# Mémoire de synthèse suite à l'obtention de l'habilitation à diriger des recherches.

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

**Fabienne Comte***

Mars 2001

*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 évanescent s'enorgueillit irrésistiblement de la présence ici de Dominique Picard, qui est pour moi un exemple. 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 tard, 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 digressions : merci.

Je remercie également la vieille URA 1321 de m'avoir accueillie jadis dans l'UFR 920, et en particulier Denis Bosq, avec qui j'ai pris plaisir à enseigner la Statistique des Processus lors de mon arrivée à Paris 6. 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 8 travaux publiés ou à paraître et 6 travaux soumis pour publication ou en révision. Des allusions sont faites à deux travaux en cours. En voici la liste :

**Articles publiés ou à paraître :**

[1]. "Simulation and estimation of long memory continuous time models", *Journal of Time Series Analysis* **17**, 1, 19-36, 1996.

[2]. "Non causality in continuous time models", en collaboration avec E. Renault, *Econometric Theory* **12**, 2, 215-256, 1996.

[3]. "Long memory continuous time models", en collaboration avec E. Renault, *Journal of Econometrics* **73**, 1, 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* **8**, 4, 291-323, 1998.

[6]. "Discrete and continuous time cointegration", *Journal of Econometrics*, **88**, 2, 207-226, 1999.

[7]. "Variance noncausality in multivariate GARCH processes", en collaboration avec O. Lieberman, *Journal of Time Series Analysis*, **21**, 5, 535-557, 2000.

[8]. "Adaptive estimation of the spectrum of a stationary Gaussian sequence", Prépublication du Laboratoire de Probabilités de Paris 6 n$^o$ 525, à paraître dans *Bernoulli*.

**Travaux soumis ou en révision:**

[9]. "Adaptive estimation in an autoregression and a geometrical $\beta$-mixing regression framework", en collaboration avec Y. Baraud et G. Viennet, Prépublication Crest-Insee 9807 en révision pour *Annals of Statistics*.

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

[11]. "Kernel estimation of a stochastic volatility model", Prépublication du Laboratoire de Probabilités de Paris 6 n$^o$ 560, en révision pour *Econometric Theory*.

[12]. "Asymptotic theory for multivariate GARCH processes", en collaboration avec O. Lieberman, Prépublication du Laboratoire de Probabilités de Paris 6 n$^o$ 593, en révision pour *Journal of Multivariate Analysis*.

[13]. "Semi-parametric estimation in the (auto)-regressive $\beta$-mixing model with errors-in-variables", en collaboration avec M.-L. Taupin, Prépublication du Laboratoire de Probabilités de Paris 6 n$^o$ 606, soumis pour publication à *Mathematical Methods of Statistics*.

[14]. "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*.

**Travaux en cours :**

[15]. "Adaptive estimation of the stationary density of mixing processes", en collaboration avec F. Merlevède.

[16]. "Affin fractional stochastic volatility models", en collaboration avec L. Coutin et E. Renault.

Les travaux [8], [11], [12], [13] sont téléchargeables sur le site www.proba.jussieu.fr.

## QUELQUES MOTS D'INTRODUCTION.

On peut regrouper mes travaux de plusieurs manières. Dans tous les cas, on ne sort jamais des Mathématiques Appliquées et Applications des Mathématiques. Néanmoins, on peut les répartir entre deux domaines en interaction et s'alimentant mutuellement : l'Économétrie Théorique, qui a connu ses mutations les plus décisives au cours des dernières années sous l'impulsion très dynamique de la Finance, et la Statistique Mathématique, qui ne peut manquer de subir un peu aussi cette influence, quoique sans doute de façon moins exclusive.

Dans mon travail économétrique, je me suis attachée à modéliser des dépendances nouvelles, dans mon travail statistique à en estimer d'anciennes par des méthodes nouvelles et plus performantes. Je vais tenter ici de développer quelques idées directrices concernant quatre thèmes dominants : le mémoire longue, la causalité, les modèles à volatilité stochastique, la sélection de modèles.

Ces quatre thèmes – les deux premiers plutôt économétriques et les deux derniers plutôt statistiques – illustrent bien les spécificités des deux disciplines : d'un côté on trouve davantage de modélisation et de réflexion sur ce dont on veut rendre compte, on étudie les propriétés du modèle en fonction de ce que l'on souhaite illustrer et on essaie de se replacer dans un contexte statistique plus ou moins standard pour appliquer des méthodes d'estimation existantes. De l'autre, le modèle est davantage considéré comme acquis et c'est précisément sur la méthode d'estimation que l'on travaille, afin de l'améliorer et de la rendre toujours plus automatique et dictée par les données – "data-driven" – pour éviter l'arbitraire des choix (de fenêtre, de dimension) qui interviennent au cours de certaines procédures. Il est à noter que l'accroissement de la puissance des ordinateurs amène désormais le statisticien autant que l'économètre à tester presque systématiquement les performances des modèles et des méthodes par le biais de la simulation ou du test sur données réelles.

## THÈME 1 : LONGUE MÉMOIRE.

Pour commencer, l'idée de mémoire longue concernant des processus qui évoluent dans le temps parle d'elle-même. Elle concerne des propriétés de dépendance forte entre des variables mesurées à des instants différents, éventuellement très éloignés. De tels phénomènes ont été mis en évidence très tôt en hydrologie, cf. Hurst (1953). Ils se caractérisent de façon mathématique par une décroissance très lente de la corrélation des processus stationnaires, i.e. tels que

$$\rho(h) = \mathrm{corr}(X_t, X_{t+h}) = \mathrm{cov}(X_t, X_{t+h})/\mathrm{Var}(X_t)$$

existe et ne dépend que de $h$ (et pas de $t$). En effet, les processus stationnaires standards sont tels que, pour $h$ grand, $\rho(h)$ est d'ordre $e^{-ah}$, pour un $a > 0$ (décroissance exponentielle) alors que les processus à longue mémoire sont tels que $\rho(h)$ est d'ordre

