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

"A few works on the study
of
dependence
in
Statistics and Econometrics"

Fabienne Comte[*]

September 2000

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

General content

The works presented in this report are split into 7 published papers and 7 papers submitted for publication. Here is the complete list:

Published papers :

1. "Simulation and estimation of long memory continuous time models", *Journal of Time Series Analysis*, 1996, 1, 19-36.
2. "Non Causality in Continuous Time Models", with Eric Renault, *Econometric Theory*, 12, 1996, 215-256.
3. "Long Memory Continuous Time Models", with Eric Renault, *Journal of Econometrics*, 76, 1996, 101-149.
4. "Regularized log-periodogram regression for Gaussian processes with bounded spectral density", with C. Hardouin, *Note aux CRAS*, t.325, Série I, p.1203-1206, 1997.
5. "Long memory in continuous time stochastic volatility models" with 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, to appear in *Journal of Time Series Analysis*.

Papers submitted for publication :

8. "Adaptive estimation in an autoregression and a geometrical β-mixing regression framework", with Y. Baraud et G. Viennet, Working paper Crest-Insee 9807 and tentatively accepted by the *Annals of Statistics*.
9. "Model selection for (auto)-regression with dependent data", en collaboration avec Y. Baraud et G. Viennet, Working paper ENS 99-12 and submitted for publication to *ESAIM*.
10. "Adaptive Estimation of the spectrum of a stationary Gaussian sequence", Working paper Paris 6 n^o 525, and tentatively accepted by *Bernoulli*.
11. "Kernel Estimation of a Stochastic Volatility Model", Working paper Paris 6 n^o 560, in revision for *Econometric Theory*.
12. "Asymptotic Theory for Multivariate GARCH Processes", with O. Lieberman, submitted for publication to *Journal of Multivariate Analysis*.
13. "Adaptive estimation of mean and volatility functions in (auto-)regressive models", with Y. Rozenholc, submitted for publication to *Stochastic Processes and Their Applications*.

14 "Semi-parametric estimation in the (auto)-regressive β-mixing model with errors-in-variables", with M.-L. Taupin, submitted for publication to *Mathematical Methods of Statistics*.

1 Introduction

There would be several ways of classifying the subjects on which I have been working during the last years: Statistics or Econometrics, Parametric methods or Model selection methods, Spectral or Time representations. The plan chosen here aims at separating the types of dependence considered between the variables. Statistical methods have first been developped to study independent identically distributed (i.i.d.) variables. But then, Economics started building some models for the random variables they studied. This is probably how Econometrics were born, which mainly aim at finding the right mathematical and especially statistical methods in order to deal with those variables and estimate the underlying parameters or functions. Then, even if some error terms remain i.i.d., the resulting variables − the output of the models − can no longer be assumed independent. Then arised the ideas of autoregressive models (dependence of the value of the variables today on their value in the past) or of stationary models (the variability is the same along time).

Another evolution, due to the emergence − and later to the hegemony − of Finance, was that the models instead of remaining in discrete time became continuous time models in order to take better in account the type of variation of price processes. This is the reason why many notions that had been first defined in discrete time must be now considered in continuous time: some new definitions must be given, whose interest must be checked and for which some characterizations must be found. On the other hand, as the observation sample remains a discrete time sample, the new continuous time models require in return to be discretized in order to find adequate methods of estimation and testing. Moreover, the validation of those methods imply some simulation study for which the rates of convergence of the discretized processes are wishable to be known.

My work is mainly built on these two directions: re-defining in continuous time notions such as causality (a dependence with a direction in time since the cause currently occurs before the effect), long memory or cointegration, with the associated questions about the statistical methods to use to estimate the parameters of the underlying continuous time models from discrete time samples; finding some new sophisticated statistical methods to study some very simple and standard models such as autoregressions (seen as a the most famous regression model with dependent variables).

Let us try to give some details. My very first work [2], written with E. Renault, was giving a new modelization of this particular dependence: causality. A definition of mean causality for discrete time variables was available, generally applied to economic series: this definition aims at measuring in mean the usefulness of the past of a variable in the prediction of another one. The recent developments of Finance required two types of extensions. First, if now the models are written in continuous time, is it possible to define, characterize and test a kind of causality which would be then more instantaneous and more structural? Second, since financial applications pay special attention to the volatility processes and their variations, is it possible to study, characterize and test some variance causality, in continuous time but also and first in discrete time, in ARCH models for instance? [2] answers mainly to the first question, [6] studies some cases of singularity in the models detailed in [2], and [7], written with O. Lieberman, studies the second question. The theoretical tools that legitimate the tests proposed in [7] are rigorously studied in [12], which is also written with O. Lieberman.
Then the works [1], [3], [4], [5], [10] study some models whose dependence properties are in some sense "stable" through some stationarity property. Note that this dependence can anyway be

very strong as in long memory models. A common point to all those contributions is also to be found in a systematic incursion in the frequency domain through a study of the spectral density. The point of view is more modelization in [3], [1] and [5] where the problem dealt with is to define stationary long memory continuous time models. The study of such models, represented as stochastic integrals with respect to the Fractional Brownian Motion has been widely developped eversince, from both the theoretical and applied point of view. [3] and [5] have been written with E. Renault. We are in a parametric setting. The estimation procedure remains however based on discrete time observations: this allows to use some statistical methods available for discrete time samples, provided that the study of the discretization and of the tranfer of long memory properties from the continuous time models to their discrete time version is lead. All those contributions end up in [5] with stochastic volatility long memory models for which we can propose both justifications and characterizations of the memory and methods of simulation and estimation. [4], written with C. Hardouin synthesizes a considerable amount of work on discrete time estimation methods based on regularized log-periodogram regression. These works did not lead up to any publications but allowed me to seriously improve my knowledges about estimation in the frequency domain. Lastly [10] develops a nonparametric adaptive method for the estimation of the spectral density of a stationary gaussian process using model selection, provided that this spectral density remains square-integrable; this implies that we remain mainly with short memory processes. Nevertheless, the method is proved to be robust to some types of long memory.

Lastly, the common thema to the works [8], [9], [11] [13], [14] lies in the dealing with a form of dependence known as β-mixing (or absolute regularity). This dependence is in all cases the result of autoregressive models with ([11], [14]) or without ([8], [9], [13]) measurement error. [8] and [9], written with Y. Baraud and G. Viennet, consider some regression models with mixing variables for error terms with constant variance. [13], written with Y. Rozenholc, is an extension of those result to autoregressions with some heteroskedasticity represented through a volatility function. An empirical study based on simulations is moreover lead in order to check the method. Moreover, when the economists want to take into account the variations of the volatility, there is a competition between the deterministic volatility function involved in an autoregressive and heteroskedastic model and the random volatility process associated to some specific source of alea involved in stochastic volatility models. The main problem in the last case is that this process is not directly observed, and this often implies slower rates of convergence of the estimators. Thus, [11] makes the link between stochastic volatility models and regressive models with measurement errors and extends to the mixing framework the properties of some deconvolution kernel estimators which are well-known in the independent setting. However the optimal rates of convergence of those estimates to the true function are often as slow as powers of $\ln(n)$. This is the reason why we look in [14], written with M.-L. Taupin, for a better specified regression function through a semiparametric study.

It is wishable to notice that went I in all my work from the theory to the applicability of the methods I proposed, and that almost all of them have been experimented either on simulated or on true data. It was indeed impossible not to go until this last step of the study to prove that all new methods were really feasible: whether to adapt old methods to new models or to developp new methods for old models.

2 Causality

2.1 Continuous time causality: [2], [6]

The first problem, studied in [2] with E. Renault, is the modelization of non causality whose most commonly used definition is based on the comparison of conditional means of a given process with respect to different σ-algebras seen as information sets.

