2 < \alpha < 0)$, soit $A(0) = 0$ $(0 < \alpha < -1/2)$. Cela signifie que l'on ne dispose plus de la décomposition semi-martingale comme dans le cas régulier étudié dans [3], et que la représentation n'est pas inversible. On constate que pour $\tilde{A} = Id$, on obtient un processus :

$$W_\alpha(t) = \int_0^t \frac{(t-s)^\alpha}{\Gamma(1+\alpha)}dW(s)$$

connu sous le nom de Brownien Fractionnaire d'ordre α, auquel on va chercher à faire jouer le rôle tenu habituellement par le Mouvement Brownien ([2]). Il faut néanmoins prendre garde que l'on n'a plus alors de propriétés de martingale ni de caractère markovien, ce qui ne simplifie ni les raisonnements ni les calculs; ces propriétés sont essentiellement remplacées par la "self-similarité" (X est H-self-similaire si $\forall a > 0,\ X(at)$ et $a^H X(t)$ ont mêmes lois de probabilité).

Le type de modèle ainsi envisagé apparaît comme l'analogue continu de modèles discrets, les ARIMA(p,α,q), $\alpha \in \mathbb{R}$, encore applés ARMA fractionnaires d'ordre α :

$$(1-L)^\alpha \Phi(L)X_t = \Theta(L)\varepsilon_t,\ \ \mathrm{d}^o\Phi = p,\ \ \mathrm{d}^o\Theta = q,\ \ (1-L)^\alpha = \sum_{j=0}^{\infty} \frac{\Gamma(j-\alpha)}{\Gamma(j+1)\Gamma(-\alpha)}L^j \tag{9}$$

développés par Granger-Joyeux (1980) et Hosking (1981). On va étudier ces processus du point de vue des représentations ainsi que des éventuelles propriétés caractérisant la mémoire longue, en gardant en tête les paradigmes apportés par le temps discret.

Ainsi, on étudie (section 2.1) sous quelles conditions on peut donner un sens dans $\mathbb{L}_2$ à des processus $Y(t) = \int_0^t D(t-s)dX(s)$ pour $X(t) = \int_0^t A(t-s)dW(s)$; cela permet en particulier de définir $\int_0^t D(t-s)dW_\alpha(s)$. Ainsi, pour X défini par (8), on peut chercher une représentation (section 2.2) :

$$X(t) = \int_0^t C(t-s)dW_\alpha(s)$$

[2]Le Brownien fractionnaire est parfois défini aussi sous la forme

$$B_H(t) - B_H(0) = \int_{-\infty}^0 \frac{(t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})}dW(s) + \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})}dW(s)$$

où le dernier terme correspond à $W_{H-1/2}(t)$, afin d'obtenir des propriétés de stationarité.

et étudier les liens entre A et C, qui sont de type dérivation/intégration d'ordre α. De plus, si C est de classe C^1 et $C(0)$ inversible, on dispose d'une inversion :

$$W_\alpha(t) = \int_0^t D(t-s)dX(s).$$

Ces deux représentations permettent de faire apparaître une décomposition de nos processus qui n'est plus semi-martingale, mais de type : brownien fractionnaire + prévisible. Enfin on définit une dérivation fractionnaire (section 3) $X^{(\alpha)}$ des processus X de type (8) qui met en évidence une dualité de représentations :

$$X(t) = \int_0^t C(t-s)dW_\alpha(s), \;\; X^{(\alpha)}(t) = \int_0^t C(t-s)dW(s)$$

qui apparaît comme l'analogue d'un résultat prouvé par Geweke et Porter-Hudak (1983). Ils établissent, en considérant les densités spectrales, une équivalence entre la classe des ARIMA(p, α, q) donnés par (9) et la classe des processus

$$\tilde{\Phi}(L)X_t = \tilde{\Theta}(L)\varepsilon_\alpha(t), \;\; \varepsilon_\alpha(t) = W_\alpha(t) - W_\alpha(t-1),$$

qui sont des processus de type ARMA par rapport à un "bruit fractionnaire", ε_α. La puissance de la modélisation en temps continu est de donner de plus une propriété d'invariance fonctionnelle (la fonction C est la même dans les deux représentations) alors qu'en temps discret, on ne sait rien des liens analytiques entre (Φ, Θ) et $(\tilde{\Phi}, \tilde{\Theta})$; le processus ε_α est en effet dépourvu de "bonnes propriétés" et dépend considérablement du pas de temps choisi.

Le cadre "mémoire longue" est fourni (section 2.3) par le sous-domaine $\alpha \in]0, 1/2[$ assorti d'une condition de stationarité asymptotique :

$$\lim_{x \to +\infty} x\tilde{A}(x) = A_\infty \neq 0. \tag{10}$$

Sous (10) et pour $\alpha \in]0, 1/2[$, on démontre que la version stationnaire de X encore notée X :

$$X(t) = \int_{-\infty}^t \frac{(t-s)^\alpha}{\Gamma(1+\alpha)} \tilde{A}(t-s)dW(s)$$

vérifie:

$$\lim_{h \to +\infty} h^{1-2\alpha}\gamma_X(h) = c.A_\infty \, {}^tA_\infty$$

où γ_X est la fonction d'autocovariance ($\gamma_X(h) = \mathrm{cov}(X(t), X(t+h))$), et c une constante, ainsi que :

$$\lim_{\lambda \to 0} \lambda^{2\alpha} f_X(\lambda) = \tilde{c}A_\infty \, {}^tA_\infty$$

où f_X est la densité spectrale continue ($f_X(\lambda) = \int e^{-i\lambda h}\gamma_X(h)dh$). Ces propriétés sont exactement celles qui caractérisent habituellement le phénomène de longue mémoire. Enfin, on peut vérifier (section 3.1) que les conditions exigées par Robinson (1996) pour une estimation semi-paramétrique convergente et asymptotiquement normale de α sont satisfaites par f_X^Δ :

$$f_X^\Delta(\lambda) = \frac{1}{\Delta} \sum_{n \in \mathbf{Z}} f_X\left(\frac{\lambda + 2\pi n}{\Delta}\right), \tag{11}$$

où f_X est définie comme précédemment; f_X^Δ est la densité spectrale du processus discret X observé à pas Δ, si l'on suppose que l'on dispose de telles observations.

Un exemple fertile de processus du type étudié est fourni par la généralisation, présentée en section 3.1, des EDS à coefficients constants (2) en EDS "fractionnaires" où l'on remplace le Brownien par un Brownien fractionnaire : si $Y(t)$ solution de (2) admet la représentation $Y(t) = \int_0^t C(t-s)dW(s)$, alors $X(t) = \int_0^t C(t-s)dW_\alpha(s)$ est considéré comme la solution de l'EDS fractionnaire :

$$dD^{p-1}X(t) = (C_0X(t) + C_1DX(t) + \ldots + C_{p-1}D^{p-1}X(t))dt + \Sigma dW_\alpha(t) \tag{12}$$

Les formules générales nous donnent l'expression de la densité spectrale :

$$f_X(\lambda) = \frac{1}{\lambda^{2\alpha}} \left[(i\lambda)^p Id - \sum_{k=0}^{p-1} C_k(i\lambda)^k\right]^{-1} \Sigma\, {}^t\Sigma \left[(i\lambda)^p Id - \sum_{k=0}^{p-1} C_k(i\lambda)^k\right]^{-1*} \tag{13}$$

La section 3.2 est consacrée à une application en Finance. Plus précisément, grâce à des modèles de ce type, on peut étendre la théorie de la structure par terme des taux d'intérêt. On fait alors apparaître des phénomènes de mémoire dans les taux à terme et les taux forward vis à vis de la maturité, ce qui est conforme à ce qui est observé.

Un autre exemple s'appuyant sur des EDS d'ordre 1 illustre enfin (section 3.3) que des comportements à mémoire longue peuvent être obtenus par agrégation d'équations individuelles à mémoire courte.

[1] propose d'étudier par des expériences de type Monte Carlo deux méthodes d'estimation des paramètres de processus du type précédent. Pour ce faire, on présente les modèles auxquels on souhaite s'intéresser ainsi que quelques unes des propriétés théoriques que l'on souhaite illustrer. Puis on se propose d'étudier des méthodes de simulation. Le modèle envisagé est une solution d'EDS fractionnaire d'ordre 1 et scalaire :

$$dx(t) = a_0x(t)dt + \sigma dw_\alpha(t), \quad a_0 < 0$$

où le paramètre θ à estimer s'écrit donc $\theta = (a_0, \alpha, \sigma)'$. Bien entendu, l'essentiel des méthodes utilisées ici se généralisent à l'ordre supérieur. Nous allons chercher à simuler $x(h), x(2h), \ldots, x(mh)$ avec h petit, afin de constituer un échantillon d'"observations" $(x(\Delta), x(2\Delta), \ldots, x(N\Delta)$ avec $\Delta = ph$ (en général, $\Delta = 10h$).

La première idée (section 3.1) consiste à exploiter le fait que l'on a un processus gaussien stationnaire et à utiliser une méthode proposée par Davis et Harte (1987) utilisant la représentation spectrale du processus.

On met ensuite en place des schémas de discrétisation de type schéma d'Euler (section 3.2); le premier est fondé sur la représentation $x(t) = \int_0^t a(t-s)dw(s)$:

$$\tilde{x}(ph) = \sum_{k=1}^{p} a((p-k)h)\sqrt{h}\varepsilon_{(h)}(k), \quad \text{où } \Delta w(kh) = \sqrt{h}\varepsilon_{(h)}(k)$$

pour lequel on montre que pour tout $t \in [0,T]$, $\mathrm{V}(x(t)-\tilde{x}(t)) \leq C.Th^{2\alpha} = C.mh^{2\alpha+1}$, $T = mh$.