$h^{2\alpha-1}$ (décroissance hyperbolique) pour un $\alpha$ compris entre 0 et 1/2.
Ces processus ont plus tard été illustrés par le mouvement Brownien Fractionnaire $B_t^\alpha$, dont les accroissements stationnaires

$$B_{t+h}^\alpha - B_t^\alpha = \int_{-\infty}^{t+h} \frac{(t+h-s)^\alpha}{\Gamma(1+\alpha)} dB_s - \int_{-\infty}^{t} \frac{(t-s)^\alpha}{\Gamma(1+\alpha)} dB_s,$$

où $B$ est un mouvement brownien et $\alpha \in (-1/2, 1/2)$, possèdent cette propriété. Le mouvement Brownien Fractionnaire cumule cette faculté de mémoire avec une $H$-autosimilarité stochastique, $H = \alpha + 1/2$, cette propriété de se reproduire à l'identique à différentes échelles, mais en loi. Son étude a été initialisée par les travaux de Mandelbrot et Van Ness (1968) et a connu un important essor au cours des dernières décennies. Ainsi, le terme de “fractionnaire” pour qualifier des processus signifie parfois qu'ils possèdent une propriété d'auto-similarité et parfois qu'ils sont à mémoire longue. Les modèles à temps discret dits “ARMA fractionnaires” qui se sont développés à partir des années 80 en économétrie (Granger et Joyeux (1980), Hosking (1981), Geweke et Porter-Hudak (1983)) n'ont par exemple que la propriété de longue mémoire. Outre leur signification particulière (oubli très lent du passé), de tels processus ont déclenché des pluies de conjectures et d'articles du point de vue statistique : en effet, ils ne vérifient pas les hypothèses standard présidant aux méthodes d'estimation classiques qui existaient alors pour des processus stationnaires (en particulier, bornitude inférieure stricte et supérieure de la densité spectrale pour l'étude du contraste de Whittle (1951)). Ils ont donc donné lieu au développement de nouvelles techniques et méthodologies (Fox et Taqqu (1986), Dahlhaus (1989), Robinson (1992)).
Parallèlement, sous l'impulsion de la Finance, la modélisation économétrique s'est mise à représenter l'évolution des processus en temps continu. Typiquement, au lieu de considérer que

$$X_{t+1} = \rho X_t + \sigma \varepsilon_{t+1}, \quad |\rho| < 1 \tag{1}$$

où $\varepsilon_t$ est une suite de termes d'erreurs stochastiques indépendants et de même loi, pour des instants discrets $t = 1, 2, 3\ldots$ où l'unité de temps est arbitraire, on tente d'écrire, de façon structurelle :

$$dX_t = -\theta X_t dt + \sigma dB_t, \quad \theta > 0, \tag{2}$$

où $B_t$ est un mouvement Brownien ($dB_t$ joue donc le rôle d'erreur aléatoire) et $t$ varie continuement.
Ainsi apparaît,de façon très logique, l'idée de construire des modèles à temps continu et à mémoire longue. D'une certaine façon, cela revient à remplacer dans (2) l'accroissement infinitésimal du mouvement Brownien par un accroissement infinitésimal du mouvement Brownien Fractionnaire. C'est ainsi que dans les travaux [1], [3] et [5], en collaboration avec Eric Renault, j'ai construit de tels modèles et j'ai fourni des réponses aux questions suivantes :

1. Les processus ainsi obtenus sont-ils mathématiquement bien définis? Le cadre Gaussien permet dans un premier temps de ne considérer que des intégrales de Wiener et de justifier que c'est en toute légitimité que l'on manipule ces objets. Des cadres beaucoup plus vastes ont été explorés depuis lors par des probabilistes ou des financiers (Alos, Mazet et Nualart (1999), Coutin et Decreusefond (1999), Carmona et Coutin (1999), Hu, Oksendal et Sulem (1999)).

2. Les modèles obtenus ont-ils des propriétés pouvant rendre compte de phénomènes observés : mémoire, régularité, comportement dans le domaine spectral?

3. Une fois un jeu d'observation recueilli − nécessairement à temps discret − pourra-t-on mener une inférence statistique donnant des estimations raisonnablement convergentes des paramètres?

Le sujet s'est révélé extraordinairement fertile et ne cesse de se développer. En fournissant des réponses aux trois questions ci-dessus, [1], [3] et [5] ouvraient une voie vers des modèles plus généraux ou plus souples permettant d'envisager de nombreuses applications en Finance. Par ailleurs, on illustre dans [3] que l'agrégation de nombreux comportements individuels "sans mémoire" peut engendrer un comportement global (somme des équations individuelles) ou moyen (moyenne des équations individuelles) avec mémoire longue : remarque riche de conséquence pour la modélisation de réseaux ou de la consommation électrique par exemple.
Un autre exemple porteur a aussi été l'extension du modèle de Hull et White (1987) à volatilité stochastique au cadre fractionnaire. Un modèle à volatilité stochastique est un modèle semblable à (2) :

$$dX_t = -\theta X_t dt + \sigma X_t dB_t,$$

où la variance $\sigma^2$, au lieu d'être constante, est un processus régi par une nouvelle équation, de même type que (2). Ces généralisations sont apparues en Finance pour rendre compte de la variabilité de la volatilité ($\sigma^2$) des prix des actifs financiers, observée sur les marchés, malgré des formules de prix fondées sur une hypothèse de type $\sigma$ = constante. On obtient ainsi le modèle :

$$\begin{aligned} dX_t &= -\theta X_t dt + \sigma_t X_t dB_t^{(1)}, && (3)\\ d\log(\sigma_t) &= -\psi \log(\sigma_t) dt + \gamma dB_t^{(2)}, && (4) \end{aligned}$$

où $B_t^{(1)}$ et $B_t^{(2)}$ sont deux mouvements Browniens souvent supposés indépendants et où la seconde équation est écrite sur $\log(\sigma)$ pour garantir la positivité de $\sigma$. L'hypothèse d'absence d'opportunité d'arbitrage (impossibilité de faire de l'argent à partir de rien) a interdit très vite de modifier trop profondément l'équation (3) : en particulier, Rogers (1997) a montré que l'on ne pouvait continuer à voir cette hypothèse satisfaite si l'on remplaçait $B_t^{(1)}$ par un mouvement Brownien Fractionnaire.

Mais il se trouve que ce sont souvent les propriétés de la volatilité qui sont au centre des préoccupations des financiers. Dans ce cadre, lorsqu'on remplace $B_t^{(2)}$ par un mouvement Brownien Fractionnaire, on généralise le processus de volatilité et on lui octroie ainsi de nouvelles propriétés intéressantes, pouvant justifier de nombreuses observations concrètes (forte dépendance dans les volatilités entre autres) tout en conservant une dynamique raisonnable pour $X_t$. Cette étude fut spécifiquement l'objet de [5]. Une collaboration nouvelle s'est même engagée sur ce thème dans [16].