More precisely, let $X(t) = (X_1(t)', X_2(t)', X_3(t)')'$, $t \in \mathbb{R}^+$, where $X_i(t)$, $i = 1, 2, 3$, are vectors of size $n_i \times 1$, $n_1 + n_2 + n_3 = n$ and $X_t = (X'_{1,t}, X'_{2,t}, X'_{3,t})'$, $t \in \mathbb{N}$, the analogous discrete time process. The third variable X_3 must be seen as an environment variable. To study the causality of X_1 onto X_2, we associate to those vectors, for each integer time $t \in \mathbb{N}$, the σ-fields:

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

By analogy, we define in continuous time, for a real time t,

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

The usual discrete time definition:

$$X_2 \text{ does not Granger-cause } X_1 \Leftrightarrow \mathbb{E}[X_{1,t+1} | \mathcal{I}_D(t)] = \mathbb{E}[X_{1,t+1} | \mathcal{I}_D^{(1,3)}(t)] \tag{1}$$

takes into account the idea that no knowledge about the past of X_2 is required for the prediction of X_1. However, this definition is not stable by temporal aggregation, that is it depend on the time unity which is chosen, denoted here (but arbitrarily) as unity. From this point of view, this notion is not structural and, for rapidly evolving processes, modelized in continuous time, it seems inadequate.

It can be seen then, thanks to Florens and Fougère (1996), that the discrete time Granger noncausality can be characterized for semi-martingale processes, in terms of the Doob decompositions associated to X_1. By analogy with this property, we work in the setting of càdlàg (right-continuous and left-limited) semi-martingales processes ($X(t) = X(0) + H(t) + M(t), H(0) = M(0) = 0$, M local martingale, H predictable), and we consider the definition:

Définition 2.1 *: X_2 does not immediately Granger-cause X_1 if the canonical decompositions of $X_1(t)$ with respect to $\mathcal{I}(t)$ and to $\mathcal{I}^{(1,3)}(t)$ are the identical (up to an evanescent process).*

Then we prove (section 2.2) that this definition can be characterized in terms of another extension which is more clearly related to the discrete time definition:

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

where lims means "limit in mean square", that is in $\mathbb{L}_2$. Of course, we must be in a setting where both limits exist: we study under what kind of regularity assumptions the notions are equivalent. Note that we work in a framework of continuous time continuous in $\mathbb{L}_2$ processes with finite variance, so that a definition in terms of the increments of the process only would lead to a trivially fulfilled condition.

Another kind of discrete time noncausality, the so-called instantaneous noncausality, is similarly extended and characterized (section 2.3). Both types of immediate (i.e. continuous time) noncausality are illustrated throughout a model which is particularly relevant in terms of a semi-martingale decomposition study (section 2.4); this is Hull and White's (1987) stochastic volatility model:

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

where $(w_1, w_2)'$ is a standard Brownian motion. This model can be related to several implicit noncausality assumptions as compared to the most general version that could have been chosen for a general two-dimensional stochastic differential equation written with respect to (S_t, σ_t). It even leads to consider another notion of noncausality namely a variance noncausality which is quite new, even in discrete time.
Next, we choose to apply our definitions (section 3.1) to processes that can more simply be written:

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

where W is an n-dimensional brownian motion and A a matricial and deterministic function. This framework is suggested by the continuous time Wold decomposition (cf. Rozanov (1967)) of gaussian stationary processes without deterministic component which can be written: $\int_{-\infty}^t A(t-s)dW(s)$. We study the semi-martingale decomposition of those processes, their invertibility i.e. the existence of a function B such that the representation

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

is available. Then we are able to characterize noncausality in terms of the matricial coefficients of the B function.

The previous results are in particular applied to processes solution of Stochastic Differential Equations (SDE from now on) with constant coefficients (section 3.4), whose theoretical framework has been presented by Bergstrom (1984):

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

where D denotes the derivation operator in $\mathbb{L}_2$. Since we know in that case the exact form of the Vector-ARMA$(p, p-1)$ process which is obtained by discrete time observation of the process, we can study the links between discrete and continuous time causality and illustrate the identifiability problem of the continuous time model from discrete time obsevrations.

The previous work studies a class of linear processes on which some regularity conditions are set: therefore, it is natural to wonder what happens is some linear singularity occurs in some of the considered representations. Such a kind of work already exists in discrete time − in the framework of ARMA processes mainly − and makes an extensive use of the very particular properties of random walks: it is gathered in the framework of its applications to Economics under the name of "Unit roots and cointegration problems".

Cointegration is the extension of the unit root notion (1 is a root of the autoregressive polynomial) to the multivariate setting where some "intermediate" singularities can occur (1 is a root of the real coefficients polynomial obtained as the determinant of the multivariate autoregressive polynomial): from the mathematical point of view, it is a problem of linear algebra in the field of polynomials with matricial coefficients. More precisely, some nonstationarity can appear in Vectorial ARMA models $\Pi(L)X_t = \varepsilon_t$ where $LX_t = X_{t-1}$, when Π has matricial coefficients $\Pi(L) = Id - A_1L - \ldots - A_pL^p$, if $\det(\Pi(1)) = 0$. If $\Pi(1) = 0$ then $(1-L)$ can be exactly factorized in the polynomial, the process is not stationary and we say that there is a "unit root". But 1 an be a root of the scalar polynomial $\det(\Pi(x))$ without $\Pi(1)$ being zero. Then the process is not stationary but some linear combinations of its components are. This happens for instance is the matrix $\Pi(1)$ is singular. From an economic point of view, the concept of cointegration is very interesting and powerful because it allows to describe a stationary relation (and thus an equilibrium) between variables which individually are not stationary. This is the reason why it seems so important to detect such properties and to study their consequences (on the properties of the estimators in particular).

[6] illustrates that similar problems can arise in a continuous time setting. We define here a notion of continuous time cointegration, we study its impact on the discretized processes and we provide characterizations, in the frequency domain in particular.

2.2 Second order causality: [7], [12]

When very volatile data are analysed, we are often more interested in conditional variance behaviour than in the conditional mean. This is obvious from the emergence of ARCH models. In such cases, we want to test variance or second order noncausality rather than mean noncausality. However, this has not been much studied, except in the bivaiate case some attempts by Granger, Robins and Engle (1986) and by Cheung and Ng (1996). This is the reason why [7], written with O. Lieberman, aims at defining studying and providing charaterizations and tests of these notions.