On vérifie alors un critère de tension donné par Billingsley (1968) pour la suite :

$$\tilde{x}_n(t) = \sum_{k=1}^{[nt]} a(t-\frac{k}{n})\Delta w(\frac{k}{n})$$

qui implique la convergence faible fonctionnelle : $\tilde{x}_n \Rightarrow x$, $n \to +\infty$.

La seconde discrétisation envisagée (section 3.2 toujours) se fonde sur le fait que $x^{(\alpha)}$, dérivée d'ordre α de x, suit une EDS du premier ordre habituelle, donc admet comme discrétisé exact un AR(1) qui s'écrit :

$$x^{(\alpha)}(t+h) = e^{a_0 h}x^{(\alpha)}(t) + \sigma\sqrt{\frac{e^{2a_0 h}-1}{2a_0}}\varepsilon(t+h).$$

Le schéma d'Euler doit donc être appliqué cette fois pour intégrer $x^{(\alpha)}$ à l'ordre α, $x(t) = \int_0^t \frac{(t-s)^\alpha}{\Gamma(1+\alpha)}dx^{(\alpha)}(s)$:

$$\overline{x}(ph) = \sum_{p=1}^{p} \frac{((p-k)h)^\alpha}{\Gamma(1+\alpha)}\Delta x^{(\alpha)}(kh).$$

$\mathrm{V}(x(t)-\overline{x}(t))$ est du même ordre que précédemment et même :

$$\overline{x}_n \Rightarrow x, \text{ avec } \overline{x}_n(t) = \sum_{k=1}^{[nt]} \frac{(t-\frac{k}{n})^\alpha}{\Gamma(1+\alpha)}\Delta x^{(\alpha)}(\frac{k}{n}).$$

C'est cette dernière méthode qui est finalement retenue.

Ces considérations amènent une question naturelle : s'il faut utiliser un filtre fractionnaire discret pour simuler les données, et peut-être pour les estimer, pourquoi ne pas se contenter du filtre fractionnaire usuel, correspondant aux modèles ARMA à mémoire longue et à l'opérateur $(1-L)^\alpha$? La section 3.3 s'attache à répondre à cette question. Elle met en évidence que le filtre $(1-L)^{-\alpha}$, appliqué à notre processus x usuel, restituerait bien un processus stationnaire à mémoire courte; néanmoins ce processus n'aurait aucune raison d'être un AR(1), et serait *a priori* un MA(∞). Un tel filtre ne respecterait donc pas la structure court-terme du processus considéré. Une illustration empirique par calcul des autocorrélations partielles moyennes sur 50 échantillons complète les arguments théoriques.

On aborde ensuite le problème de l'estimation. La première méthode (section 4.1), proposée par Fox-Taqqu (1986) et Dahlhaus (1989), cherche à minimiser en θ l'approximation de la vraisemblance dans le domaine des fréquences (estimateur de type Whittle) :

$$\mathcal{L}_n^\Delta(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\ln f_{X,\theta}^\Delta(x) + \frac{I_n(x)}{f_{X,\theta}^\Delta(x)}\right)dx$$

où $I_n(x)$ est le périodogramme et $f_{X,\theta}^\Delta$ toujours donnée par la formule (13) pour $p=1$.

L'autre méthode testée (section 4.2) consiste à utiliser les résultats semi-paramétriques de Robinson (1996) déjà cités, pour l'estimation de α : il s'agit d'une amélioration de la régression du log-périodogramme sur $\ln\lambda$ pour λ petit (exploitant $f^1_{X,\theta}(\lambda) \sim d.\lambda^{2\alpha}$, $\lambda \to 0$), proposée à l'origine par Geweke et Porter-Hudak (1983). L'idée est alors que par dérivation approchée à l'ordre approché :

$$\hat{x}^{(\hat{\alpha})}(t) = \sum_{k\Delta<t} \frac{(t-k\Delta)^{\hat{\alpha}}}{\Gamma(1+\hat{\alpha})}(x(k\Delta) - x((k-1)\Delta)),$$

on obtient une approximation de $x^{(\alpha)}$, que l'on asssimile à un AR(1), ce qui permet d'estimer c_0 et σ. Les performances de cette dérivation approchée ont été testées dans la première partie de l'étude et se sont avérées très convenables.

Les résultats d'estimation (section 4.3) par méthode de Monte Carlo sont assez bons pour a_0 et α, mais l'estimation de σ est biaisée inférieurement. Enfin, la seconde méthode est incomparablement plus rapide que la première pour des résultats similaires. La différence en faveur de la première méthode se ferait probablement sur des échantillons plus grands.

Dans [5], en collaboration avec E. Renault, nous étudions un modèle d'actif financier dit "à volatilité stochastique" :

$$\begin{cases} \dfrac{dS(t)}{S(t)} = \mu(t, S(t))dt + \sigma(t)dw^1(t) \\ d(\ln\sigma(t)) = k(\theta - \ln\sigma(t))dt + \gamma dw^2_\alpha(t) \end{cases} \tag{14}$$

où $S(t)$ désigne le prix de l'actif sous-jacent, $\sigma(t)$ sa volatilité instantanée, $(w^1(t), w^2(t))$ est un Brownien bidimensionnel standard et w^2_α est le brownien fractionnaire obtenu par l'intégration *ad hoc* de w_2. Pour $\alpha = 0$, c'est-à-dire sans aspect fractionnaire, ces modèles ont été étudiés par Renault et Touzi (1996) qui ont en particulier mis en évidence qu'ils rendaient compte de l'effet "smile" observé sur les marchés. Ils ont été proposés dans leur première version pour rendre compte d'un aléa spécifique à la volatilité par Hull-White (1987), Scott (1987) et Melino et Turnbull (1990). [5] cherche à étendre ces modèles pour rendre compte de plus de la persistance de la volatilité constatée en pratique.

En effet, pour les données haute fréquence, on trouve des indices de comportements proches des racines unités dans le processus de variance conditionnelle. En temps discret, de nombreux modèles de type ARCH avec racine unité ont été ainsi étudiés. Dans ce travail, on introduit de la mémoire longue dans le processus de volatilité modélisé en temps continu en remplaçant simplement le brownien standard w_2 du modèle à volatilité stochastique par le brownien fractionnaire w^2_α, avec α restreint à $0 \leq \alpha < \frac{1}{2}$ (au lieu de $|\alpha| < \frac{1}{2}$ autorisé par la définition générale parce que la longue mémoire n'intervient que sur ce domaine). Bien sûr pour α non nul w^2_α n'est plus une semi-martingale (cf. Rogers (1997)), et la théorie de l'intégration stochastique usuelle n'est plus valable. Mais, en procédant comme dans [3], on n'utilise que la théorie $\mathbb{L}_2$ et l'intégrale de Wiener pour l'intégration de processus gaussiens. De plus, les prix d'options, eux, et bien que fonctions du processus de volatilité, conservent la propriété de semi-martingale (grâce à l'intégration), ainsi que le processus de prix. Cette propriété de semi-martingale est cruciale pour les prix d'actifs car les processus non semi-martingales n'admettent pas de mesure martingale équivalente. On sait en effet d'après Delbaen et Schachermayer (1994) que l'absence d'opportunité d'arbitrage entraîne l'existence d'une mesure martingale équivalent (i.e. d'une probabilité risque-neutre). Signalons que nous sommes néanmoins confrontés au problème, habituel avec ces modèles, de la non unicité de la probabilité risque-neutre équivalente (marché incomplet).

Ainsi, nous étudions tout d'abord dans [5] (section 2) les propriétés probabilistes de notre modèle de Volatilité Stochastique Fractionnaire. Ensuite nous donnons en section 3 les bases nécessaires à une étude empirique en étudiant des approximations discrètes de ces modèles. Cela permet de mettre en évidence les avantages de la modélisation en temps continu par rapport à ses analogues discrets tels que les FIGARCH (Fractionally Integrated GARCH) rencontrés dans la littérature. La section 4 établit une formule de pricing d'options et étudie les caractéristiques des volatilités implicites. En particulier, la mémoire longue modifie le comportement des coefficients d'hétéroscédasticité conditionnelle et de kurtosis par rapport à la date d'échéance. Enfin, on s'intéresse en section 5 à l'aspect statistique du problème. Si la volatilité instantanée était observée, les méthodes proposées dans [3] seraient applicables. A défaut, on peut songer à utiliser la méthode proposée par Pastorello, Renault and Touzi (1999) dans le contexte du modèle de Hull et White (1987). Mais comme, pour $\alpha \neq 0$, le processus de volatilité n'est plus Markovien, l'utilisation de la formule de pricing d'option devient beaucoup moins maniable. De façon alternative, on suggère aussi d'utiliser des discrétisations approchées de $S(t)$ pour obtenir des approximations des volatilités instantanées et de travailler avec des vraisemblances approchées. Pour finir, on utilise en section 6 les discrétisations de σ calculées en section 3 pour réaliser quelques simulations visant à mettre en évidence les propriétés théoriques précédemment étudiées et à comparer les procédures d'estimation suggérées.