Du point de vue mathématique, les techniques mises en oeuvre dans ces travaux relèvent du calcul stochastique et de la théorie de l'intégrale de Wiener pour ce qui est de la définition des processus en jeu. Nous avons eu aussi besoin des méthodes, courantes en Finance, de changement de probabilité pour le calcul des prix d'option. Nous avons étudié des schémas d'Euler de discrétisation afin de pouvoir simuler les processus, procéder à des dérivations fractionnaires approchées et tester les méthodes d'estimation. Du point de vue statistique, nous nous basons toujours sur un échantillon à temps discret et, en reliant la densité spectrale du processus discrétisé à la densité spectrale du processus à temps continu, nous vérifions que l'on est dans le cadre d'applications de méthodes semi-paramétriques proposées par Robinson (1992) pour estimer le paramètre de mémoire ou des méthodes paramétriques de contraste de Whittle généralisées par Fox et Taqqu (1986) et Dahlhaus (1989).

## THÈME 2 : CAUSALITÉ

Un second thème, développé dans les travaux [2] et [7], est celui de la causalité. La question est simple : a-t-on besoin d'observer $Y$ pour prévoir $X$, en moyenne ? C'est encore une manière d'étudier la dépendance, mais non plus d'une variable vis à vis d'elle-même au cours du temps comme précédemment, mais de l'avenir d'une variable par rapport au passé d'une autre. La notion de "non causalité au sens de Granger" se formalise alors ainsi. Si $\mathcal{I}(t)$ désigne l'information disponible à l'instant $t$ lorsqu'on ne tient pas compte du passé de $Y$ et $\mathcal{I}_Y(t)$ l'information disponible à l'instant $t$ compte-tenu du passé de $Y$, alors on dit que $Y$ ne cause pas $X$ au sens de Granger si

$$\mathbb{E}(X_{t+1}|\mathcal{I}(t)) = \mathbb{E}(X_{t+1}|\mathcal{I}_Y(t)), \tag{5}$$

où $\mathbb{E}(.|.)$ désigne l'espérance conditionnelle. Sous l'impulsion de la Finance encore, de nombreux modèles se sont développés pour rendre compte de la dynamique des processus non seulement en moyenne, mais aussi en variance : on est ainsi passé de modèles homoscédastiques à des modèles hétéroscédastiques. Schématiquement, cela revient à remplacer (1) par

$$X_{t+1} = \rho X_t + \sigma_t \varepsilon_{t+1} \tag{6}$$

et à construire un nouveau modèle pour $\sigma_t^2$, comme, par exemple

$$\sigma_t^2 = \omega + a\sigma_{t-1}^2 + b\varepsilon_{t-1}^2. \tag{7}$$

Des modèles tels que (6)-(7) sont appelés modèles GARCH (Generalized AutoRegressive Conditionally Heteroskedastic) et ont été introduits par Engle (1982). Ils ont été généralisés par Bollerslev (1986) et très intensément étudiés par l'ensemble de la littérature économétrique et financière depuis leur création. Le formalisme multivarié final n'a été mis en place par Engle et Kroner qu'en 1995.
Il s'agit de rendre compte d'une dynamique spécifique de la variance $\sigma_t^2$ conditionnelle au passé du processus $X$. Dans un tel modèle, étendu éventuellement au cadre multivarié, il apparaît que les causalités peuvent jouer en moyenne mais aussi en variance. Ainsi [7], en collaboration avec Offer Lieberman, étudie une notion de non causalité au second ordre, sa formalisation et ses caractérisations, notamment dans des classes de modèles de type modèles GARCH. Plus tard, dans [12], nous nous attachons à résoudre un problème spécifiquement statistique lié à ces modèles : il s'agit de prouver une conjecture − classiquement utilisée dans la littérature et en particulier dans [7] − de normalité asymptotique d'un estimateur des paramètres d'un modèle GARCH multivarié dans sa représentation d'Engle et Kroner (1995).
La seconde question que l'on peut se poser vis à vis de la causalité concerne la façon la plus structurelle de la définir. La relation (5) dépend de l'unité de temps choisie, car des chaînes de causalité peuvent se créer artificiellement par le biais de l'environnement. Cette dépendance peut se mettre en évidence en considérant comme dans le cadre envisagé pour la mémoire longue, des processus à temps continu, pour lesquels on peut alors faire aisément varier le pas de discrétisation. La question que nous étudions alors dans [2], en collaboration avec Eric Renault, est la suivante : quel sens donner à une non causalité qui serait instantanée ? On voit que dans (5), si on remplace $X_{t+1}$ par $X_{t+\delta}$ avec $\delta$ petit, on va tendre vers une propriété trivialement vérifiée dans tout contexte de processus et de filtrations continus. Pourtant, puisque dans le modèle :

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix}$$

où $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t})'$ est i.i.d., $\mathbb{E}(\varepsilon_t) = 0$, $\mathrm{Var}(\varepsilon_t) = I_2$, avec $I_2$ matrice identité d'ordre 2, la non causalité discrète de $Y$ sur $X$ est caractérisée par $a_{12} = 0$ (Boudjellaba, Dufour et Roy (1992)), dans son analogue à temps continu

$$d\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} dt + \begin{pmatrix} dB_t^{(1)} \\ dB_t^{(2)} \end{pmatrix},$$

on imagine aisément que la non causalité structurelle doit entraîner $b_{12} = 0$. C'est pourquoi, suivant une voie ébauchée par Florens et Fougère (1996), [2] s'attache à considérer les décompositions semi-martingales des processus par rapport à différentes filtrations et à caractériser la notion ainsi mise en place via des espérances conditionnelles de dérivées de processus, le cas échéant. Les différentes propriétés impliquées par cette définition nouvelle de non causalité dite “immédiate”[1] sont alors diversement illustrées, dans des modèles à volatilité stochastiques tels que (3)-(4), ou des modèles linéaires à temps continu. Dans ce dernier cas, on dispose de discrétisations exactes des processus qui permettent d'étudier le lien entre causalité immédiate et causalité à temps discret. Si des causalités discrètes apparaissent d'une variable $Y$ vers une variable $X$ alors que le critère à temps continu indique que $Y$ ne cause pas $X$ de façon immédiate, c'est qu'une troisième variable $Z$, environnementale, véhicule des effets de $Y$ vers $X$ : mais la chaîne de causalité n'est en fait pas de $Y$ vers $X$ mais de $Y$ vers $Z$ et de $Z$ vers $X$, de sorte que, pour agir sur l'avenir de $X$, c'est sur le présent de $Z$ et non celui de $Y$ qu'il faut faire porter ses efforts. Une fois la notion mise en place et caractérisée dans certains cadres paramétriques, nous avons montré qu'il était possible de développer alors des méthodologies standard pour l'estimation et des tests de causalité.