For sake of simplicity, we work in a framework without environment variable. All variables are also assumed to be square integrable. Let X_t, $t \in \mathbf{Z}$, an n-dimensional process with components $(X_{i,t})_{1\leq i\leq 2}$, each with dimension n_i, $i = 1,2$, $n_1 + n_2 = n$ et $X_{i,t} = (X_{ij,t})_{1\leq j\leq n_i}$. We denote here by $\mathcal{I}_D^{(1)}(t)$ the σ-field generated by X_1 and by $\mathbb{E}(.|\mathcal{I}(t))$ the conditional mean knowing $\mathcal{I}(t)$. We consider $\overline{X(t)}$, the Hilbert space generated by the variables $X_{i,s}$, $s \leq t$, $i = 1,2$ and $\overline{X_1(t)}$ the Hilbert space generated by the variables in $X_{1,s}$, $s \leq t$. We denote by $\Pi(X_{ij,t+1}|\overline{X(t)})$ the orthogonal projection of $X_{ij,t+1}$ on $\overline{X(t)}$ and $\Pi(X_{i,t+1}|\overline{X(t)})$ the vector $(\Pi(X_{i1,t+1}|\overline{X(t)}), \ldots, \Pi(X_{in_i,t+1}|\overline{X(t)}))'$. $\Pi(X_{i,t+1}|\overline{X_1(t)})$ is defined in a similar way. Note that $\mathbb{E}(X_{t+1}|\mathcal{I}(t))$ is the best predictor of X_{t+1}, whereas $\Pi(X_{t+1}|\overline{X(t)})$ is its best linear predictor. We recall with our simplified notations that:

$$(i)\ \ X_2 \begin{array}{c}\text{does not}\\ \text{Granger cause}\end{array} X_1 (X_2 \overset{G}{\not\to} X_1) \text{ if } \mathbb{E}(X_{1,t+1}|\mathcal{I}_D(t)) = \mathbb{E}(X_{1,t+1}|\mathcal{I}_D^{(1)}(t)),\ \ \forall t \in \mathbf{Z}.$$

$$(ii)\ \ X_2 \begin{array}{c}\text{does not linearly}\\ \text{Granger-cause}\end{array} X_1 (X_2 \overset{LG}{\not\to} X_1) \text{ if } \Pi(X_{1,t+1}|\overline{X(t)}) = \Pi(X_{1,t+1}|\overline{X_1(t)}),\ \ \forall t \in \mathbf{Z}.$$

If X_t is a gaussian process, both definitions (i) and (ii) are of course equivalent. Some difficulties arise when we want to generalize those definitions to the second order. The squares imply some nonlinearities that increase the difference between projections and conditional means. Whereas the σ-algebras need not being modified, the projection spaces must be adapted. Let $\overline{X^2(t)}$ the Hilbert space generated by the variables $X_{ij,s}X_{kl,s}$, $s \leq t$, $1 \leq i,k \leq 2, 1 \leq j \leq n_i, 1 \leq l \leq n_k$, and $\overline{X_1^2(t)}$ the Hilbert space generated par $X_{1j,s}X_{1l,s}$, $s \leq t$, $1 \leq j,l \leq n_1$. It is thus natural to consider the following definitions:

Définition 2.2 *(i) X_2 does not cause X_1 in variance ($X_2 \overset{V}{\not\to} X_1$), if $V(X_{1,t+1}|\mathcal{I}_D(t)) = V(X_{1,t+1}|\mathcal{I}_D^{(1)}(t)),\ \ \forall t \in \mathbf{Z}$.*

(ii) X_2 does not cause X_1 linearly in variance ($X_2 \overset{LV}{\not\to} X_1$), if $\forall t \in \mathbf{Z}$,

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

In the definition 2.2 (*ii*) the square of the vector is set for the product of the vector by its transpose. We denote by X' the transpose of X.

Granger, Robins and Engle (1986) gave a related definition of variance noncausality. To be precise, their definition amounts to second-order rather than variance noncausality:

Définition 2.3

(i) There is no second-order noncausality from X_2 onto X_1 ($X_2 \overset{GRE}{\not\to} X_1$) if $\forall t \in \mathbf{Z}$,

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

(ii) There is no linear second order noncausality from X_2 onto X_1 ($X_2 \overset{LGRE}{\not\to} X_1$) if $\forall t \in \mathbf{Z}$,

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

Equivalence relations between mean, second-order and variance noncausality are given below:

Proposition 2.1
$$\begin{array}{lcl} X_2 \overset{V}{\not\to} X_1 & \Leftrightarrow & (X_2 \overset{G}{\not\to} X_1 \text{ and } X_2 \overset{GRE}{\not\to} X_1). \\ X_2 \overset{LV}{\not\to} X_1 & \Leftrightarrow & (X_2 \overset{LG}{\not\to} X_1 \text{ and } X_2 \overset{LGRE}{\not\to} X_1). \end{array}$$

It follows from Proposition 2.1 that if mean causality is established, then so is variance causality, whereas there still can be or not second-order noncausality.
As an example, we can see that in vectorial ARMA models with martingale-difference innovations with constant variance, there is always second order noncausality. Therefore, the testing procedures developped by Boudjellaba, Dufour and Roy (1992, Section 4) for mean noncausality would be also relevant for testing linear variance noncausality, in this particular framework.
Moreover, we can give sufficient conditions for second order noncausality in processes with time-varying conditional variance matrix:

Proposition 2.2 *Let X_t be a sequence of random variables with $V(X_t|\mathcal{I}_D(t-1)) = H_t$, where H_t is a matrix process measurable with respect to $\mathcal{I}_D(t-1)$; Then*

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

We illustrate with some examples the usefulness and limits of this result. In particular, we give some sufficient conditions of noncausality in bivariate GARCH(1,1) models as well as in the ARCH factor model proposed by Engle, Ng and Rothschild (1990).
If the model is more precisely defined, in mean and not only in variance, then we are able to

provide equivalence relations for both mean and second-order noncausality. Thus, in the case of GARCH(p, q) models, we obtain a characterization of second order linear noncausality in terms of the matricial coefficients of the polynomials involved in the representation of the process. We can also reformulate (this is done in two corollaries of our main theorem) this noncausality condition as nullity conditions of some determinants. The interest of those results is that, if some estimators of these coefficients are available, then we can do some nullity and therefore causality tests.
All possible tests of nullity are discussed in Section 4. We consider likelihood ratio tests, provided that the asymptotic normality of the corresponding maximum likelihood estimators for multivariate GARCH(p, q) models is admitted. Since it was established in univariate or very simple cases only, this explains why [12], written with O. Lieberman again, was done to complete the theoretical part of our study.
Note that we lastly develop (section 5) an empirical 3-dimensional example concerning daily stock returns.

As mentioned right above, the maximum likelihood estimators for GARCH(p, q) models were in practice considered as consistent and asymptotically normal; but the theoretical part was somehow a little late on practitioners since only very simple cases had been completely studied, from the point of view of the existence of stationary ergodic solution as well as from the asymptotic properties point of view. Nelson (1990) gave stationarity conditions for the univariate GARCH(1,1), Lumsdaine (1996) proved the consistance and asymptotic normality of the quasi-maximum likelihood estimators for univariate GARCH(1,1) processes under a moment condition of order 32 and Lee and Hansen (1994) obtained the same result under slightly less strong conditions; Bougerol and Picard (1992) gave an equivalent condition of stationarity for a univariate GARCH(p, q) process and Boussama (1998) studied the consistency and asymptotic normality of the same model. Jeantheau (1998) was interested from the statistical point of view in the multivariate case, but he gave consistency criteria only and applied them to some particular case as well. [12] is interested in the asymptotic theory of multivariate GARCH(p, q).
Let $(X_t)_{t\in\mathbf{Z}}$ be a sequence of random vectors in $\mathbb{R}^d$ and let $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$. We assume that X_t is square integrable and satisfies:

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

Let

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

where X' is the transpose of X. The Engle and Kroner (1993) representation, called BEKK repesentation, of the multivariate GARCH(p, q) model is given by

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

where C is positive definite and the matrices A_{ij} and B_{ij} are $d \times d$ square matrices. k is an integer less than $\frac{d(d+1)}{2}$. The main advantage of this model is that it guarantees that H_t is positive definite. We denote by *vech* the operator that stacks the columns corresponding to the lower triangular part of a matrix. Then the model (5) can be written

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

for $\tilde{A}_i$, $\tilde{B}_i$'s deduced from the A_{ij} 's and B_{ij}'s. Without loss of generality, we take $k = 1$.