3.2 Régression sur le log périodogramme : [4]

La méthode étudiée dans [4], en collaboration avec C. Hardouin, n'a pas donné lieu à publication. Pourtant, les documents de travail des études sous-jacentes sont parfois encore citées dans certains articles d'Économétrie et Statistique (Velasco (2000)).
L'estimation par régression sur le log périodogramme a été proposée à l'origine par Kashyap et Eom (1988) et Geweke et Porter-Hudak (1983), mais dans le cadre fractionnaire de processus ARIMA(p, d, q) avec $-\frac{1}{2} < d < 0$. Cependant, dans le cas Gaussien, les estimateurs obtenus ne sont pas efficaces. Nous avons constaté dans [4], en reprenant cette méthode dans le cadre usuel (densité spectrale bornée) que nous obtenions la consistance, la normalité et l'efficacité asymptotiques, à condition de "régulariser" au préalable le périodogramme. Cette procédure convient pour une grande classe de modèles et est tout particulièrement adaptée à des modèles faisant intervenir des paramètres en exposant. Nous pensons par exemple aux processus MADE de Gonçalves [3] : transformation d'un bruit blanc par des filtres du type $\prod_j (1 - \alpha_j L)^{d_j}$, où L est l'opérateur retard

usuel ($LX_t = X_{t-1}$). Les cas $|\alpha_j| < 1, d \in \mathbb{R}$ appartiennent à notre classe de modèles. La procédure d'estimation est alors très rapide puisque l'équation de régression devient linéaire en d_j. Néammoins, elle nécessite dans tous les cas un grand nombre d'observations. On considère le modèle suivant :

$$X_t = A(L)\ \varepsilon_t,\ \ (\varepsilon_t) \text{ i.i.d. } (0,1),\ A(L) = \sum_{j\geq 0} a_j L^j,\ \ a_0 = 1,\ \ \sum_{j\geq 0} a_j^2 < \infty. \tag{15}$$

observé sur $t = 1, \cdots, n$, n pair. Soient $X_n(\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^n X_t\ e^{it\lambda}$, $\varepsilon_n(\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \varepsilon_t\ e^{it\lambda}$ les transformées de Fourier discrètes normalisées de X et ε observés sur $\{1, \ldots, n\}$, et $I_{nX}(\lambda) = |X_n(\lambda)|^2$, $I_{n\varepsilon}(\lambda) = |\varepsilon_n(\lambda)|^2$ les périodogrammes de X et ε. Nous définissons le pseudo-périodogramme de X : $J_{nX}(\lambda) = |A(e^{i\lambda})|^2\ I_{n\varepsilon}(\lambda) = 2\pi F(\lambda) I_{n\varepsilon}(\lambda)$, où F est la densité spectrale de X. Alors, l'équation de régression s'écrit :

$$\ln J_{nX}(\lambda) = \ln 2\pi F(\lambda)\ +\ \ln I_{n\varepsilon}(\lambda) \tag{16}$$

La procédure "estime" les paramètres en considérant la régression (16) aux fréquences $\lambda_k = \frac{2\pi k}{n}$, $k = 1, \ldots \frac{n}{2} - 1$. C'est l'idée proposée par Kashyap et Eom (1988) pour un ARIMA$(0, d, 0)$, à notre régularisation et au remplacement de J par I près.

Dans [4], nous étudions au paragraphe 2 l'estimation de modèles paramétriques gaussiens (15) basée d'abord sur le pseudo-périodogramme. Pour obtenir l'efficacité, nous remplaçons le pseudo-périodogramme J_{nX} par sa version "régularisée" et considérons la régression (16) modifiée en conséquence. Nous pouvons alors utiliser des résultats sur l'approximation de I_{nX} par J_{nX} pour étudier la procédure de régression basée sur le vrai périodogramme (paragraphe 3). Cette étape est cruciale puisque J_{nX} n'est pas observable. Enfin, nous montrons au paragraphe 4 que notre méthode et celle de Whittle (1953) sont asymptotiquement équivalentes. Des extensions au cadre fractionnaire ou au cadre non gaussien (thème très en vogue aujourd'hui) étaient alorsabordées dans les preprints.

3.3 Estimation adaptative de la densité spectrale d'un processus stationnaire gaussien : [10]

La connaissance approfondie de la densité spectrale ainsi acquise explique que j'aie songé à appliquer dans ce cadre, de façon fructueuse, les méthodes d'estimation adaptative en train de faire leur chemin en statistique, et en particulier la méthode de sélection de modèles exposée ci-après.

Je considère dans [10] le problème de l'estimation non paramétrique de la densité spectrale f d'une suite réelle gaussienne stationnaire à partir d'un échantillon $X = (X_1, \ldots, X_n)$.

$$f(\omega) = \frac{1}{2\pi} \sum_{k\in Z} \gamma_k e^{i\omega k},\ \omega \in [-\pi, \pi],$$

où $\gamma_k = \text{cov}(X_t, X_{t+k})$, $t \in \mathbf{Z}$. Si $\hat{f}$ est un estimateur de f, on mesure le risque de $\hat{f}$ via la norme $\mathbb{L}_2$:

$$\mathbb{E}\|f - \hat{f}\|^2 = \mathbb{E} \int_{-\pi}^{\pi} \left(f(\omega) - \hat{f}(\omega)\right)^2 d\omega,$$

où $\|g\|$ désigne la norme $\mathbb{L}_2$ usuelle : $\|g\|^2 = \int_{-\pi}^{\pi} g^2(\omega)d\omega$.

[10] estime f grâce à un estimateur par projection et en choisissant le meilleur espace de projection possible. Dans ce but, on introduit une collection d'espaces linéaires de dimension finie $\{S_m, m \in \mathcal{M}_n\}$ (les S_m sont appelés modèles) et nous associons à chaque S_m l'estimateur par projection $\hat{f}_m$ de f sur cet espace, i.e. $\hat{f}_m$ minimise, pour tous les t dans S_m, la fonction de contraste:

$$\gamma_n(t) = \int_{-\pi}^{\pi} t^2(\omega)d\omega - 2\int_{-\pi}^{\pi} t(\omega)I_n(\omega)d\omega$$

où

$$I_n(\omega) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n}(X_t - \bar{X}_n)e^{i\omega t}\right|^2, \quad \bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t. \tag{17}$$

Sous des hypothèses adéquates, on montre aisément que, en notant s une constante dépendant de f, le risque de $\hat{f}_m$ est borné par :

$$\mathbb{E}\|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + s\frac{\dim(S_m)}{n} \tag{18}$$

où f_m est la projection $\mathbb{L}_2$-orthogonale de f sur S_m.
Le but poursuivi est de proposer une procédure de sélection de modèles fondée sur les données seules, et capable de choisir un $\hat{m}$ dans $\mathcal{M}_n$ de façon à ce que l'estimateur par projection $\hat{f}_{\hat{m}}$ réalise une performance presque aussi bonne que le meilleur de tous les $\hat{f}_m$ de la collection, cette performance étant mesurée en terme de risque $\mathbb{L}_2$. La procédure de sélection considérée est un critère pénalisé :

$$\hat{m} = \text{Argmin}_{m\in\mathcal{M}_n}[\gamma_n(\hat{f}_m) + \text{pen}(m)],$$

où pen est une fonction de pénalité de $\mathcal{M}_n$ dans $\mathbb{R}^+$. Bien sûr le problème central réside ici dans la détermination de la pénalité. Au vu de (18), le type de performance recherchée et obtenue s'écrit :

$$\mathbb{E}\|f - \hat{f}_{\hat{m}}\|^2 \leq C \inf_{m\in\mathcal{M}_n}\left[\|f - f_m\|^2 + \|f\|_\infty^2 L_m \frac{\dim(S_m)}{n}\right] \tag{19}$$

où C est une constante multiplicative. On trouve $L_m = 1$ ou $\ln^2(n)$ selon la collection de modèles considérée. Ainsi lorsque $L_m = 1$, notre estimateur minimise le risque sur toute la collection, à une constante multiplicative près : cela correspond au cadre des modèles réguliers **(T)**, **(RP)**, **(W)** décrits en appendice. On trouve $L_m = \ln(n)$ pour les collections générales (**(GP)**, **(GW)**).
La méthode de sélection de modèles ainsi que les techniques utilisées dans les preuves sont inspirées par les travaux de Birgé et Massart (1997, 1998) et Barron, Birgé et Massart (1999). Le problème de l'estimation non paramétrique de la densité spectrale a été abordé par Bentkus et Rudzkis (1976) et Bentkus (1985) avec des méthodes non adaptatives (la régularité de f doit être connue a priori). Ils établissent néanmoins l'ordre des vitesses

minimax des estimateurs sur des classes de régularité de f. Golubev (1993) propose une procédure pour un périodogramme pondéré, adaptive du point de vue du choix des poids. Son estimateur est asymptotiquement minimax sur des familles de voisinages de la vraie densité spectrale, sous des hypothèses sur f légèrement plus forte que celles requises dans ce travail. Enfin Neumann (1996) considère un estimateur par projection pondéré dans un contexte non gaussien. Ses conditions sur f sont données pour l'essentiel en terme des cumulants de la suite (X_i). Il utilise une méthode de seuillage à seuil fixe (mais inconnu) puis aléatoire inspirée des travaux de Donoho et Johnstone (1994, 1998), et Donoho, Johnstone, Kerkyacharian, Picard (1996), et il atteint la vitesse optimale usuelle sur les Besov à un facteur $\ln(n)$ près. La méthode adaptative utilisée ici permet de regagner ce facteur $\ln(n)$. De plus nos résultats sont non asymptotiques.
Disons quelques mots des hypothèses de ce travail. On suppose que les X_i vérifient :

$(\mathbf{H}_X)$ La suite $(X_1, \ldots, X_n)$ est une suite stationnaire de variables aléatoires gaussiennes.

De plus, on fait l'hypothèse standard suivante sur f, qui entrîne sa bornitude :

$(\mathbf{H}_f)$ La fonction d'autocovariance est telle que $\sum_{k\in Z} |\gamma_k| = M < +\infty$ et $\sum_{k\in Z} |k\gamma_k^2| = M_0 < +\infty$.

Soit ensuite $(S_m, m \in \mathcal{M}_n)$ une collection de modèles tels que les S_m sont des espaces linéaires de fonctions, de dimension D_m. On peut donner une formulation générales des hypothèses qu'ils doivent vérifier mais disons seulement pour simplifier qu'elles sont vérifiées par tous les modèles décrits en Appendice de ce mémoire.

En utilisant des méthodes de chaînage $\mathbb{L}_2$-$\mathbb{L}_\infty$, nous prouvons notre théorème principal pour des modèles généraux (i.e. non nécessairement réguliers), en supposant d'abord que la borne de f est connue. L'inégalité de Bernstein au départ de la preuve utilise de façon cruciale le caractère gaussien du processus, de sorte qu'une extension au cadre non gaussien ne semble pas immédiate.