## THÈME 3 : VOLATILITÉ STOCHASTIQUE

Nous avons rencontré à plusieurs reprise lors de l'analyse des thèmes précédents, un exemple qui est, aujourd'hui encore, l'objet de nombreuses études : il s'agit des modèles à volatilité stochastique décrits par (3)-(4). Ces modèles, introduits par Hull et White (1987), ont été largement exploités par Renault et Touzi (1998) qui en ont montré la pertinence pour rendre compte de nombreux effets observés sur les marchés − notamment l'effet “smile” sur la courbe de volatilité (la volatilité implicite est une fonction convexe du prix d'exercice de l'option construite sur l'actif dont la dynamique est modélisée par (3)-(4)). Ces modèles ont aussi été étudiés dans [2] du point de vue des hypothèses de non causalité implicites qu'ils contiennent dans leur formulation la plus standard et dans [5], du point de vue des généralisations fractionnaires ou longue mémoire. La difficulté statistique majeure pour mener à bien l'estimation des paramètres ou des fonctions impliqués dans ces modèles provient du fait que $\sigma_t$ n'est pas observé. Seul $X_t$ est mesuré.
Considérons par exemple le modèle à volatilité stochastique suivant :

$$\begin{aligned} d\log(X_t) &= -\mu dt + \sigma_t dB_t^{(1)}, \\ d\log(\sigma_t) &= -\psi \log(\sigma_t)dt + \gamma dB_t^{(2)}, \end{aligned}$$

dans le cas où, $\mu$ pouvant être estimé, on se ramène à supposer $\mu = 0$. Une

[1] Le terme instantanée, plus satisfaisant, n'a pu être retenu car il désigne déjà une autre forme de causalité à temps discret.

discrétisation au pas $h$ conduit aux équations approchées

$$\begin{array}{rcl}\log(X_{t+h}/X_t) & \sim & \sigma_{t+h}(B^{(1)}_{t+h} - B^{(1)}_t),\\ \log(\sigma_{t+h}) & = & e^{-\psi h}\log(\sigma_t) + \gamma\int_t^{t+h} e^{-\psi(t+h-s)}dB^{(2)}_s.\end{array}$$

En posant $Y_n = \log(X_{nh}/X_{(n-1)h})$, $\log(\sigma_{nh}) = Z_n$,

$$v_n = B^{(1)}_{nh} - B^{(1)}_{(n-1)h},\ \ \eta_n = \gamma\int_{(n-1)h}^{nh} e^{-\psi(nh-s)}dB^{(2)}_s,$$

on obtient le modèle discrétisé approchant

$$\left\{\begin{array}{l} Y_n = \exp(Z_n)v_n,\\ Z_{n+1} = e^{-\psi h}Z_n + \eta_{n+1}.\end{array}\right.$$

En posant de plus $U_n = \log(|Y_n|) - \mathbb{E}|v_n|$ et $\varepsilon_n = |v_n| - \mathbb{E}|v_n|$, on trouve :

$$\left\{\begin{array}{l} U_n = Z_n + \varepsilon_n,\\ Z_n = e^{-\psi h}Z_{n-1} + \eta_n,\end{array}\right. \tag{8}$$

où les $\varepsilon_n$ sont i.i.d. centrées de variance notée $\sigma^2_\varepsilon$ et les $\eta_n$ i.i.d. centrées de variance notée $\sigma^2_\eta$. Il est clair avec cette construction que les variables $\varepsilon$ et $\eta$ ne sont pas gaussiennes.
Le modèle (8) est connu sous le nom de modèle d'autorégression avec erreurs-sur-les-variables. Il sert, dans la littérature, à modéliser une autorégression où le processus est observé avec une erreur de mesure. D'où l'idée de considérer des modèles de ce type dans un cadre non paramétrique:

$$\left\{\begin{array}{l} U_n = Z_n + \varepsilon_n,\\ Z_n = m(Z_{n-1}) + \eta_n,\end{array}\right.$$

où l'on cherche à estimer la fonction $m$ et où les bruits $\varepsilon$ et $\eta$ sont i.i.d. centrés (et non nécessairement gaussiens). On peut appliquer dans ce cas une méthode de noyau de déconvolution existante, dans un cadre de variables dépendantes. Indiquons brièvement comment est construit l'estimateur issu de cette méthode. Comme

$$\mathbb{E}[U_{j+1}|Z_j = z] = m(z) = \frac{\int u f_{(Z,U)}(z,u)du}{f_Z(z)}$$

où $f_{(Z,U)}$ désigne la loi jointe de $(Z_1, U_1)$ et $f_Z(z)$ la loi de $Z_1$, si les $Z_i$ étaient observées en même temps que les $U_i$, on pourrait utiliser l'estimateur

$$\hat{m}(z) = \frac{\dfrac{1}{nh_n}\displaystyle\sum_{j=1}^n U_{j+1}K\left(\frac{z - Z_j}{h_n}\right)}{\dfrac{1}{nh_n}\displaystyle\sum_{j=1}^n K\left(\frac{z - Z_j}{h_n}\right)} \tag{9}$$

où $K$ est un noyau et $h_n$ une fenêtre. Le dénominateur en particulier est un estimateur standard de la densité $f_Z$. Bien entendu, puisqu'on n'observe pas les $Z_i$, on doit le modifier. On procède de la façon suivante. Comme $f_U(u) = f_Z * f_\varepsilon(u)$ où $*$ désigne le produit de convolution : $g * h(x) = \int g(x-u)h(u)du$, on en déduit $f_U^\star(u) = f_Z^\star(u) f_\varepsilon^\star(u)$ où $g^\star$ désigne la transformée de Fourier: $g^\star(x) = \int e^{itx} g(t)dt$. Par conséquent, en utilisant la formule d'inversion de Fourier, on trouve

$$f_Z(z) = \frac{1}{2\pi}\left(\frac{f_U^\star}{f_\varepsilon^\star}\right)^\star(-z) = \frac{1}{2\pi}\int e^{-itz}\frac{f_U^\star(t)}{f_\varepsilon^\star(t)}dt.$$