We denote by θ the vector of the parameters of the process, built with the coefficients of the matrices C, $\tilde{A}_i$ and $\tilde{B}_i$. The model is not assumed to be Gaussian but we work with the gaussian log-likelihood. Thus, the quasi-maximum likelihood estimator is defined as maximizing:

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

where $\det(A)$ denotes the determinant of the matrix A. We moreover denote by $\rho(A)$ the spectral radius of A. First we establish the strong consistency of our estimator using some consistency conditions given by Jeantheau (1998). Let

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

and assume that the ε_t's are i.i.d. Without loss of generality, we can choose $H_t^{1/2}$ symmetric. Let Θ be the space parameter and $\theta_0 \in \Theta$ the true value of the parameter. Then we have

Théorème 2.1 *For the GARCH(p, q) process defined by (4)–(7) and for $\hat{\theta}_n$ as defined above, we assume that:*

1. Θ is compact,

2. the model is identifiable in the sense of Engle and Kroner (1995),

3. the errors ε_t are i.i.d. and admit a density absolutely continuous with respect to the Lebesgue measure, positive in a neighborhood of the origin,

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

Then $\hat{\theta}_n$ is strongly consistent, that is: $\hat{\theta}_n \longrightarrow_{n\to+\infty} \theta_0,\ I\!P_{\theta_0} - a.s.$

The condition $\rho(\sum_{i=1}^{q} \tilde{A}_i + \sum_{i=1}^{p} \tilde{B}_i) < 1$ is a consequence of a result given by Boussama (1998) who establishes that this is a necessary and sufficient condition for the existence of a stationary and ergodic solution to the model for errors satisfying *2.*
Next we prove the asymptotic normality of $\hat{\theta}_n$, assuming first that the initial conditions for H_t are drawn in the stationary distribution. Basawa, Feigin and Heyde (1976) gave conditions to obtain the asymptotic normality of maximum likelihood estimators for general processes. Using the ergodicity and strict stationarity of our process (here is the interest of our particular initial condition), as well as a central limit theorem given by Billingsley (1968), the problem reduces to the checking of some simple conditions for our contrast process. Therefore, with standard tools of parametric statistics, we obtain:

Théorème 2.2 *Under the assumptions of Theorem 2.1, assuming moreover that the components of ε_t are independent, that X_t admits moments of order 8 and that the initial condition (for H) follows the stationary law,*

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

where C_0 and C_1 are some well defined matrices.

The result is then extended to the case of any deterministic initial condition for H (the link (7) between X and H allows to deduce the initial random value of X).

3 Long memory and stationarity

3.1 Continuous time long memory models: [1], [3], [5]

The very starting point of [3], written with E. Renault, is the study of a functional singularity in some continuous time moving average representations; this ended up by quite complete studies on a thema which has been eversince widely developped by other authors, from the point of view of stochastic calculus (Coutin-Decreusefond (1999), Carmona-Coutin (1999)) as well as in Finance (Hu, Oksendal et Sulem (1999)) where high frequency data are very convenient for both modelisations (long memory behaviour, continuous time models) and for the estimation procedures requiring very long data sets.
The problem was to consider processes $X(t) = \int_0^t A(t-s)dW(s)$ of the same kind as in [2], with now:

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

In such cases, either $A(0)$ does not exist $(-1/2 < \alpha < 0)$, or $A(0) = 0$ $(0 < \alpha < -1/2)$. This means that the semimartingale decomposition of the regular case studied in [3] is no longer available and that the representation no longer can be inverted. For $\tilde{A} \equiv Id$, we obtain in particular the process:

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

known as fractional Brownian motion of order α, to which we want to give the role usually played by the standard brownian motion ([1]). Nevertheless, we must be aware that the deduced processes no longer have any martingale nor markovian properties, which makes both arguments and computations more complicated; those properties are mainly replaced by a self-similarity property (X is H-self-similar if $\forall a > 0$, $X(at)$ and $a^H X(t)$ have the same probability distribution).

The kind of models found in that way appears as the continuous time analogous of discrete time ARIMA(p, α, q), $\alpha \in \mathbb{R}$, also called fractional ARMA processes of order α:

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

developped by Granger-Joyeux (1980) and Hosking (1981).

We study (section 2.1) under which conditions we can define in $\mathbb{L}_2$ processes $Y(t) = \int_0^t D(t-s)dX(s)$ with $X(t) = \int_0^t A(t-s)dW(s)$; this allows in particular to define $\int_0^t D(t-s)dW_\alpha(s)$. Thus, for X defined by (8), we can look for a representation (section 2.2) :

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

[1] The fractional brownian motion is sometimes defined as

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

where the last term corresponds to $W_{H-1/2}(t)$, in order to obtain stationarity properties.

and study the link between the functions A and C: these are derivation/integration of order α links. Moreover, if C is continuously differentiable and if the matrix $C(0)$ is invertible, we can invert the representation into:

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

Both of these representation lead to a decomposition of the processes which is now of type: fractional brownian motion + predictable process. Lastly we define a fractional derivative (section 3) $X^{(\alpha)}$ of processes X of type (8) which allows to compare the representations:

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

This appears as the analogous of a result proved by Geweke and Porter-Hudak (1983). They establish by considering the class of the corresponding spectral densities, an equivalence between the class of ARIMA(p,α,q) processes given by (9) and the class of processes

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

which are ARMA type processes, with respect to a "fractional noise", ε_α. The power of the continuous time modelling lies in the fact that it also implies a functional invariance property (the fonction C is the same in both representations) whereas in discrete time, nothing is known about the links between (Φ, Θ) and $(\tilde{\Phi}, \tilde{\Theta})$; indeed, the process ε_α has no "good" property and considerably depends on the chosen time unity.

The long memory framework is met (section 2.3) when $\alpha \in]0, 1/2[$ under an asymptotic stationarity condition:

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

Under (10) and for $\alpha \in]0, 1/2[$, we prove that the stationary version of X, still denoted by X:

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

satisfies:

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

where γ_X is the autocovariance function ($\gamma_X(h) = \mathrm{cov}(X(t), X(t+h))$), and c is a constant, and also satisfies

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

where f_X is the continuous spectral density ($f_X(\lambda) = \int e^{-i\lambda h}\gamma_X(h)dh$). These properties are exactly the ones which usually characterize long memory phenomena. Lastly, we can check (section 3.1) that the conditions required by Robinson (1996) for a consistent and asymptotically normal semiparametric estimation of α are fulfilled by f_X^Δ :

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

where f_X is defined above; f_X^Δ is the spectral density of the discrete time process X observed every $k\Delta$ times, for k nonnegative integer, if these are the available observations.

An interesting example of the previous kind of processes is given by the extension of SDE's with constant coefficients (2) into "fractional" SDE in which the usual brownian motion is replaced by a fractional brownian motion: if $Y(t)$, solution of (2), admits the representation $Y(t) = \int_0^t C(t-s)dW(s)$, then $X(t) = \int_0^t C(t-s)dW_\alpha(s)$ is considered as the solution of the following fractional SDE:

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

The spectral density of X is then given by

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

Section 3.2 is dedicated to an application to Finance. More precisely we can, thanks to such models, extend the theory of the term structure of interest rates. In the models built in that way, long memory phenomena arise in the behaviour of the yield of a long bond and of the forward rate: this is in accordance with what is observed.

Lastly, another example built using SDE of order 1 illustrates (section 3.3) that long memory behaviour can be obtained from the aggregation of short memory individual equations.

Then [1] aims at studying by Monte Carlo experiments two methods of estimation for the parameters of processes of the previous type. To do so, we present the models which are considered and the theoretical properties that should be illustrated. Then we study some simulation methods. We consider the solution of a fractional scalar SDE of order 1:

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

where the parameter θ is $\theta = (a_0, \alpha, \sigma)'$. Of course, the methods can be generalized to higher order SDE's. We want to generate some sample paths $x(h), x(2h), \ldots, x(mh)$ for some small h, in order to build an "observation" sample $(x(\Delta), x(2\Delta), \ldots, x(N\Delta)$ with $\Delta = ph$ (generally $\Delta = 10h$).