Théorème 3.1 *Supposons* $(\mathbf{H}_X)$ *et* $(\mathbf{H}_f)$ *vérifiées et considérons une famille de modèles généraux vérifiant* $\sum_{m\in\mathcal{M}_n} e^{-L_m D_m} \leq \Sigma < +\infty$, *pour des poids positifs* L_m. *Alors l'Estimateur par Projection Pénalisé (EPP) défini par :*

$$\tilde{f} = \hat{f}_{\hat{m}} \text{ où } \hat{m} = \operatorname{argmin}_{m\in\mathcal{M}_n}\left(\min_{t\in S_m}\int_{-\pi}^{\pi} t(\omega)(t(\omega) - 2I_n(\omega))d\omega + \operatorname{pen}(m)\right) \qquad (20)$$

avec

$$\operatorname{pen}(m) \geq \kappa \|f\|_\infty^2 (1 + C_{\bar{r}}^2)(1 + L_m)^2 D_m/n$$

où κ est une constante numérique (universelle) et $C_{\bar{r}}$ dépend de la collection de modèle, vérifie

$$\mathbb{E}(\|f - \tilde{f}\|^2) \leq C \inf_{m\in\mathcal{M}_n}\left[\|f - f_m\|^2 + \operatorname{pen}(m)\right] + 8\frac{M_1}{n} + C'\frac{(1 + C_{\bar{r}}^2)\Sigma\|f\|_\infty^2}{n} \qquad (21)$$

où C et C' sont des constantes universelles.

Bien sûr, le principal problème du Théorème 3.1 réside dans le fait que la pénalité (déterministe) dépend d'une quantité inconnue : $\|f\|_\infty$. On la remplace donc par un estimateur, ce qui nous amène à considérer une pénalité aléatoire en remplaçant $\|f\|_\infty$ par la norme infinie de $\hat{f}_n^P$, la projection orthogonale (empirique) du périodogramme sur un espace de polynômes par morceaux réguliers (famille **(P)**) de dimension dépendant de n et notée N_n. On note f_n^P la projection orthogonale $\mathbb{L}_2$ de f sur ce même espace. On obtient alors une extension du résultat (21) précédent, sous les hypothèses du théorème précédent, et si $\|f - f_n^P\|_\infty \leq (1/4)\|f\|_\infty$ et $N_n \leq \frac{1}{36(r+1)^2}\frac{n}{\ln(n)}$, pour l'EPP $\tilde{f}$ défini par (20) avec

$$\mathrm{pen}(m) \geq \kappa \|\hat{f}_n^P\|_\infty^2 (1 + C_{\bar{r}}^2)(1 + L_m)^2 D_m/n.$$

Bien sûr, la condition $\|f - f_n^P\|_\infty \leq (1/4)\|f\|_\infty$ doit être étudiée. Si f est supposée vérifier des conditions de régularité, alors elle est satisfaite d'après des résultats de théorie de l'approximation. En effet si $f \in \mathcal{B}_{\beta,p,\infty}$, $\beta > 0$, on sait, d'après Devore and Lorentz (1993) que pour $\beta > 1/p$,

$$\|f - f_n^P\|_\infty \leq C(\beta)|f|_{\beta,p} N_n^{-(\beta - 1/p)}$$

où les définitions sont rappelées brièvement plus bas. Cela implique une contrainte $N_n \geq C(\beta, f)$ qui est satisfaite quand n est assez grand.

On peut affiner ces résultats dans le cadre spécifique des bases trigonométriques, particulièrement adaptées pour des fonctions développables en séries de Fourier. Ainsi, dans ce cadre, on peut prolonger les résultats de type (19) à des densités spectrales non bornées, tant qu'elles restent de carré intégrable. De plus, les espaces **(T)** dans le cas général voyaient leur dimension maximale restreinte à être inférieure à $\sqrt{n}$ alors que dans ce nouveau cadre, on atteint $n/\ln(n)$. Ces résultats nécessitent des calculs plus précis en terme de densité spectrale mais s'appuient sur des inégalités de déviations plus directes, obtenues sans chaînage, d'où leur attractivité.

Enfin, illustrons comment on peut dériver, à partir d'inégalités du type de (21), des résultats d'adaptation au degré de régularité inconnu de f. Rappelons simplement qu'une fonction f appartient à l'espace de Besov $\mathcal{B}_{\beta,p,\infty}([0,1])$ si elle vérifie

$$|f|_{\beta,p} = \sup_{y>0} y^{-\beta} w_d(f,y)_p < +\infty, \quad d = [\beta] + 1,$$

où $w_d(f,y)_p$ désigne le module de régularité ("smoothness"). Pour une définition précise, voir DeVore et Lorentz (1993), Chapitre 2, Section 7, où il est de plus démontré que $\mathcal{B}_{\beta,p,\infty}(A) \subset \mathcal{B}_{\beta,2,\infty}(A)$ pour $p \geq 2$. Cela justifie que nous restreignions notre attention à $\mathcal{B}_{\beta,2,\infty}(A)$.

Si les S_m sont l'une des trois familles **(T)**, **(RP)** et **(W)**. Il résulte de résultats classiques en théorie de l'approximation que, si $r \geq \beta$ et $f \in \mathcal{B}_{\beta,2,\infty}([-\pi,\pi])$,

$$\|f - f_m\| \leq C(\beta)|f|_{\beta,2} D_m^{-\beta}, \tag{22}$$

où r est la régularité. Voir par exemple Barron, Birgé, Massart (1999) Lemma 12. p.404. Voir aussi Birgé et Massart (2000). Ainsi on obtient :

Corollaire 3.1 *Supposons* $(\mathbf{H}_X)$ *et* $(\mathbf{H}_f)$ *vérifiées et supposons que* $f(2\pi.)$ *appartient à un espace* $\mathcal{B}_{\beta,2,\infty}([0,1]) \cap \mathbb{L}_\infty([0,1])$*. Supposons de plus que* $\{S_m\}_{m\in\mathcal{M}_n}$ *est l'une des familles* **(P)** *ou* **(W)** *avec* $\beta > 1/2$*. Soit* $\tilde{f}$ *l'EPP défini par (20). Alors*

$$\mathbb{E}\left[\|f-\tilde{f}\|^2\right] \leq C|f|_{\beta,2}^{\frac{2}{1+2\beta}} n^{-\frac{2\beta}{2\beta+1}}, \tag{23}$$

où la constante C *ne dépend que de* $\beta, M_1, \|f\|_\infty$ *et* $C_{\bar{r}}$*.*

En d'autres termes

$$\sup_{f\in B_{\beta,2}(R)} \mathbb{E}\left[\|f-\tilde{f}\|^2\right] \leq Cn^{-\frac{2\beta}{2\beta+1}},$$

où la constante C dépend ici de $\beta, R, M_1, \|f\|_\infty$ et $\bar{r}$ et où $B_{\beta,2}(R) = \{f \in \mathcal{B}_{\beta,2,\infty}, |f|_{\beta,2} \leq R\}$.
On sait d'après Bentkus (1986) que le risque minimax sur une boule Besov $\{f/|f|_{\beta,2} \leq L\}$ ne peut pas être plus petit que $C'L^{\frac{2}{1+2\beta}}n^{-\frac{2\beta}{2\beta+1}}$ où C' est une constante positive. Par conséquent, notre estimateur est minimax (à des constantes près) sur toutes ces boules simultanément.

4 Régressions et β-mélange

4.1 Modèles de régression : [8], [9], [13]

Dans les articles [8], [9] et [13], on étudie le modèle suivant:

$$X_{i+1} = b(X_i) + \sigma(X_i)\varepsilon_{i+1}, \tag{24}$$

où les ε_i sont des variables aléatoires i.i.d. centrées et de variance 1. Ce modèle est un cas particulier du modèle de régression standard :

$$Y_i = b(X_i) + \sigma(X_i)u_i, \tag{25}$$

où les u_i sont i.i.d. centées réduites mais où les (X_i, Y_i) ne sont pas supposées indépendantes; au contraire elles sont autorisées à être β-mélangeantes. [8] et [9], en collaboration avec Y. Baraud et G. Viennet, cherchent à estimer de façon adaptative b en supposant $\sigma(x) \equiv \sigma_2$ où la variance σ_2^2 est supposée connue. [13], en collaboration avec Y. Rozenholc, se place dans le cadre général et cherche à estimer de façon adaptative à la fois les fonctions b et σ^2.
Il y a eu de nombreuses études concernant la sélection de modèles, fondées sur des critères pénalisés de type C_p de Mallows, tels que les critères d'Akaike ou BIC, pour des modèles régressifs ou autorégressifs (Akaike (1973), Shibata (1976), Li (1987), Polyak et Tsybakov (1992), parmi d'autres). Tous les résultats sont de nature asymptotique. L'extension

de ces critères pénalisés à des procédures de sélection de modèles basées sur les données ont été effectuées par Barron (1991), Barron et Cover (1991), Rissanen (1984), puis plus récemment et dans le cadre d'une démarche dépassant le contexte de la simple régression, par Birgé et Massart (1997) et Barron, Birgé et Massart (1999). Le point de vue de ces derniers est non asymptotique. Néanmoins leur généralité même les amène à imposer des restrictions non souhaitables (fonction de régression bornée, par exemple) à leur modèle. [8] et [9] relaxent ces restrictions tout en continuant à donner des résultats non asymptotiques; ils s'appuient sur des résultats obtenus en "fixed design" par Baraud (2000) et des méthodes de couplage développées dans Viennet (1997). [13] pose le problème de l'estimation de la moyenne en présence de volatilité ainsi que l'estimation de la volatilité σ^2. Yao et Tong (1994) et Härdle et Tsybakov (1997) ont étudié l'estimation de σ^2 de façon adaptative mais ils ne se préoccupent pas d'être simultanément adaptatifs vis à vis de la moyenne. Hall et Carroll (1989) utilisent des estimateurs analogues aux nôtres, mais dans un cadre non adaptatif. Les résultats auxquels nous nous comparons le plus volontiers sont ceux de Fan et Yao (1998), du point de vue théorique (mais leurs hypothèses de régularité sont fortes et leurs résultats asymptotiques) aussi bien que du point de vue des performances empiriques (comparaison des résultats de simulation). Citons enfin Hoffman (1999) qui obtient dans un cadre de seuillage en ondelettes des résultats asymptotiques correspondant aux vitesses minimax en moyenne et à un facteur $\ln(n)$ près en volatilité.
Reprenons notre modèle (24). Si $\hat{f}$ est un estimateur de f, où f vaut b ou σ^2, alors on mesure le risque de $\hat{f}$ par le biais de la norme $\mathbb{L}_2$-empirique :

$$\mathbb{E}[\|\hat{f} - f\|_n^2] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(\hat{f}(X_i) - f(X_i))^2\right].$$