D'où l'idée due à Stefanski et Carroll (1990) de considérer l'estimateur suivant de $f_Z(z)$ :

$$\hat{f}_n(z) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itz} K^\star(th_n)\frac{\hat{\Phi}_n(t)}{f_\varepsilon^\star(t)}dt,$$

où $K$ est un noyau et $\hat{\Phi}_n$ est la fonction caractéristique empirique de $U$ :

$$\hat{\Phi}_n(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itU_j}.$$

Ainsi, si on définit

$$K_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-itx)\frac{K^\star(t)}{f_\varepsilon^\star(t/h_n)}dt, \tag{10}$$

on a aussi

$$\hat{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K_n\left(\frac{z-U_i}{h_n}\right). \tag{11}$$

En substituant $K_n$ à $K$ dans la formule générale (9), on obtient :

$$\hat{m}_n(z) = \frac{\sum_{j=1}^n K_n((z-U_j)/h_n)U_{j+1}}{\sum_{i=1}^n K_n((z-U_i)/h_n)} = \frac{1}{nh_n}\sum_{j=1}^{n}\frac{K_n((z-U_j)/h_n)U_{j+1}}{\hat{f}_n(z)} \tag{12}$$

où $K_n$ est donné par (10) et $\hat{f}_n(x)$ par (11).
C'est le but de [11] que de faire le lien entre la littérature financière et la littérature relative au traitement statistique non paramétrique des modèles avec erreur-sur-les-variables, effectué par Fan et Truong (1993) pour des variables indépendantes et par Masry (1993) pour des variables dépendantes.
La dépendance ici se mesure par le biais de coefficients dits "de mélange". Nous allons seulement mentionner pour fixer les idées la définition des processus mélangeants correspondant aux deux types de mélange que nous évoquerons. Soit $\mathcal{F}_i^k$ la $\sigma$-algèbre

des événements engendrés par des variables aléatoires $\{X_j, i \leq j \leq k\}$. Un processus strictement stationnaire $\{X_j\}$ est dit fortement mélangeant ou $\alpha$-mélangeant (Rosenblatt (1956)) si

$$\sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| = \alpha_k \to 0 \text{ quand } k \to +\infty.$$

Il est dit absolument régulier ou $\beta$-mélangeant (Kolmogorov and Rozanov 1960) si

$$\frac{1}{2} \sup \left\{ \sum_{i=1}^{I} \sum_{j=1}^{J} |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)| \right\} = \beta_k \to 0 \text{ quand } k \to +\infty.$$

où le sup ci-dessus est pris sur toutes les partitions finies $(A_i)_{1 \leq i \leq I}$ et $(B_j)_{1 \leq j \leq J}$ de l'espace de probabilité ambiant $\Omega$ respectivement $\mathcal{F}_{-\infty}^0$ et $\mathcal{F}_k^\infty$ mesurable. Notons que le mélange $\beta$ est plus fort et entraîne le mélange $\alpha$.
Du point de vue technique, je considère dans [11] l'estimateur à noyau (12) étudié par Fan et Truong (1993) dans un cadre de variables indépendantes (le modèle d'autorégression est alors remplacé par un modèle de régression) et je généralise leurs résultats de contrôle de la distance quadratique intégrée entre $f$ et son estimateur ($\mathbb{E} \int (f(x) - \hat{f}(x))^2 w(x) dx$ où $\hat{f}$ est l'estimateur de $f$ et $w$ une fonction de poids) au cas de variables $\alpha$-mélangeantes en utilisant des inégalités de covariance désormais classiques dues à Davydov (1970). Le résultat est sans surprise, sensiblement plus simple que celui de Masry (1993) qui contrôlait la distance entre $f$ et son estimateur en norme infinie. Par ailleurs, j'ai expérimenté sur des simulations que sa mise en œuvre pratique était extrêmement aisée.
Ces méthodes impliquent néanmoins pour certains types de lois d'erreurs $\varepsilon_n$ des vitesses de convergence de l'estimateur non paramétrique qui, tout en étant les meilleures possibles, sont inéluctablement très lentes : par exemple, si les $\varepsilon_n$ sont gaussiennes, la vitesse de convergence est en $1/\log(n)^2$. C'est pourquoi il paraît intéressant d'exploiter la structure naturellement paramétrique des modèles obtenus par discrétisation pour mettre en œuvre d'autres méthodes, plus spécifiquement paramétriques et susceptibles d'atteindre la vitesse $1/\sqrt{n}$. Ainsi [13], en collaboration avec Marie-Luce Taupin, généralise un estimateur proposé par Taupin (2000) pour des modèles de régression avec erreur-sur-les-variables, dans un cadre de variables indépendantes et de bruit $\varepsilon$ gaussien. L'extension porte tout à la fois sur l'abandon de l'hypothèse d'indépendance et sur la généralisation de la loi des $\varepsilon_n$. Cette démarche devrait permettre le traitement statistique de modèles à volatilité stochastique plus complexes et éventuellement aussi leurs extensions fractionnaires. Mathématiquement, nous travaillons dans un cadre paramétrique avec un estimateur nouveau pour des modèles paramétriques généraux : il est fondé sur un contraste des moindres carrés dans lequel les variables non observées sont remplacées par une approximation de leur espérance conditionnelle sachant les observées; cette

approximation est aussi basée sur une formule de Fourier inverse. On peut citer seulement deux concurrents éloignés de cet estimateur. L'estimateur proposé par Chanda (1995) atteint la vitesse $n^{-1/2}$ paramétrique mais est spécifique à une classe de modèles linéaires avec erreurs-sur-les-variables. Douc et Matias (2000) ont montré la $n^{-1/2}$-consistance de l'estimateur du maximum de vraisemblance pour des chaînes de Markov cachées : leurs résultats s'appliqueraient à nos modèles à condition de paramétrer la loi du second bruit $\eta$. Ceci n'est pas nécessaire dans notre cadre de travail et c'est une contrainte à laquelle il est satisfaisant d'échapper.
La technique utilisée est une méthode de développement de Taylor assez standard en statistique paramétrique, mais qui fait apparaître un grand nombre de termes à contrôler, dont certains en norme infinie (norme sup). Nous sommes donc amenées à évaluer, dans un cadre $\beta$-mélangeant, les déviations du supremum d'un processus par rapport à son espérance en étendant une inégalité de Massart et Rio (1998) à des variables $\beta$-mélangeantes par la technique d'approximation de Viennet (1997). Nous appliquons par ailleurs des inégalités de covariance à la Davydov (1970) en cascade pour majorer des sommes de produits d'ordres divers de variables mélangeantes. Nous obtenons un résultat de consistance de l'estimateur sous des conditions de régularité standard de la fonction de régression par rapport aux paramètres, des conditions sur la vitesse de mélange et sur la vitesse de décroissance de la fonction caractéristique du bruit $\varepsilon$ à l'infini. Nous donnons de plus, sous des conditions légèrement renforcées, une borne supérieure pour la vitesse de convergence. Nous vérifions que l'ordre obtenu peut correspondre à la vitesse paramétrique pour certaines classes de fonctions de régression.