The first idea (section 3.1) consists in exploiting the fact that we have a stationary gaussian process and to use a method proposed in that case by Davies and Harte (1987) based on the spectral representation of the process.

Next we study some discretisation schemes of Euler type (section 3.2); the first scheme is based on the representation $x(t) = \int_0^t a(t-s)dw(s)$:

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

for which we prove that for all $t \in [0,T]$, $\mathrm{V}(x(t) - \tilde{x}(t)) \le C.Th^{2\alpha} = C.mh^{2\alpha+1}$, $T = mh$.

Then we check a tightness criterion given by Billingsley (1968) for the sequence:

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

which implies the weak functional convergence $\tilde{x}_n \Rightarrow x, \ \ n \to +\infty$.

The second discretisation scheme is based on the fact that the derivative of order α, $x^{(\alpha)}$, of x follows a usual SDE of order 1 and therefore admits an exact AR(1) discretization which can be written:

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

Thus, the Euler scheme here must be applied to the integration of order α of $x^{(\alpha)}$ since $x(t) = \int_0^t \frac{(t-s)^\alpha}{\Gamma(1+\alpha)} dx^{(\alpha)}(s)$:

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

$\mathrm{V}(x(t) - \overline{x}(t))$ has the same order as previously and

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

This last method is the one finally used.

This implies the following question: if we need a discrete fractional filter to generate the data and maybe to estimate the parameters, why not use the usual fractional filter, corresponding to long memory ARMA models, that is the operator $(1-L)^\alpha$? Section 3.3 answers to this question by showing that the filter $(1-L)^{-\alpha}$ applied to our long memory process x gives indeed in output a short memory process, but that this process is not and has no reason to be an AR(1): this gives *a priori* an MA(∞) process. This means that such a filter does not preserve the short term structure of the considered process. An empirical illustration of this theoretical argument is given.

Next we study the estimation problem. The first method (section 4.1) proposed by Fox-Taqqu (1986) and Dahlhaus (1989), minimizes with respect to θ an approximation of the likelihood in the frequency domain (Whittle-type estimator):

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

where $I_n(x)$ is the periodogram and $f_{X,\theta}^\Delta$ is still given by (13) for $p = 1$.

The other experimented method (section 4.2) consists in using the semiparametric results of Robinson (1996) for the estimation of α: it is an improvement of the log-periodogram regression of $\ln\lambda$ for small λ (taking advantadge of the fact that $f_{X,\theta}^1(\lambda) \sim d.\lambda^{2\alpha}, \ \ \lambda \to 0$) originally proposed by Geweke and Porter-Hudak (1983). The idea then is that by approximate derivation to the approximate order:

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

we should obtain an approximation of $x^{(\alpha)}$, which can be considered as AR(1) for the estimation of c_0 and σ. The performances of this approximate derivation have been experimented in the first part of the study and found to be very acceptable.

The estimation results (section 4.3) found by Monte Carlo experiments are quite good for a_0 and α, but σ is underevaluated. At last, the second method is much faster that the first one for similar results.

In [5], written with E. Renault, we study a model for stock prices called "stochastic volatility model":

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

where $S(t)$ denotes the underlying stock price and $\sigma(t)$ its instantaneous volatility, $(w^1(t), w^2(t))$ is a 2-dimensional brownian motion and w^2_α is the fractional brownian motion obtained by fractional integration of w_2. For $\alpha = 0$, that is without the fractional aspect, these models have been studied by Renault and Touzi (1996): in particular, they illustrated that they could with them give account for the so-called "smile" effect observed on the markets by practitioners. Those models have been first proposed to take into account some specific alea of the volatility by Hull and White (1987), Scott (1987) and Melino and Turnbull (1990). [5] aims at an extension of these models in order to also take into account the volatility persistence met by practtioners on the markets.

Indeed, high frequency data often exhibit near integrated behaviours in conditional variance processes. In discrete time, many ARCH type models with unit roots have consequently been studied. In this work, we introduce long memory in the volatility process modelized in continuous time by simply replacing the usual brownian motion w_2 driving the stochastic volatility equation by the fractional brownian motion w^2_α, with α in $0 \leq \alpha < \frac{1}{2}$ (instead of $|\alpha| < \frac{1}{2}$ as authorized by the general definition because long memory occurs on that field only). Of course for nonzero α, w^2_α is no longer a semi-martingale (see Rogers (1997)), and the standard theory of stochastic integration is no longer available. But proceeding as in [3], we only use $\mathbb{L}_2$ theory and Wiener integrals for the integration of gaussian processes. Moreover, the option prices, despite they are functions of the volatility, keep the semi-martingale property, as well as the price process. This semi-martingale property is crucial for stock prices since non semi-martingale processes do not admit an equivalent martingale mesure. Indeed we know from Delbaen and Schachermayer (1994) that the NFLVR (No Free Lunch with Vanishing Risk) property implies the existence of a martingale equivalent measure (i.e. of a neutral risk probability). We must mention that we do not avoid the standard problem of such models: the non-uniqueness of the neutral-risk equivalent measure (incomplete markets).

Thus we study first in [5] (section 2) the probabilistic properties of our fractional stochastic volatility model. We give in section 3 the useful tools for an empirical study by looking for discrete approximations of these models. This allows to illustrate the advantages of the continuous time modelization with respect to its discrete time versions such as the FIGARCH (Fractionally Integrated GARCH) models that can be met in the literature. Section 4 proves an option pricing formula and studies the characteristics of implies volatilities. In particular, the long memory property changes the behaviour of conditional heteroskedasticity and kurtosis coefficients with respect to the maturity time. Next, we study in section 5 the statistical aspect of the problem. If the instantaneous volatility process was observed, the methods proposed in [3] may have been applied. Since it is not, we can think of using the indirect inference method proposed Pastorello, Renault and Touzi (1999) in the context of an Hull and White (1987) type model. But, since for $\alpha \neq 0$, the volatility process is no longer markovian and the use of the option pricing formula is much less convenient. Thus we alternatively suggest to use the approximate discretization of $S(t)$ to obtain approximations of the instantaneous volatilities and to work with approximate likelihood as well. At last, we use in section 6 the discretizations of σ computed in section 3 to generate some sample paths in order to illustrate the theoretical properties that have been previously studied and to compare the suggested estimation procedures.