Le choix de cette mesure de risque est longuement discutée dans [8] : disons, de façon schématique que le minimisateur de $(1/n)\sum_{i=1}^{n}(X_{i+1} - t(X_i))^2$, pour $t \in \mathbb{L}_2(\mathbb{R}, dx)$ sur un espace linéaire $S \subset \mathbb{L}_2(\mathbb{R}, dx)$, noté $\hat{f}_S$, peut ne pas être unique, alors que le vecteur $(\hat{f}_S(X_1), \ldots, \hat{f}_S(X_n))$ lui l'est toujours. De plus, sous des hypothèses adaptées, ce risque peut être décomposé en une somme biais + variance via:

$$\mathbb{E}[\|f - \hat{f}_S\|_n^2] \leq \|f - f_S\|_\mu^2 + \kappa\frac{\dim(S)}{n}, \tag{26}$$

où f_S est la projection $\mathbb{L}_2(dx)$-orthogonale de f sur S, $\|t\|_\mu^2 = \mathbb{E}(t^2(X_1))$ et κ dépend de constantes du problème.
Voici donc la procédure que nous utilisons. Elle est analogue à chaque étape à celle utilisée dans [10]. Nous considérons une ou deux collections finies de modèles notées $\{S_m^{(i)}, m \in \mathcal{M}_n^{(i)}\}$, $i = 1$ pour b et $i = 2$ pour σ^2; chaque $S_m^{(i)}$ est un sous-espace de dimension finie de $\mathbb{L}_2(\mathbb{R})$. Bien sûr, on ne demande pas que b ni σ^2 appartienne à aucun de ces modèles. Soit $\hat{b}_m$ l'estimateur des moindres carrés de b sur $S_m^{(1)}$ associé au contraste :

$$\gamma_n^{(1)}(t) = \frac{1}{n}\sum_{i=1}^{n}(X_{i+1} - t(X_i))^2$$

pour le modèle (24) basé sur les observations $X_1, \ldots, X_{n+1}$. Nous utilisons une procédure qui choisit $\hat{m}_1$ dans $\mathcal{M}_n^{(1)}$ et qui réalise un bon compromis entre les termes de biais et de variance de la façon suivante :

$$\mathbb{E}\left[\|b - \hat{b}_{\hat{m}_1}\|_n^2\right] \leq C \min_{m \in \mathcal{M}_n} \left\{\|b - b_m\|_\mu^2 + \frac{\dim(S_m^{(1)})}{n}\right\}, \tag{27}$$

où b_m est la projection orthogonale de b sur $S_m^{(1)}$ et C est une constante multiplicative dépendant des données du problème. Cela signifie que, à cette constante multiplicative C près, l'estimateur choisit le meilleur modèle de la collection. On obtient ce résultat en choisissant $\hat{m}_1$ minimisant $\gamma_n^{(1)}(\hat{b}_m) + \mathrm{pen}^{(1)}(m)$ parmi tous les m dans $\mathcal{M}_n^{(1)}$, où $\mathrm{pen}^{(1)}$ est une fonction à déterminer et est appelée fonction de pénalité.
Dans un second temps, $\hat{\sigma^2}$ est estimée de façon analogue à partir du contraste :

$$\gamma_n^{(2)}(t) = \frac{1}{n} \sum_{i=1}^{n} \left[X_{i+1}^2 - \hat{b}_{\hat{m}_1}^2(X_i) - t(X_i)\right]^2,$$

pour $t \in S_m^{(2)}$, avec un objectif similaire à (27), b étant remplacée par σ^2. On utilise là une fonction de pénalité $\mathrm{pen}^{(2)}(m)$.
On trouve une ([8]) ou deux ([13]) fonctions de pénalités $\mathrm{pen}^{(i)}(m)$, $i = 1, 2$ d'ordre $\dim(S_m^{(i)})/n$. Les modèles envisagés dans [8] et [13] sont exclusivement réguliers. Dans le cadre général, [9] fait apparaître une pénalité en $L_m \dim(S_m^{(1)})/n$ où (cf. [10]) L_m est d'ordre $\ln(n)$ (cadre général) ou 1 (cadre régulier).
Il est important de noter que des estimateurs vérifiant des inégalités telles que (27) ont des propriétés intéressantes vis à vis des familles de modèles que nous avons en tête (polynômes par morceaux, ondelettes, polynômes trigonométriques). En particulier, de tels estimateurs sont adaptatifs au sens du minimax sur de nombreuses classes de régularité (cf. Barron *et al.* (1999), Birgé et Massart (1997)) et on peut passer, comme exposé en section 3.3 pour [10], d'inégalités de type (27) à des Corollaires analogues au Corollaire 3.1, et à des inégalités de type (23). Cela signifie que pour $L_m = 1$, on obtient les vitesses minimax dans les Besov.
Plus précisément [8] se place dans un cadre d'erreurs ε admettant des moments d'ordre $p > 2$ et de variables X_i (autorégression) ou (X_i, Y_i) (régression) géométriquement ou arithmétiquement β-mélangeantes. La fonction de variance σ est constante. L'obtention d'inégalités de type (27) nécessite alors un contrôle précis et délicat de

$$\mathbb{E}\left[\sup_{t \in B_{m'}(0,1)} \sum_{i=1}^{n} \varepsilon_i t(X_i) - \mathbb{E}\left(\sup_{t \in B_{m'}(0,1)} \sum_{i=1}^{n} \varepsilon_i t(X_i)\right)\right]_+^p$$

où $B_{m'}(0,1)$ est la boule unité de $S_{m'}$, ainsi qu'une gestion détaillée du mélange s'appuyant sur le lemme de Berbee et les méthodes d'approximation des variables mélangeantes mises en place dans Viennet (1997). Les contrôles du risque ne sont obtenus que dans le cas de modèles réguliers. On illustre grâce à quelques exemples détaillés le compromis à réaliser entre la vitesse de mélange, la dimension maximale autorisée pour les espaces linéaires d'estimation et l'ordre p des moments exigé.

[9] reste dans un cadre où la fonction de variance σ est constante et où les variables X_i ou (X_i, Y_i) sont géométriquement ou arithmétiquement β-mélangeantes. Mais à présent, on suppose les erreurs sous-gaussiennes i.e. telles que, pour tout u réel,

$$\mathbb{E}[\exp(u\varepsilon_1)] \leq \exp\left(\frac{u^2 s^2}{2}\right)$$

pour un s positif ($s = \sigma_2$ si les variables sont gaussiennes ou correspond à la borne pour des variables bornées). On peut, à ce prix, obtenir des résultats dans le cadre de collections de modèles générales. Notons que les techniques mises en œuvre dans les preuves pour les inégalités de déviation sont très différentes de celles utilisées dans [8] puisqu'on s'appuie désormais sur la propriété de martingale de la suite $\sum_{i=1}^n \varepsilon_i t(X_i)$ pour établir une inégalité de Bernstein donnant lieu à un chaînage $\mathbb{L}_2$. De plus, on développe dans [9] des inégalités d'adaptation au sens du minimax précises et détaillées (en particulier dans les Besov $\mathcal{B}_{\beta,p,\infty}$ pour $p < 2$).
Enfin [13] étend les résultats précédents en présence d'une fonction de volatilité et estime σ^2 dans une seconde étape. On se place sous des conditions de moments d'ordre p suffisants sur les erreurs ε: $p > 8$ ou $p > 16$ selon les cas et on suppose le β-mélange des X_i géométrique.
Dans la première étape, on a

$$\tilde{b} := \hat{b}_{\hat{m}_1} \text{ où } \hat{m}_1 = \operatorname{argmin}_{m \in \mathcal{M}_n^{(1)}} \left(\gamma_n^{(1)}(\hat{b}_m) + \operatorname{pen}^{(1)}(m)\right) \tag{28}$$

et

$$\operatorname{pen}^{(1)}(m) = \kappa \Phi_1^2 \hat{r}_n^2 \frac{D_m^{(1)}}{n}. \tag{29}$$

où κ est une constante numérique, Φ_1^2 une constante liée à la collection de modèles et

$$\hat{r}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_{i+1} - \hat{b}_{m_n}(X_i))^2. \tag{30}$$

$S_{m_n}^{(1)}$ est un espace de la famille de modèles de dimension $D_{m_n}^{(1)}$ à préciser. Ainsi on établit que sous des conditions standard, si p est assez grand (disons $p \geq 8$), alors $\tilde{b}$ défini par (28)-(29) vérifie

$$\mathbb{E}[\|b\mathbf{1}_A - \tilde{b}\|_n^2] \leq C \inf_{m \in \mathcal{M}_n^{(1)}} \left(\|b\mathbf{1}_A - b_m\|_\mu^2 + \frac{D_m^{(1)}}{n}(\|b\|_\mu^2 + \|\sigma\|_\mu^2)\right) + \frac{R}{n}$$

où C est une constante universelle et R est une constante qui dépend des données du problème, à condition que $\hat{r}_n^2$ soit calculée sur un espace de dimension inférieure a $n^{1/2-2/p}$. A est le compact sur lequel on souhaite effectuer l'estimation. b_m désigne la projection