## THÈME 4 : SÉLECTION DE MODÈLES

Le dernier thème dont je vais parler est plus exclusivement statistique. Supposons que l'on veuille estimer une certaine fonction $f$ : densité de variables strictement stationnaires de loi $f$ dans [15], densité spectrale d'une suite stationnaire gaussienne dans [8] c'est-à-dire

$$f(\lambda) = \frac{1}{2\pi} \sum_{k \in Z} \gamma_k e^{i\lambda k} \quad \text{avec} \quad \gamma_k = \text{cov}(X_1, X_k),$$

fonction de régression ou d'autorégression dans le modèle autorégressif

$$X_{t+1} = f(X_t) + \sigma\varepsilon_{t+1}, \quad \varepsilon_t \text{ i.i.d.}, \ \mathbb{E}(\varepsilon_t) = 0, \ \text{Var}(\varepsilon_t) = 1,$$

dans [9] et [10], ou fonction de volatilité dans le modèle

$$X_{t+1} = m(X_t) + f(X_t)\varepsilon_{t+1}, \quad \varepsilon_t \text{ i.i.d.}, \ \mathbb{E}(\varepsilon_t) = 0, \ \text{Var}(\varepsilon_t) = 1,$$

dans [14].
Dans tous les cas, nous allons tout d'abord définir un estimateur privilégié comme une fonction choisie dans un espace linéaire de dimension finie de fonctions et minimisant un certain contraste. On considère par exemple des espaces de polynômes par morceaux ou d'ondelettes. Prenons, pour illustrer la méthode, le cas où nous cherchons cette fonction sous la forme d'un histogramme. En d'autres termes, si $S_D$ est l'espace engendré par $\varphi_{1,D}, \varphi_{2,D}, \ldots, \varphi_{D,D}$,

$$\varphi_{j,D} = \sqrt{D}\mathbb{1}_{[(j-1)/D, j/D[}, \quad j = 1, 2, \ldots, D,$$

base orthonormée de $\mathbb{L}_2([0,1])$, nous cherchons le $t = \sum_{j=1}^{D} a_{j,D}\varphi_{j,D}$ de $S_D$ qui minimise un contraste pertinent. Considérons pour simplifier le cas (développé par Birgé et Massart (1997)) de variables $X_1, \ldots, X_n$ i.i.d. de densité $f$ par rapport à la mesure de Lebesgue, à support dans $[0,1]$ et pour lesquelles on veut estimer $f$. Dans ce cas, le contraste (de projection) est :

$$\gamma_n(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\|t\|^2 - 2t(X_i)\right] \quad \text{où} \quad \|t\|^2 = \int_0^1 t^2(x)dx.$$

On obtient sans difficulté l'expression de l'estimateur $\hat{f}_D$ de $f$ qui minimise $\gamma_n(t)$ sur $S_D$ :

$$\hat{f}_D = \sum_{j=1}^{D} \hat{a}_{j,D}\varphi_{j,D},$$

avec

$$\hat{a}_{j,D} = \frac{1}{n}\sum_{i=1}^{n}\varphi_{j,D}(X_i) = \frac{\sqrt{D}}{n}\text{card}\left\{i \;/\; X_i \in \left[\frac{j-1}{D}, \frac{j}{D}\right[\right\}.$$

De même, on montre aisément que, si $f_D = \sum_{j=1}^{D} a_{j,D}(f)\varphi_j$ est la projection $\mathbb{L}_2$-orthogonale de $f$ sur $S_D$ :

$$\|f - \hat{f}_D\|^2 = \|f - f_D\|^2 + \|f_D - \hat{f}_D\|^2$$

et que

$$\|f_D - \hat{f}_D\|^2 = \sum_{j=1}^{D}\left[a_{j,D}(f) - \hat{a}_{j,D}\right]^2.$$

Comme $a_{j,D}(f) = <f, \varphi_{j,D}> = \mathbb{E}(\hat{a}_{j,D})$, on obtient :

$$\begin{aligned}
\mathbb{E}\|f_D - \hat{f}_D\|^2 &= \sum_{j=1}^{D}\text{Var}(\hat{a}_{j,D}) = \frac{1}{n^2}\sum_{j=1}^{D}\sum_{i=1}^{n}\text{Var}(\varphi_j(X_i)) = \frac{1}{n}\sum_{j=1}^{D}\text{Var}(\varphi_j(X_1)) \\
&\leq \frac{1}{n}\sum_{j=1}^{D}\mathbb{E}(\varphi_j^2(X_1)) = \frac{D}{n}\sum_{j=1}^{D}\int_{(j-1)/D}^{j/D} d(x)dx = \frac{D}{n}\int_0^1 f(x)dx \\
&\leq \frac{D}{n}.
\end{aligned}$$

On fait ainsi apparaître une décomposition du risque quadratique $\mathbb{E}\|f - \hat{f}_D\|^2$ particulièrement simple et extrêmement générale de type biais$^2$+ variance :

$$\mathbb{E}\|f - \hat{f}_D\|^2 \leq \|f - f_D\|^2 + \frac{D}{n}.$$

Par ailleurs, des résultats classiques de théorie de l'approximation (cf. DeVore et Lorentz (1993)) nous disent que si $f$ est de régularité $\alpha$, alors $\|f - f_D\|^2$ est d'ordre $D^{-2\alpha}$. Par conséquent, la meilleure vitesse est obtenue pour $S_D$ choisi de dimension $n^{1/(1+2\alpha)}$ et vaut $n^{-2\alpha/(2\alpha+1)}$. On voit que pour minimiser le risque, il faut $D$ en fonction de $\alpha$, c'est-à-dire que l'on souhaite *sélectionner* un espace $S_D$ de dimension optimale, d'où le terme de "sélection de modèles". Malheureusement, ce choix est en général inaccessible, car $f$ étant inconnue, sa régularité $\alpha$ l'est également, et elle est impossible à estimer.
Le but des techniques de sélection de modèles est de proposer une méthode − en général un contraste judicieusement pénalisé − permettant d'atteindre automatiquement la meilleure vitesse possible sans connaître $\alpha$. Les premiers critères pénalisés visant à faire de la sélection de modèles sont apparus dans les années 70 (par exemple le critère AIC d'Akaike (1973) ou le $C_p$ de Mallows (1973)). Les premières études formalisées de ces critères, notamment en ce qui concerne l'étude de la pénalité en fonction du cardinal de la collection de modèles considérée, ont été réalisées dans un contexte de régression par Li (1987) ou Polyak et Tsybakov (1990), entre autres. Ensuite Barron et Cover (1991) ont essayé de trouver une approche globale de la sélection de modèles dans des cadres particuliers. La méthode a enfin été généralisée et systématisée par les travaux de Birgé et Massart (1997, 1998) et Barron, Birgé, Massart (1999). Remarquons que, dans leur approche unificatrice, Birgé et Massart (1997) notent que les méthodes de seuillage en ondelettes développées auparavant par Donoho et Johnstone (1994, 2000) et Donoho, Johnstone, Kerkyacharian et Picard (1996) peuvent être décrites simplement avec le formalisme qu'ils mettent en place et sont étroitement apparentées à la sélection de modèles par pénalisation.
Illustrons la démarche et les objectifs de cette méthode. Dans le problème ci-dessus, en notant toujours $S_D$ l'espace engendré dans $\mathbb{L}_2$ par les $\varphi_{j,D}, j = 1, \ldots, D$ pour $1 \leq D \leq n$ variable, on va sélectionner