3.2 Log-periodogram regression: [4]

The method studied in [4], written with C. Hardouin, did not end up with any published paper. The underlying working papers however are still quoted in some papers of econometric or statistical reviews (Velasco (2000)).
Estimation procedures based on log-periodogram regression have been first proposed by Kashyap and Eom (1988) and Geweke and Porter-Hudak (1983), but in the fractional framework of ARIMA(p,d,q) processes, with $-\frac{1}{2} < d < 0$. The estimator obtained by this method are however not efficient, even in the Gaussian case. We proved in [4] that we could recover consistency, asymptotic normality and asymptotic efficiency in the usual framework (bounded spectral density) provided that a "regularization" of the periodogram is performed. This procedure works for a wide class of models and is particularly relevant for models involving exponent parameters. We think for instance of MADE processes described by Gonçalves (1987): these are transformations of white noise processes by filters of type $\prod_j (1-\alpha_j L)^{d_j}$, where L is the usual lag operator ($LX_t = X_{t-1}$). The cases $|\alpha_j| < 1, d \in \mathbb{R}$ belong to our class of models. The estimation procedure is then very fast since the regression equation becomes a linear function of the d_j's. Nevertheless, it requires in all cases a great number of data.
More precisely, we consider the following model:

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

observed for $t = 1, \cdots, n$, n even. Let $X_n(\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t\; e^{it\lambda}$, $\varepsilon_n(\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t\; e^{it\lambda}$ be the discrete normalized Fourier transforms of X and ε observed on $\{1, \ldots, n\}$, and $I_{nX}(\lambda) = |X_n(\lambda)|^2$, $I_{n\varepsilon}(\lambda) = |\varepsilon_n(\lambda)|^2$ the periodograms of X and ε. We define the pseudo-periodogram of X: $J_{nX}(\lambda) = |A(e^{i\lambda})|^2\; I_{n\varepsilon}(\lambda) = 2\pi F(\lambda) I_{n\varepsilon}(\lambda)$, where F is the spectral density of X. Then the regression equation can be written:

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

The procedure provides an estimator of the parameters by considering the regression equation (16) at frequencies $\lambda_k = \frac{2\pi k}{n}$, $\;k = 1, \ldots \frac{n}{2} - 1$. This is the principle of the idea proposed by Kashyap et Eom (1988) for an ARIMA$(0,d,0)$ process, up to our regularization (by mean on sub-blocks of the observation set) and up to the replacement of J by I (which is fundamental since only I is observed).

In [4], we study in section 2 the estimation of parametric gaussian models (15) based first on the pseudo-periodogram. To obtain efficiency we replace the pseudo-periodogram J_{nX} by its regularized version and consider the regression (16) consequently modified. Then we can use some results about the order of the approximation of I_{nX} by J_{nX} to study the regression procedure based on the true periodogramm (section 3). This step is crucial since J_{nX} is unknown. Lastly we prove in section 4 that our method and Whittle's (1953) are asymptotically equivalent. Some extensions to the fractional or the non gaussian framework were considered in the preprints.

3.3 Adaptive estimation of the spectral density of a stationary gaussian process: [10]

The knowledge I had after [4] about the spectral density explains why I thought of applying in that framework the methods of adaptive estimation which were always more widely used in statistics, and in particular the method of model selection which is explained below.

I consider in [10] the problem of nonparametric estimation of the spectral density f of a real gaussian stationary sequence from a discrete sample $X = (X_1, \ldots, X_n)$.

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

where $\gamma_k = \mathrm{cov}(X_t, X_{t+k}),\ t \in \mathbf{Z}$. If $\hat{f}$ is an estimator of f, the risk of $\hat{f}$ is measured in $\mathbb{L}_2$-norm:

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

where $\|g\|$ denotes the usual $\mathbb{L}_2$-norm: $\|g\|^2 = \int_{-\pi}^{\pi} g^2(\omega) d\omega$.

In [10], I estimate f thanks to a projection estimator and by choosing the best projection space as possible. For that purpose, I introduce collections of finite dimensional linear spaces $\{S_m, m \in \mathcal{M}_n\}$ (the S_m's are called models) and I associate to each S_m the projection estimator $\hat{f}_m$ of f on this space, that is $\hat{f}_m$ minimizes for all t in S_m, the contrast function:

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

where

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

Under adequate assumptions, it is easily shown that, if s denotes a constant depending on f, the risk of $\hat{f}_m$ is bounded by:

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

where f_m is the orthogonal $\mathbb{L}_2$-projection of f on S_m.
The aim here is to propose a model selection procedure based on data only and that would be able to choose some $\hat{m}$ in $\mathcal{M}_n$ in such a way that the selected projection estimator $\hat{f}_{\hat{m}}$ has a better $\mathbb{L}_2$-risk of all the $\hat{f}_m$'s of the collection. This is performed thanks to a penalized criterion:

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

where pen is a penalty function from $\mathcal{M}_n$ into $\mathbb{R}^+$. Of course, the main difficulty lies in the choice of the penalty. From (18), we can see that the kind of performance that we are looking for (and that we obtain) can be written:

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

where C is some multiplicative constant. We find $L_m = 1$ or $\ln^2(n)$ depending on the collections of models which are considered (and in particular on the size of the collection). Thus, when

$L_m = 1$, our estimator minimizes the risk over the whole collection, up to a multiplicative constant: this corresponds to the case of regular models **(T)**, **(RP)**, **(W)** described in Appendix. We find $L_m = \ln(n)$ for general collections (**(GP)**, **(GW)**).
The method of model selection and the techniques used in the proofs are inspired by the works of Birgé and Massart (1997, 1998) and Barron, Birgé and Massart (1999). The problem of nonparametric estimation of spectral density was considered by Bentkus and Rudzkis (1976) and Bentkus (1985) with non-adaptive methods (the regularity of f must be a priori known). Nevertheless they compute the minimax rates of the estimators over some classes of smoothness of f. Golubev (1993) proposes a procedure for a weighted periodogram, which is adaptive with respect to the choice of the weights. His estimator is asymptotically minimax over some collection of neighbourhoods of the true spectral density, under assumptions which are slightly stronger than the one set in this work. Lastly Neumann (1996) considers a projection estimator in a non gaussian framework. His conditions on f are mainly given in terms of the cumulants of the sequence (X_i). He uses a thresholding wavelet method with first a fixed (but unknown) threshold and next a random threshold; this method is inspired by the works of Donoho and Johnstone (1994, 1998), and Donoho, Johnstone, Kerkyacharian, Picard (1996), and reaches the optimal rate on usual Besov spaces up to a $\ln(n)$ factor. The adaptive method that we use here allows to recover the true rate.
Let us say a few words about the assumptions. We assume that the X_i satisfy:

($\mathbf{H}_X$) The sequence $(X_1, \dots, X_n)$ is a stationary gaussian sequence.

Moreover, the following assumption on f implies that it is bounded:

($\mathbf{H}_f$) The autocovariance function is such that $\sum_{k\in Z} |\gamma_k| = M < +\infty$ and $\sum_{k\in Z} |k\gamma_k^2| = M_0 < +\infty$, $\gamma_k = cov(X_t, X_{t+k})$.

Let then $(S_m, m \in \mathcal{M}_n)$ be a collection of models such that the S_m's are linear spaces of functions with dimension D_m. We can give a general formulation of the assumptions that they must fulfill but let us only say for simplicity that these assumptions are fulfilled by all the models described in Appendix.

Using an $\mathbb{L}_2$-$\mathbb{L}_\infty$ chaining method, we prove our main theorem for general models (that is non necessarily regular models); we assume first that the bound on f is known. The Bernstein inequality which is used at the very begining of the proof makes a crucial use of the gaussian assumption; therefore, an extension to the non-gaussian framework would not be straightforward.

Théorème 3.1 *Assume that* ($\mathbf{H}_X$) *and* ($\mathbf{H}_f$) *hold and consider a general collection of models satisfying* $\sum_{m\in\mathcal{M}_n} e^{-L_m D_m} \le \Sigma < +\infty$, *for some positive weights* L_m. *Then the penalized projection estimator defined by:*

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

with

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

where κ is a (universal) numerical constant and $C_{\bar{r}}$ depends on the collection of models, satisfies

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

where C and C' are universal constant.