$\mathbb{L}_2(dx)$-orthogonale de b sur S_m.
En seconde étape, pour le contraste

$$\gamma_n^{(2)}(t) = \frac{1}{n}\sum_{i=1}^{n}\left[X_{i+1}^2 - \tilde{b}^2(X_i) - t(X_i)\right]^2 \tag{31}$$

on montre que, si p est assez grand (disons $p \geq 16$), alors l'estimateur pénalisé, noté $\widetilde{\sigma^2}$ vérifie

$$\mathbb{E}[\|\sigma^2\mathbf{1}_A - \widetilde{\sigma^2}\|_n^2] \leq C \inf_{m\in\mathcal{M}_n^{(2)}}\left(\|\sigma^2\mathbf{1}_A - \sigma_m^2\|_\mu^2 + R\frac{D_m^{(2)}}{n}\right) + \frac{R'}{n} + C'\mathbb{E}[\|b^2\mathbf{1}_A - \tilde{b}^2\|_n^2] \tag{32}$$

où C et C' sont des constantes universelles, et R et R' des constantes qui dépendent des données du problème.
La première étape amenait à la vitesse minimax pour $b \in \mathcal{B}_{\alpha,2,\infty}(A)$. Ici, en utilisant un nouveau contrôle de $\mathbb{E}[\|b^2\mathbf{1}_A - \tilde{b}^2\|_n^2]$ on obtient alors des vitesses minimax pour σ^2 à condition que b appartienne à un espace de Besov $\mathcal{B}_{\alpha,2,\infty}(A)$ et σ^2 à $\mathcal{B}_{\beta,2,\infty}(A)$ avec $\alpha > 2\beta+3/4$. Notons que Hall and Carroll (1989) ont obtenu la même vitesse : $n^{-2\beta/(2\beta+1)}$ mais dès que $\alpha > 1/2$. Néanmoins ils sont dans un cadre non adaptatif, de régression avec variables indépendantes.

Une importante étude par des simulations a de plus été menée dans [13] afin d'étudier la qualité des estimateurs, dans un cadre de polynômes par morceaux de degré inférieur à 3, le degré étant automatiquement choisi sur chaque sous-intervalle par l'algorithme. On illustre ainsi les bons résultats de la méthode de sélection de modèles. L'enjeu est d'importance puisqu'il s'agit de valider la faisabilité et les bonnes performances de la sélection de modèles dans un cadre régulier pour commencer, et de démarrer la mise en place d'une boîte à outils susceptible de fournir aux statisticiens appliqués et aux économètres une nouvelle méthode d'estimation souple, efficace et "prête à l'emploi". Les premières expériences ont été menées en estimation de densité par histogrammes par Birgé et Rozenholc (2000). L'étude menée ici complète la mise en place dans le cadre de polynômes par morceaux réguliers. Pour l'avenir, il reste encore à implémenter un algorithme capable d'affronter le cadre irrégulier dans des temps raisonnables.

4.2 Modèles de régression avec erreurs dans les variables

Je me suis tout d'abord intéressée à la version discrète des modèles à volatilité stochastique étudiés à temps continu dans [2], section 2.4 puis dans [5]. Dans le cadre du temps discret, ces modèles sont un cas particulier des modèles de régression avec erreurs dans les variables :

$$X_t = m(X_{t-1}) + \eta_t \tag{33}$$
$$Y_t = X_t + \varepsilon_t \tag{34}$$

avec ε_t i.i.d. $(0, \sigma_\varepsilon^2)$, et η_t i.i.d. $(0, \sigma_\eta^2)$, les deux suites étant indépendantes. Seules les Y_t sont observées et on souhaite estimer m. De la même façon que Fan and Truong (1993), on prend une fonction m de classe C^k sur $[a,b]$, $a<b$ et un noyau K d'ordre k (i.e. tel que $\int K(y)dy = 1, \int y^j K(y)dy = 0, j = 1,\ldots,k-1, \int y^k K(y)dy \neq 0$). On suppose que $\Phi_\varepsilon(t) = \mathbb{E}(e^{it\varepsilon}) \neq 0, \ \ \forall t \in \mathbb{R}$. Alors on considère l'estimateur :

$$\hat{m}_n(x) = \frac{\sum_{j=1}^n K_n((x-Y_j)/h_n)Y_{j+1}}{\sum_{i=1}^n K_n((x-Y_i)/h_n)} = \frac{1}{nh_n}\sum_{j=1}^n \frac{K_n((x-Y_j)/h_n)Y_{j+1}}{\hat{f}_n(x)} \tag{35}$$

où:

$$\hat{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^n K_n(\frac{x-Y_i}{h_n}) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\Phi_K(th_n)\frac{\hat{\Phi}_n(t)}{\Phi_\varepsilon(t)}dt, \ \ \hat{\Phi}_n(t) = \frac{1}{n}\sum_{j=1}^n e^{itY_j},$$

Φ_K désigne la transormée de Fourier de K, $\Phi_\varepsilon(t) = \mathbb{E}(e^{it\varepsilon})$ et

$$K_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\frac{\Phi_K(t)}{\Phi_\varepsilon(t/h_n)}dt.$$

On se place dans le cadre de X_i stationnaires, de loi $d\mu(x) = f_X(x)dx$, où f_X est strictement minorée sur $[a,b]$ et admet k dérivées bornées. De plus, la suite $(X_i)_{i\in\mathbb{N}}$ est supposée absolument régulière avec des coefficients de β-mélange β_k vérifiant $\sum_{k\geq 0}(k+1)^c\beta_k < \infty$ pour un $c>0$. Enfin, nous considérons comme Fan and Truong (1993) ou Masry (1993) deux classes d'erreurs ε : les "Super Smooth d'ordre α" notées SSE(α) à décroissance en $\exp(-|t|^\alpha)$ à l'infini et les "Ordinary Smooth d'ordre α" notées OSE(α) à décroissance en $|t|^{-\alpha}$ à l'infini.
Pour des erreurs *super smooth*, Fan and Truong (1993) obtenu des vitesses optimales pour leur estimateur d'ordre $\ln(n)^{k/\alpha}$. Leurs résultats ont été étendus à un modèle de régression en contexte α-mélangeant par Masry (1993) sans perte de vitesse. C'est donc sans surprise que l'on conserve cette vitesse pour le modèle d'autorégression. Pour des erreurs *ordinary smooth*, nous conservons encore les vitesses optimales obtenues par Fan and Truong (1993) en régression indépendante : $n^{-k/[(2k+\alpha)+1]}$. Mais cette fois notre résultat est meilleur que celui de Masry (1993) qui avait une perte d'un facteur $\ln(n)$ dans la vitesse en α-mélange. C'est le contexte de β-mélange et l'utilisation des Inégalités de Viennet (1997) qui nous permet ce gain. On le voit, il est souhaitable dans ce cadre de travailler avec des erreurs ordinary smooth. C'est pourquoi l'on étudie ensuite, dans le modèle à volatilité stochastique défini par $h_t = \exp(X_t/2)\xi_t$ et (33) (qui amène à l'équation (34) en posant $Y_t = \ln(h_t^2) - \mathbb{E}(\ln(\xi_1^2))$ et $\varepsilon_t = \ln(\xi_t^2) - \mathbb{E}(\ln(\xi_1^2)))$, quelles lois d'erreurs ξ_t permettent d'obtenir des ε *ordinary smooth*.

Un autre angle pour améliorer les vitesses, notamment en *super smooth* (par exemple dans le cas d'erreurs gaussiennes) est de se placer dans un cadre où la fonction m est mieux spécifiée. C'est pourquoi [14], en collaboration avec M.L. Taupin, cherche à estimer $m = m_\theta$ dans un cadre semi-paramétrique. Deux types de résultats pré-existent dans un cadre analogue : ceux de Taupin (1998), qui se place dans un cadre indépendant avec erreurs gaussiennes et de Chanda (1995), qui se place dans le cas dépendant particulier

d'un modèle ARMA(p,q) avec erreurs dans les variables. Bien entendu, nous sommes ici dans le prolongement du travail de Taupin (1998) dont nous étendons l'estimateur et le cadre d'hypothèses.
Si la loi des X_i était connue, on étudierait le critère suivant :

$$\widetilde{S}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} W(Y_{i-1})\left(Y_i - \mathbb{E}(m_\theta(X_i)|Y_{i-1})\right)^2, \tag{36}$$

où $W(.)$ est une fonction de poids à support compact notée S_W. En effet, la connaissance de f_X et f_ε permet d'utiliser la formule :

$$\mathbb{E}(m_\theta(X)|Y) = \frac{\int m_\theta(x) f_\varepsilon(Y-x) f_X(x)dx}{\int f_\varepsilon(Y-x) f_X(x)dx} := \frac{\Gamma_{m_\theta}(Y)}{h(Y)} := \Phi_\theta(Y), \tag{37}$$

où $f_Y = f_\varepsilon * f_X$, $*$ désignant le produit de convolution, est la densité de Y. Ainsi, le critère (36) est associé à l'estimateur :

$$\widetilde{\theta} = \text{argmin}_{\theta\in\Theta}\widetilde{S}_n(\theta). \tag{38}$$

On vérifie à titre préliminaire la consistance et la normalité asymptotique de cet estimateur sous les conditions

$$\sum_{k\geq 0} k\beta_k < +\infty \text{ et } \mathbb{E}[|(Y_1 - \Phi_\theta(Y_0))\Phi_{\theta^0}^{(1)}(Y_0)|^4] < \infty, \tag{39}$$

Dans le cas général $\mathbb{E}(m_\theta(X_i)|Y_{i-1})$ n'est pas calculable et on la remplace par un estimateur déduit de la formule (37).