$$\hat{D} = \text{argmin}_{D=1,\ldots,n} \left( \gamma_n(\hat{f}_D) + C\frac{D}{n} \right), \tag{13}$$

où $C$ est une constante vérifiant $C \geq C_0$ où $C_0$ est connue ou estimable. La quantité $\text{pen}(D) = CD/n$ qui apparaît dans (13) est appelée fonction de pénalité. On peut alors montrer que l'estimateur résultant $\hat{f}_{\hat{D}}$ vérifie :

$$\mathbb{E}\|f - \hat{f}_{\hat{D}}\|^2 \leq K \inf_{D \in \{1,\ldots,n\}} \left( \|f - f_D\|^2 + \frac{D}{n} \right), \tag{14}$$

où $K$ est une constante dépendant éventuellement des données du problème (de $\|f\|$, de la base choisie, ...). On le voit, (14) est la garantie du compromis automatique − et non asymptotique − entre le carré du biais $\|f - f_D\|^2$ et la variance $D/n$.
On a défini ici une démarche consistant à :

1. Choisir un contraste $\gamma_n$ adapté au problème,

2. Choisir une collection de modèles $(S_D)$ sur lesquels on définit des estimateurs $\hat{f}_D$ minimisant $\gamma_n$ et parmi lesquels on souhaite opérer la sélection,

3. Trouver une fonction de pénalité pertinente $\text{pen}(D)$ pour choisir

$$\hat{D} = \text{argmin}_{D=1,\ldots,n} \left(\gamma_n(\hat{f}_D) + \text{pen}(D)\right),$$

4. Vérifier que l'estimateur $\hat{f}_{\hat{D}}$ vérifie une inégalité de type (14).

C'est là le programme mis en oeuvre dans [8], [9], [10], [14], [15] pour les fonctions précédemment énumérées. Les contrastes sont différents, les bases $\varphi_{j,D}$ éventuellement plus élaborées (polynômes par morceaux, ondelettes, polynômes trigonométriques) ou plus nombreuses (cas de subdivisions non régulières), mais l'objectif est toujours une inégalité de type (14). La difficulté provient du fait que les variables considérées ne sont jamais indépendantes : la suite est stationnaire dans [8], géométriquement $\beta$-mélangeante (i.e. avec des coefficients de mélange à décroissance exponentielle) dans [14], arithmétiquement (i.e. avec des coefficients de mélange à décroissance hyperbolique) ou géométriquement $\beta$-mélangeante dans [9], [10], arithmétiquement ou géométriquement $\beta$- ou $\alpha$- mélangeante dans [15].
Du point de vue technique, on se ramène à contrôler les déviations par rapport à son espérance du supremum d'un processus empirique recentré, $\sup_{t\in B_D(0,1)} |\nu_n(t)|$ où $B_D(0,1) = \{t \in S_D, \|t\| = 1\}$; par exemple en densité comme ci-dessus

$$\nu_n(t) = \frac{1}{n}\sum_{i=1}^{n} [t(X_i) - \mathbb{E}t(X_i)] .$$

En d'autre termes, on cherche à majorer

$$\mathbb{P}\left(\sup_{t\in B_D(0,1)} |\nu_n(t)| \geq \mathbb{E}\left[\sup_{t\in B_D(0,1)} |\nu_n(t)|\right] + \lambda\right)$$

par des termes exponentiellement petits.
L'inégalité de Talagrand (1994) fournit une telle majoration pour des variables indépendantes. Dans [15], en collaboration avec Florence Merlevède, les résultats sont obtenus en étendant cette inégalité au cas de variables dépendantes par des techniques d'approximations dues à Rio (1993) en $\alpha$-mélange et à Viennet (1997) en

$\beta$-mélange. Dans [9], en collaboration avec Yannick Baraud et Gabrielle Viennet, on utilise une inégalité de moment (de type Rosenthal pour des supremum), découlant elle-même de l'inégalité de Talagrand (1994) et démontrée par Baraud (2000) : on l'étend au cadre dépendant en utilisant encore la méthode d'approximation de Viennet (1997). Dans [10], en collaboration avec Yannick Baraud et Gabrielle Viennet encore, on utilise des techniques de martingales exponentielles pour établir une inégalité de type Bernstein (borne exponentielle pour la probabilité de déviation du processus $\nu_n$ sans sup) et un chaînage $\mathbb{L}_2$ pour passer au supremum par recouvrement de la boule unité de $\mathbb{L}_2[0,1]$. Dans [8], j'établis une inégalité de Bernstein encore, en utilisant les propriétés gaussiennes de la suite, et je passe au supremum par un chaînage $\mathbb{L}_2$-$\mathbb{L}_\infty$ (c'est-à-dire en contôlant aussi la norme infinie des termes du recouvrement et plus seulement leur norme 2).
Bien sûr, chacun des problèmes traités par les méthodes que l'on vient de présenter est à replacer dans son contexte. [9] et [10] systématisent les résultats de Polyak et Tsybakov (1990) afin de traiter des problèmes autorégressifs, donc avec dépendance, et non plus seulement régressifs en contexte indépendant. [14], en collaboration avec Yves Rozenholc, donne des résultats en risque global que l'on peut comparer à ceux obtenus point par point par Härdle et Tsybakov (1997) avec des polynômes locaux et par Fan et Yao (1998) avec des méthodes de noyaux. Neuman (1996) cherchait comme [8] à estimer une densité spectrale de façon adaptative, en utilisant une méthode de seuillage en ondelettes, sous des conditions de cumulants assez complexes exigées par son cadre de travail non nécessairement gaussien. En me cantonnant à des suites gaussiennes, j'envisage dans [8] des collections de modèles plus générales et des jeux d'hypothèses amenant à des résultats dans certains contextes de longue mémoire que les hypothèses usuelles excluent. La méthode est aussi à relier à l'idée de Golubev (1993) de sélectionner des poids pondérant le périodogramme (l'estimateur empirique usuel de la densité spectrale) parmi différentes famille de poids. Enfin, [15] généralise les résultats de Leblanc (1997) pour des processus à temps continu et de Tribouley et Viennet (1998) pour des processus à temps discret. Ces quelques dates illustrent que le domaine est en pleine mutation et au coeur d'une évolution sans doute encore inachevée. L'épreuve du feu sera celle de l'expérimentation et de la validation numérique (par simulation) et empirique (sur de vraies données). Bien entendu, de tels essais ont déjà été effectués (dans [8] et [14], voir aussi Birgé et Rozenholc (2000)) et d'autres sont en cours. Les résultats sont encourageants mais l'ajustement des constantes de pénalisation se révèle malgré tout un exercice délicat et numériquement lourd si l'on souhaite une procédure "quasi-universelle" c'est-à-dire si l'on veut trouver les facteurs multiplicatifs constants (la bonne valeur du $C$ de (13) qu'il faut utiliser dans la pénalité pour obtenir des résultats satisfaisants quelle que soit la fonction à estimer et notamment quelle que soit sa régularité.

**QUELQUES MOTS POUR CONCLURE.**

J'ai procédé pour cette présentation à une fragmentation de mon travail. Mais bien sûr ces différents thèmes ont été rencontrés en formalisant des préoccupations qui ne sont pas dépourvues d'unité, notamment en ce qui concerne les modèles envisagés : la plupart ont leur point de départ dans l'autorégression − d'ordre supérieur avec les modèles dits "ARMA", d'ordre 1 mais avec erreur de mesure, paramétriques ou non paramétriques, associé à une modélisation de la variance conditionnelle avec les GARCH, en temps continu avec les diffusions et les modèles à volatilité stochastique. C'est davantage le point de vue qui change : plus modélisateur ou plus statistique selon les cas.
De plus, la motivation constante qui a dicté le choix du titre de ce mémoire est claire : la gestion de la dépendance. Dépendance forte ou mémoire longue, dépendance directionnelle ou causalité, mélange $\alpha$ ou $\beta$, géométrique ou arithmétique, stationnarité : ce sont différentes facettes d'un même problème. Dans tous les cas, les questions qui se posent sont directement liées à cette dépendance : comment la modéliser, comment l'estimer ou comment estimer des paramètres ou des fonctions malgré l'absence d'indépendance.

## RÉFÉRENCES


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.

Alos, E., O. Mazet and D. Nualart (1999) Stochastic calculus with respect to fractional brownian motion with Hurst parameter less than 1/2. Preprint.

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

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. and Cover, T.M. (1991) Minimum complexity density estimation. *IEEE Transactions on Information Theory* 37, 1034-1054.

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. (2000) Gaussian models selection. A parître dans *J.E.M.S.*

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

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307-327.

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, 1082–1090.

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

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

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

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

Davydov,Yu A. (1970) The invariance principle for stationary processes. *Theory Probab. Appl.* 15, 487-498.

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.

Douc, R. and Matias, C. (2000) Asymptotics of the maximum likelihood estimator for general hidden Markov models. To appear in *Bernoulli.*

R. F. Engle (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* 50, 987–1007.

Engle, R. F., and Kroner, K. F. (1995), “Multivariate simultaneous generalized ARCH,” *Econometric Theory*, 11, 122–150.

Fan, J. and Truong, Y.K. (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.

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

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

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

Hosking, J.M.R. (1981) Fractional differencing. *Biometrica*, 68, 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.

Hurst, H.E. (1951) Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* 116, 770-779.

Kolmogorov, A.R. and Rozanov, Y.A. (1960) On the strong mixing conditions for stationary gaussian sequences. *Theor. Probab. Appl.*, 5, 204-207.

Leblanc, F. (1997) Density estimation for a class of continuous time processes. *Math. Methods of Stat.*, 6, 171-199.

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

Mandelbrot, B.B. and Van Ness (1968) Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10, 422-437.

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

P. Massart and Rio, E. (1998) A uniform Marcinkiewicz-Zygmund strong law of large numbers for empirical processes. *Asymptotic Methods in Probability and Statistics,*

*A volume in Honour of Miklós Csörgo*, 199-211.
Nelson, D.B. (1990) Stationarity and Persistence in the GARCH(1,1) Model. *Econometric Theory*, 6, 318–334.
Neuman, M.H. (1996) Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *J. Time Series Anal.*, 17, 601-633.
Polyak, B.T. and Tsybakov, A. (1990) A family of asymptotically optimal methods for choosing the order of a projective regression estimate. *Theory Probab. Appl.*, 37, 471-481.
Renault, E. and Touzi, N. (1998) Option hedging and implicit volatilities in a stochastic volatility model. *Mathematical Finance*, 6, 279-302.
Rio, E. (1993) Covariance inequalities for strongly mixing processes. *Ann. Inst. Henri Poincaré*, 29, 587-597.
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.
Rosenblatt, M. (1956) A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci.*, 4, 43-47.
Rozanov, Yu. A. (1967) *Stationary Random Processes*, Holden Day.
Stefanski, L.A. & Carroll, R.J. (1990) Deconvolution kernel density estimators. *Statistics*, 21, 169-184.
Talagrand, M. (1994) Sharper bounds for gaussian and empirical processes. *Ann. Probab.*, 22, 28-76.
Taupin, M.L. (1998) Semi-parametric estimation in nonlinear structural errors-in-variables model. *Preprint Université Paris-Sud* 98-35, tentitavely accepted to *Ann. Statist.*
Tribouley, K. and Viennet, G. (1998) $\mathbb{L}_p$ adaptive density estimation in a $\beta$-mixing framework, *Ann. Inst. Henri Poincaré*, 34, 179-208.
Viennet, G. (1997) Inequalities for absolutely regular sequences: application to density estimation. *Probab. Th. Rel. Fields*, 107, 467-492.
Whittle, P. (1953) The analysis of multiple stationary time series. *J. R. Statist. Soc.*, **B**, 15, 125-139.