D'abord on s'occupe du numérateur. On note $g^\star$ la transformée de Fourier de g, $g(x) = \int g(u)e^{ixu}du$, et

$$T_{m_\theta,y}(x) = m_\theta(x) f_\varepsilon(z-x).$$

Comme

$$\Gamma_{m_\theta}(y_0) = \frac{1}{2\pi} < T^\star_{m_\theta,y_0}, f^\star_X >= \frac{1}{2\pi} < T^\star_{m_\theta,y_0}/f^\star_\varepsilon, f^\star_Y >,$$

on prend

$$\widehat{\Gamma}_{m_\theta}(y_0) = \frac{1}{2\pi n}\sum_{j=1}^{n} Re < T^\star_{m_\theta,y_0}/f^\star_\varepsilon, h^\star_{nj} > \text{ avec } h_{nj}(t) = e^{itY_j}K^\star_n(t) \tag{40}$$

où $Re(z)$ désigne la partie réelle de z et K_n est défini par $K_n(x) = C_n K(C_n x)$. K est un noyau supposé vérifier des hypothèses simples. Par exemple, le noyau naïf $K(x) = \sin(x)/(\pi x)$ vérifie ces hypothèses ainsi que :

$$V(x) = \frac{\cos(x) - \cos(2x)}{\pi x^2}. \tag{41}$$

Pour le dénominateur, on choisit de façon standard

$$\widehat{f}_Y(y_0) = \frac{1}{n}\sum_{i=1}^{n} V_n(y_0 - Y_j) \tag{42}$$

où $V_n(x) = h_n^{-1}V(x/h_n)$, $h_n^{-1} \to +\infty$ reste à choisir et $V(x)$ est défini par (41). Ainsi, l'estimateur global est $\widehat{\theta}$:

$$\widehat{\theta} = \mathrm{argmin}_{\theta\in\Theta} S_n(\theta) \tag{43}$$

où

$$S_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} W(Y_{i-1})[Y_i - \widehat{\phi}_\theta(Y_{i-1})]^2, \text{ avec } \widehat{\phi}_\theta(Y_{i-1}) = \frac{\widehat{\Gamma}_\theta(Y_{i-1})}{\widehat{f}_Y(Y_{i-1})}$$

où $\widehat{\Gamma}_\theta$ est défini par (40) et $\widehat{f}_Y$ est défini par (42).

Une étude détaillée du numérateur et du dénominateur nous permet alors de montrer que $\widehat{\theta}$ est consistant, pour des erreurs *ordinary* ou *super smooth*, que le mélange soit géométrique ou arithmétique $\beta_k \leq Mk^{-\rho}$ avec $\rho > 2$.

Le calcul de la vitesse est plus délicat. Néanmoins, on l'exprime de façon synthétique en fonction de quelques quantités ainsi définies. Pour un fonction g donnée, on note $\bar{\mathbb{I}}_{C_n}(t) = \mathbb{I}_{|t|\geq C_n}$ et $\mathbb{I}_{C_n}(t) = \mathbb{I}_{|t|\leq C_n}$ et on pose :

$$\lambda_n(g,y) = \mathbb{E}\left|\left\langle T^\star_{g,y}(f^\star_\varepsilon)^{-1}, h^\star_{n,j}\right\rangle\right|^2, \sigma_n^2(g) = \sup_{y\in S_W}\lambda_n(g,y), B_n(g) = \sup_{y\in S_W}\|T^\star_{g,y}\bar{\mathbb{I}}_{C_n}\|_1$$

$$M_n(g) = C_n \sup_{y\in S_W}\sup_{|t|\leq C_n}|T^\star_{g,y}(t)(f^\star_\varepsilon)^{-1}(t)|.$$

Ainsi, Sous quelques hypothèses techniques (mais standard) de régularité de m_θ, en supposant les conditions de consistance remplies et en imposant quelques contraintes : $M_n^2(m_\theta) = o(n)$, $M_n^2(m_\theta^{(1)}) = o(n)$, et pour $\rho > 2$, $\alpha > 2$, ou $\alpha > 11/5$ en mélange arihtmétique et pour des erreurs *Ordinary smooth*, on a :

$$\widehat{\theta} - \theta^0 = O_p\left[\delta_{1,n}(m_{\theta^0}) + n^{-1/2} + \delta_{2,n}(m_{\theta^0})\right],$$

où

$$\delta_{1,n}(m_{\theta^0}) = n^{-1/2}\mathbb{E}^{1/2}\left\{F_n^2(Y_2^\star)\right\} + B_n(m_{\theta^0}),$$

$$\delta_{2,n}(m_{\theta^0}) = O_p\left(\frac{\sigma_n^2(m_\theta)}{n} + \frac{\sigma_n^2(m_\theta^{(1)})}{n}\right)$$

où $(Y_j^\star)_j$ est une suite de variables indépendantes telles que $Y_j^\star$ ait la même loi que Y_j, où $F_n(Y_2^\star)$ est une espérance conditionnelle parfois évaluable et dans tous les cas telle que $\mathbb{E}^{1/2}(D_n^2(Y_2^\star)) \leq \sigma_n(m_\theta)$.
En d'autres termes, on a, de façon moins précise,

$$\widehat{\theta} - \theta^0 = O_p\left[B_n(m_{\theta^0}) + \frac{\sigma_n(m_{\theta^0})}{\sqrt{n}}\right].$$

Globalement, notre théorème central fournit donc une méthode de calcul de vitesse, que l'on peut mettre en oeuvre assez commodément dès que l'on se donne un type assez précis de fonction de régression m_θ. Les quantités à évaluer sont clairement désignées.

Le résultat est illustré dans quelques exemples : $m_\theta(x) = \sum_{k=0}^{p}\theta_k x^k$, $\sum_{k=1}^{p}\theta_k^{(1)}e^{\theta_k^{(2)}x}$ en régression et $e^{-\theta x^2}$ $0 < \theta < 1/e$ en autorégression, $m_\theta(x) = \sum_{k=1}^{p}\theta_k\cos(2\pi kx)$. On réalise des gains de vitesse importants par rapport notamment aux vitesses logarithmiques du non paramétrique avec erreur SSE.

APPENDICE
Exemples de collections de modèles

On considèrera trois sortes de familles de modèles $(S_m)_{m\in\mathcal{M}_n}$.

(T) Polynômes Trigonométriques : On considère les espaces S_m engendrés par les fonctions $\varphi_0(x) = 1/\sqrt{2\pi}$, $\varphi_{2j}(x) = \sqrt{1/\pi}\cos(jx)$, $\varphi_{2j+1}(x) = \sqrt{1/\pi}\sin(jx)$ pour $j = 0, \ldots, d_m$, où $D_m = 2d_m + 1$ est la dimension de S_m. La famille $\mathcal{M}_n$ est dans ce cas l'ensemble des dimensions possibles : $\mathcal{M}_n = \{1, \ldots, n\}$. $\mathcal{S}_n$ est engendré par tous les φ_j, $2j + 1 \leq n$.

(RP) Polynômes par morceaux Réguliers : on considère la partition de $[0, 1]$, $\mathcal{I}_m = \{[j/m, (j + 1)/m), j = 0, \ldots m - 1\}$ et un entier $r \geq 0$ donné. On définit alors S_m comme l'espace des polynômes par morceaux de degré inférieur ou égal à r sur la partition $\mathcal{I}_m$ Alors $D_m = (r + 1)m$. L'espace associé à la partition la plus fine est noté $\mathcal{S}_n$. On note **(DP)** les espaces à noeuds diadiques (i.e. aux point $j/2^m$).

(GP) Polynômes par morceaux Généraux (de degré $\leq r$). Ces espaces sont engendrés par des polynômes de degré au plus r sur tout sous-intervalle de la grille dont l'ensemble des noeuds est n'importe quel sous-ensemble de la grille la plus fine : $\{\pi j/2^{J_n}, j = 0, 1, \ldots, J_n\}, (r + 1)2^{J_n} \leq n$ i.e. $J_n = [\ln(n/r)/\ln(2)]$ en diadique.

(W) Ondelettes à support compact régulières. voir Daubechies (1992), Chapitre 9.3.

(GW) Ondelettes générales à support compact : pour la description de ces espaces, on réfère à Barron, Birgé and Massart (1999), Section 3.2.1.

RÉFÉRENCES

Adenstedt, R.K. (1974) On large sample estimation for the mean of a stationary sequence. *Ann. Statist.*, 2, 1095-1107.

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. *Proceedings 2nd International Symposium on Information Theory*, P.N. Petrov and F. Csaki (Eds.), Akademia Kiado, Budapest, 267-281.

Anderson, T.W. (1971) *The Statistical Analysis of Time Series*, Wiley.

Amemiya, T. (1985), *Advanced Econometrics*, Blackwell.

Baraud, Y. (2000) Model selection for regression on a fixed design, *Probab. Theory Relat. Fields*, 117, 467-493.

Barron, A.R. (1991) Complexity regularization with application to artificial neural networks, *Proceedings NATO Advanced Study Institute on Nonparametric Functional estimation*, G.Roussas, Ed., Dordrecht, The Netherlands: Kluwer, 561-576.

Barron, A.R., Birgé, L. et Massart, P. (1999) Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields*, 113, 301-413.

Barron, A.R. et Cover, T.M. (1991) Minimum complexity density estimation. *IEEE;Trans. Inform. Theory*, 37, 1034-1054.

Basawa, I.V., Feigin, P.D. and Heyde, C.C. (1976) Asymptotic properties of maximum likelihood estimators for stochastic processes. *Sankhya*, 38, Series A, Pt. 3, 259–270.

Bentkus, R. (1985) Rate of uniform convergence of statistical estimators of spectral density in spaces of differentiable functions. *Lith. Math. J.*, 26, 209-219.

Bentkus, R. and Rudzkis, R. (1976) Large deviations for the spectral estimate of a stationary Gaussian sequence. *Lith. Math. J.*, 16, 519-529.

Bergstrom, A.R. (1984). Continuous time stochastic models and
ssues of aggregation over time, in *Handbook of Econometrics*, vol.II, chap.20, eds Z. Griliches and M. Intrilligator, Amsterdam, North-Holland.

Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley.

Birgé, L. et Massart, P. (1997) From model selection to adaptive estimation, in *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics* (D. Pollard, E. Torgersen and G. Yang, eds), 55-87, Springer-Verlag, New-York.

Birgé, L. et Massart, P. (1998) Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3), 329-375.

Birgé, L. et Massart, P. (2000) An adaptive compression algorithm in Besov spaces. *Constructive Approximation*, 16, 1-36.

Birgé, L. et Rozenholc, Y. (2000) How many bins should be put in a regular histogram. *Working paper.*

Birman, M.S. and Solomjak, M.Z. (1967) Piecewise polynomials approximation of functions of the classes W_p. *Mat. Sbornik* 73, 295-317.

Boudjellaba, H., Dufour, J. M., and Roy, R. (1992) Testing Causality Between Two Vectors in Multivariate Autoregressive Moving Average Models. *Journal of the American Statistical Association*, 87, 420, 1082–1090.

Bougerol, P. et Picard, N. (1992) Stationarity of GARCH Processes and Some Nonnegative Time Series. *Journal of Econometrics*, 52, 115–127.

Boussama, F. (1998), "Ergodicity, Mixing and Estimation in GARCH Models", Ph.D. dissertation, University Paris 7.

Brockwell, J.B. et Davis, R.A. (1991), *Time Series : Theory and Methods*, 2nd edition, Springer Verlag.

Carmona, P. et Coutin, L. (1999) Stochastic integration with respect to fractional Brownian motion. Preprint Univ. Toulouse-Paul Sabatier.

K.C. Chanda (1995) Large sample analysis of autoregressive moving average models with errors-in-variables. *Journal of Time Series Analysis*, 16, 1, 1-15.

Cheung, Y. W., et Ng, L. K. (1996) A Causality in Variance Test and Its Application to Financial Market Prices. *Journal of Econometrics*, 72, 33–48.

Coutin, L. et Decreusefond, L. (1999) Stochastic differential equations driven by a fractional Brownian motion. *to appear in Stochastic Processes and Their Applications.*

Dacuhna-Castelle, D. et Duflo, M. (1993) *Probability and Statistics*, Vol.2, Springer Verlag.

Dahlhaus, R. (1989) Efficient parameter estimation for self-similar processes. *Ann. Statist.*, 17, 1749-1766.

Daubechies, I. (1992) *Ten lectures on wavelets*, SIAM: Philadelphia.

Davies, R.B. (1973) Asymptotic inference in stationary Gaussian time-series, *Adv. in Appl. Probab.*, 5, 469-497.

Davies, R.B. and Harte, D.S. (1987) Tests for the Hurst effect. *Biometrika*, 74(1), 95-101.

Delbaen, F. and Schachermayer, W (1994) A general version of the fundamental theorem of asset pricing. *Math Annals* 300, 3, 463-520.

DeVore, R.A. and Lorentz, G.G. (1993) *Constructive approximation.* Springer-Verlag.

Donoho, D.L. and Johnstone, I.M. (1994) Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81, 425-455.

Donoho, D.L. and Johnstone, I.M. (1998) Minimax estimation via wavelet shrinkage. *Ann. Statist.*, 26, 879-921.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996) Density estimation by wavelet thresholding. *Ann. Statist.*, 24, 508-539.

Engle, R. F., and Kroner, K. F. (1995), "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, 122–150.

Engle, R. F., Ng, V. K., and Rotschild, M. (1990), "Asset Pricing with a Factor-ARCH Covariance Structure", *Journal of Econometrics*, 45, 213-237.

J. Fan and Y.K. Truong (1993) Nonparametric regression with errors-in-variables. *Ann. Statist.* **21**, 4, 1900-1925.

Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85, 3, 645-660.

Florens, J.P. and Fougère, D. (1996). Non causality in continuous time. *Econometrica*, 64, 5, 1195-1212.

Fox, R. and Taqqu, M.S. (1986) Large sample properties of parameter estimates for strongly dependent stationary time series. *Ann. Statist.*, 14, 2, 517-532.

Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long memory time series models, *J. Time Series Anal.*, 4, p. 221-238.

Gobulev, G.K. (1993) Nonparametric estimation of smooth spectral densities of Gaussian stationary sequences, *Theory Probab. Appl.*, 38, 4, 630-639.

Gonçalves, E. (1987) Une généralisation des processus ARMA, *Annales d'Economie et de Statistique*, 5, p.109-145.

Granger, C.W.J. and Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis*, Vol.1, 1, 15-29.

Granger, C. W. J., Robins, R. P., and Engle, R. F. (1986) Wholesale and Retail Prices: Bivariate Time Series Modeling with Forecastable Error Variances, in *Model Reliability*, eds. D. A. Belsey and E. Kuh, MIT Press.

Guyon, X. (1995) *Random fields on a network. Modeling, Statistics and Applications*, Springer Verlag.

Hall, P. and Carroll, R.J. (1989). Variance function estimation in regression: the effect estimation of the mean. *J.R. Statist. Soc.*, **B** 51, 3-14.

Hall, P. and Hart, J.D. (1990) Nonparametric regression with long-range dependence. *Stochastic Process. Appli.*, 36, 339-351.

Hannan, E.J. (1970) *Multiple Time Series*, Wiley, New-York.

Härdle, W and Tsybakov, A. (1997), Local polynomial estimators of the volatility function in nonparametric regression. *Journal of Econometrics*, 81, 223-242.

Hardouin, C. (1992) Quelques résultats nouveaux en statistique des processus : contraste fort, régression à résidus à longue portée, estimation par log-périodogramme, *PhD University Paris 7.*

Harrison, J. and Kreps, D. (1979). Martingale and Arbitrage in Multiperiods Securities Markets. *Journal of Financial Economic Theory*, 20, 380-408.

Hoffman, M. (1999). On nonparametric estimation in nonlinear AR(1)-models. *Statistics and Probability Letters*, 44, 29-45.

Hosking, J.M.R. (1981). Fractional differencing. *Biometrica*, 68, 1, 165-176.

Hu, Y., Oksendal, B. and Sulem, A. (1999) Optimal protfolio in a fractional Black and Scholes market. Preprint , Dep. of Mathematics, University of Oslo, no 13.

Hull, J. and White, A. (1987). The Pricing of Options on Assets with Stochastic Volatilities. *The Journal of Finance*, 3, 281-300.

Jeantheau, T. (1998) Strong Consistency of Estimators for Multivariate ARCH models. *Econometric Theory*, 14, 70–86.

Kashyap, R.L. and Eom, K.B. Estimation in long memory time series model. *J. Time Series Anal.*, 1, 1988, p. 36-41.

Lee, S.W. and Hansen, B.E. (1994), "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator", *Econometric Theory*, 10, 29–52.

Li, K.C. (1987). Asymptotic optimality for C_p, C_l cross-validation and genralized cross-validation: discrete index set. *Ann. Statist.*, 15, 958-975.

Lumsdaine, R. L. (1996) Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models. *Econometrica*, 64, 575–596.

E. Masry (1993) Multivariate regression estimation with errors-in-variables for stationary processes. *J. Nonparametric Statist.* **3**, 13-26.

Melino, A. and Turnbull S.M. (1990) Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45, 239-265.

Nelson, D.B. (1990) Stationarity and Persistence in the GARCH(1,1) Model. *Econometric Theory*, 6, 318–334.

M.B. Priestley, *Spectral analysis and time series.* Vol 1, Academic Press: London, 1981.

Neuman, M.H. (1996) Spectral density estimation via nonlinear wavelet methos for stationary non-Gaussian time series. *J. Time Series Anal.*, 17, 6, 601-633.

Pinkus, A. (1985). *n-widths in Approximation Theory.* Springer-Verlag, Berlin.

Polyak, B.T. and Tsybakov, A. (1992). A family of asymptotically optimal methods for choosing the order of a projective regression estimate. *Theory Probab. Appl.*, 37, 471-481.

Renault, E. et Touzi, N. (1998). Option hedging and implicit volatilities in a stochastic volatility model. *Mathematical Finance*, 6, 279-302.

Rissanen, J. (1984) Universal coding, information, prediction and estimation. *IEEE Trans. Inform. Theory*, 30, 629-636.

Robinson, P.M. (1992) Semiparametric analysis of long memory time series. *Ann. Statist*, 22, 515-539.

Rogers, L.C.G. (1997) Arbitrage with Fractional Brownian Motion. *Mathematical Finance*, 7, 95-105.

Rozanov, Yu. A. (1967). *Stationary Random Processes*, Holden Day.

Scott, L. (1987) Option pricing when the variance changes randomly: estimation and an application. *Journal of Financial and Quantitative Analysis*, 22, 419-438.

Shibata , R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, 63, 117-126.

Taupin, M.L. (1998) Semi-parametric estimation in nonlinear structural errors-in-variables model. *Preprint Université Paris-Sud* 98-35.

Velasco, C. (2000) Non-Gaussian log-periodogram regression. *Econometric Theory*, 16, 1, 44-79.

Whittle, P. (1953) The analysis of multiple stationary time series. *J. R. Statist. Soc.*, **B**, 15, 125-139.

Yao, Q. and Tong, H. (1994). Quantifying the Influence of Initial Values on Non-linear Prediction. *J. R. Statist. Soc.* **B**, 4, 701-725.

Table des matières