Of course, the main problem of Theorem 3.1 lies in the fact that the (deterministic) penalty depends on an unknown quantity: $\|f\|_\infty$. It is thus replaced by an estimator, which leads me to consider a random penalty where $\|f\|_\infty$ is replaced by the supremum norm of $\hat{f}_n^P$, the orthogonal projection of the periodogram on a space of regular piecewise polynomials (family **(DP)**) of dimension depending on n and denoted by N_n. We denote by f_n^P the $\mathbb{L}_2$ orthogonal projection of f on the same space. Then we find an extension of the inequality (21), under the same assumptions as previously and if $\|f - f_n^P\|_\infty \leq (1/4)\|f\|_\infty$ and $N_n \leq \dfrac{1}{36(r+1)^2}\dfrac{n}{\ln(n)}$, for the estimator $\tilde{f}$ defined by (20) with

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

Of course, the condition $\|f - f_n^P\|_\infty \leq (1/4)\|f\|_\infty$ must be studied. But it is fulfilled as soon as f is assumed to satisfy some smoothness assumptions. Indeed if $f \in \mathcal{B}_{\beta,p,\infty}$, $\beta > 0$, then we know from Devore and Lorentz (1993) that for $\beta > 1/p$,

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

where the definitions are briefly recalled below. This implies a constraint $N_n \geq C(\beta, f)$ which is satisfied when n is great enough.

Some other results can be proved in the specific case of trigonometric bases, which are particularly relevant for functions admitting Fourier developments. In that framework, we can extend results of the kind of (19) to nonbounded spectral densities provided that they remain square-integrable. These extensions require some more precise computations in terms of the spectral density but rely on direct deviation inequality which are obtained without the use of chaining method.

Lastly, let us give an illustration of how inequalities as (21), can lead to adaptation results with respect to the unknown regularity of f. Let us simply recall that f belongs to some Besov space $\mathcal{B}_{\beta,p,\infty}([0,1])$ if

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

where $w_d(f,y)_p$ denotes the modulus of smoothness. For a precise definition, see DeVore and Lorentz (1993), Chapter 2, Section 7, where it is moreover proved that $\mathcal{B}_{\beta,p,\infty}(A) \subset \mathcal{B}_{\beta,2,\infty}(A)$ for $p \geq 2$. This is the reason why we restrict our attention to $\mathcal{B}_{\beta,2,\infty}(A)$.

If the S_m's are one of the three families **(T)**, **(RP)** or **(W)**, it results from classical theorems of approximation theory that, if $r \geq \beta$ and $f \in \mathcal{B}_{\beta,2,\infty}([-\pi,\pi])$,

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

where r is the regularity. See for instance Barron, Birgé, Massart (1999) Lemma 12. p.404. See also Birgé et Massart (2000). Thus we obtain:

Corollaire 3.1 *Assume that* $(\mathbf{H}_X)$ *and* $(\mathbf{H}_f)$ *hold and assume that* $f(2\pi.)$ *belongs to a space* $\mathcal{B}_{\beta,2,\infty}([0,1]) \cap \mathbb{L}_\infty([0,1])$. *Assume moreover that* $\{S_m\}_{m\in\mathcal{M}_n}$ *is one of the families* **(P)** *or* **(W)** *with* $\beta > 1/2$. *Let* $\tilde{f}$ *be defined by (20). Then*

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

where the constant C *depends only on* $\beta, M_1, \|f\|_\infty$ *and* $C_{\bar{r}}$.

In other words

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

where the constant C depends here on $\beta, R, M_1, \|f\|_\infty$ and $\bar{r}$ and where $B_{\beta,2}(R) = \{f \in \mathcal{B}_{\beta,2,\infty}, |f|_{\beta,2} \leq R\}$.
We know from Bentkus (1986) that the minimax risk over a Besov ball $\{f/|f|_{\beta,2} \leq L\}$ can not be smaller that $C'L^{\frac{2}{1+2\beta}}n^{-\frac{2\beta}{2\beta+1}}$ where C' is a positive constant. Consequently, our estimator is minimax (up to some constants) over all these balls simultaneously.

4 Regression in a β-mixing context

4.1 Regression models: [8], [9], [13]

In the works [8], [9] et [13], we study the following model:

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

where the ε_i are i.i.d. centered random variables with unit variance. This model is a particular case of the standard regression model:

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

where the u_i's are i.i.d. centered with unit variance but where the (X_i, Y_i) are not assumed to be independent; on the contrary, they are assumed to be β-mixing. [8] and [9], written with Y. Baraud and G. Viennet, aim at adaptively estimating b under the assumption that $\sigma(x) \equiv \sigma_2$ where the variance σ_2^2 is known. [13], written with Y. Rozenholc, considers the general case and aims at adaptively estimating both functions b and σ^2.
There have been many studies concerning model selection, based on penalized criteria of the same type as Mallows' C_p, such as Akaike or BIC criteria, for both regressive or autoregressive models (Akaike (1973), Shibata (1976), Li (1987), Polyak et Tsybakov (1992), among others). All the results are asymptotic. The extension of the penalized criteria to model selection procedure based on the data only has been done by Barron (1991), Barron and Cover (1991), Rissanen (1984), and more recently, in the context of a more general framework than the only regressive one, by Birgé and Massart (1997) and Barron, Birgé and Massart (1999). Their point of view is now non asymptotic. Nevertheless, the very general aspect of their work leads them to set on their model some restrictions which are not wishable (bounded regression function, for instance). [8] et [9] relax these restrictions but keep giving non asymptotic results; these works lie on other results obtained by Baraud (2000) in a fixed design regression context and on coupling methods developped in Viennet (1997). [13] studies the problem of the estimation of the mean in presence

of a volatility and the problem of the estimation of the volatility σ^2 itself. Yao and Tong (1994) and Härdle and Tsybakov (1997) studied adaptive estimation of σ^2 but they do not manage to be simultaneously adaptive with respect to the mean and the variance. Hall et Carroll (1989) use estimators which are similar to ours but in a non adaptive framework. We mainly compare our results to those given by Fan et Yao (1998), from the theoretical point of view (but their regularity assumptions are strong and their results are asymptotic) and from the empirical point of view (since they give some simulation results to which we compare ours). Let us mention that Hoffman (1999) obtains by wavelet thresholding some asymptotic results leading to the minimax rate for the mean and to the minimax rate up to a $\ln(n)$ factor for the volatility.
Let us consider model (24). If $\hat{f}$ is an estimator of f, where f is b or σ^2, then the risk of $\hat{f}$ is measured with the $\mathbb{L}_2$-empirical norm:

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

The choice of this measure for the risk is discussed at length in [8]: let us schematically say that the minimizer of $(1/n)\sum_{i=1}^{n}(X_{i+1} - t(X_i))^2$, for $t \in \mathbb{L}_2(\mathbb{R}, dx)$ on a linear subspace $S \subset \mathbb{L}_2(\mathbb{R}, dx)$, denoted by $\hat{f}_S$, can not be unique whereas the vector $(\hat{f}_S(X_1), \dots, \hat{f}_S(X_n))$ always is. Moreover, under adequate assumptions, this risk can be decomposed into a sum bias + variance term:

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

where f_S is the $\mathbb{L}_2(dx)$-orthogonal projection of f onto S, $\|t\|_\mu^2 = \mathbb{E}(t^2(X_1))$ and κ depends on some constants of the problem.
Here is the procedure we use. It is similar at each step to the method used in [10]. We consider one or two collections of models denoted by $\{S_m^{(i)}, m \in \mathcal{M}_n^{(i)}\}$, $i = 1$ for b and $i = 2$ for σ^2; each $S_m^{(i)}$ is a finite dimensional subspace of $\mathbb{L}_2(\mathbb{R})$. Of course, it is not required that b nor σ belong to any of these models. Let $\hat{b}_m$ be the mean squares estimator of b on $S_m^{(1)}$ associated with the contrast:

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

for the model (24) based on the observations $X_1, \dots, X_{n+1}$. We use a procedure that automatically chooses $\hat{m}_1$ in $\mathcal{M}_n^{(1)}$ and which reaches the best compromise between the bias and variance terms:

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

where b_m is the orthogonal projection of b on $S_m^{(1)}$ and C is a multiplicative constant depending on the problem. This means that, up to a multiplicative constant, the procedure chooses the best model in the collection. This is obtained by choosing $\hat{m}_1$ as minimizing $\gamma_n^{(1)}(\hat{b}_m) + \mathrm{pen}^{(1)}(m)$ among all m in $\mathcal{M}_n^{(1)}$, where $\mathrm{pen}^{(1)}$ is a function to be determined, called a penalty function.
In a second stage, $\hat{\sigma^2}$ is estimated in an analogous way with respect to the contrast:

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

for $t \in S_m^{(2)}$, with an aim similar to (27), b being replaced by σ^2. We use here a penalty function $\mathrm{pen}^{(2)}(m)$.
We find one ([8]) or two ([13]) penalty functions $\mathrm{pen}^{(i)}(m)$, $i = 1, 2$ of order $\dim(S_m^{(i)})/n$. The models that are considered in [8] and [13] are exclusively regular. In the general case, we are lead to choose in [9] a penalty of order $L_m \dim(S_m^{(1)})/n$ where (see [10]) L_m is of order $\ln(n)$ (general models) or 1 (regular models).
It is important to notice that estimators satisfying inequalities as (27) have interesting properties: they are adaptive in the minimax sense over numerous classes of smoothness (see Barron *et al.* (1999), Birgé and Massart (1997)) and we can go as explained in section 3.3 for [10] from inequalities of type (27) to corollaries similar to Corollary 3.1, and to inequality of type (23). This means again that for $L_m = 1$, we obtain the minimax rates simultaneously over classes of Besov balls.
More precisely, [8] is in a framework of errors ε admitting moments of order $p > 2$ and of variables X_i (autoregression) or (X_i, Y_i) (regression) geometrically or arithmetically β-mixing. The variance function is constant. To obtain inequality of type (27) requires a precise control of

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

where $B_{m'}(0,1)$ is the unit ball of $S_{m'}$. It also requires a tedious treatment of the mixing that relies on Berbee's Lemma and on approximation methods for mixing variables as presented in Viennet (1997). The bound for the risk is obtained in the case of regular models only. We illustrate throughout some detailed examples the delicate compromise that must be performed between the rate of mixing, the maximal dimension that can be considered for the linear spaces of the collection of models and the minimal order p required for the errors.

[9] remains in a framework where the variance is constant and known and where the variables X_i or (X_i, Y_i) are geometrically or arithmetically β-mixing. But now, we assume that the errors are sub-gaussian i.e. such that for any real u,

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

for some positive s ($s = \sigma_2$ if the variables are gaussian or corresponds to the bound for bounded variables). Then we can consider general collections of models. Let us mention that the tools used in the proofs to obtain deviation inequality are very different from the one used in [8] since we use now the martingale property of the sequence $\sum_{i=1}^n \varepsilon_i t(X_i)$ to establish a Bernstein inequality leading to an $\mathbb{L}_2$-chaining method. Moreover, we develop in [9] some precise and detailed adaptation results in the minimax sense (in particular in "wild" Besov spaces $\mathcal{B}_{\beta,p,\infty}$ for $p < 2$).
Lastly [13] extends the previous results to the case of a volatility function and estimates σ^2 in a second step. We are under moment conditions on the errors of order p: $p > 8$ or $p > 16$ depending on the function to be estimated and we assume that the β-mixing is geometrical.
In the first step, we have

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

and

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

where κ is a numerical constant, Φ_1^2 a constant depending on the collection of models and

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

$S_{m_n}^{(1)}$ is a space of the collection of models whose dimension $D_{m_n}^{(1)}$ is to be precised. Thus we prove that under some standard condition and if p is great enough, (say $p \geq 8$), then $\tilde{b}$ defined by (28)-(29) satisfies

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

where C is a universal constant and R is a constant which depends on the problem, provided that $\hat{r}_n^2$ is computed on a space with dimension less than $n^{1/2-2/p}$. A is the compact set on which we wish to estimate b. b_m denotes the $\mathbb{L}_2(dx)$-orthogonal projection of b on S_m.
In the second step, for the contrast

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

we prove that if p is great enough (say $p \geq 16$), then the penalized estimator denoted by $\widetilde{\sigma^2}$ satisfies

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

where C and C' are universal constants, and R and R' are constants depending on the problem. The first step leads to minimax rates for $b \in \mathcal{B}_{\alpha,2,\infty}(A)$. Here, by using another control of $\mathbb{E}[\|b^2\mathbf{1}_A - \tilde{b}^2\|_n^2]$ we obtain minimax rates for σ^2 provided that b belongs to a Besov space $\mathcal{B}_{\alpha,2,\infty}(A)$ and σ^2 to $\mathcal{B}_{\beta,2,\infty}(A)$ with $\alpha > 2\beta + 3/4$. Note that Hall and Carroll (1989) obtained the same rate: $n^{-2\beta/(2\beta+1)}$ but under the lighter condition $\alpha > 1/2$. Nevertheless, they are not in an adaptive framework, and they consider regression with independent variables.

An important simulation study has moreover been lead in [13] in order to study the performance of the estimators, using piecewise polynomials of degrees up to 3, the degree being automatically chosen by the algorithm on each sub-interval of the subdivision. This allows to illustrate the good performances of the model selection procedure. The first experimentation of that kind has been done in a context of density estimation with histograms by Birgé et Rozenholc (2000).

4.2 Regression models with erros-in-variables: [11], [14]

At the begining, I was interested in the discrete time version of stochastic volatility models: their continuous time versions have been studied in [2], section 2.4 and in [5]. In the discrete time framework, it appears that those models are particular cases of regression models with error-in-variables:

$$\begin{aligned} X_t &= m(X_{t-1}) + \eta_t \quad (33) \\ Y_t &= X_t + \varepsilon_t \quad (34) \end{aligned}$$

with ε_t i.i.d. $(0, \sigma_\varepsilon^2)$, and η_t i.i.d. $(0, \sigma_\eta^2)$, both sequences being independent. The Y_t's only are observed and we aim at estimating the function m. In the same way as Fan and Truong (1993), we assume that m can be k times continuously differentiable on $[a, b]$, $a < b$ and we take a kernel K of order k (i.e. such that $\int K(y)dy = 1, \int y^j K(y)dy = 0, j = 1, \dots, k-1, \int y^k K(y)dy \neq 0$). We assume that $\Phi_\varepsilon(t) = \mathbb{E}(e^{it\varepsilon}) \neq 0, \ \forall t \in \mathbb{R}$. Then we consider the estimator:

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

where:

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

Φ_K denotes the fourier transform of K, $\Phi_\varepsilon(t) = \mathbb{E}(e^{it\varepsilon})$ and

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

We assume that the X_i's are stationary, with common distribution $d\mu(x) = f_X(x)dx$, where f_X is strictly bounded away from zero on $[a, b]$ et and admits k bounded derivatives. Moreover, the sequence $(X_i)_{i \in \mathbb{N}}$ is assumed to be absolutely regular with β-mixing coefficients β_k such that $\sum_{k \geq 0} (k+1)^c \beta_k < \infty$ pour un $c > 0$. Lastly, we consider in the same way as Fan and Truong (1993) or Masry (1993) two classes of errors ε : "Super Smooth errors of order α" denoted by SSE(α) whose characteristic function has exponential rate of decay $\exp(-|t|^\alpha)$ for t going to infinity, and "Ordinary Smooth errors of order α" denoted by OSE(α) with hyperbolic decay: $|t|^{-\alpha}$ when t goes to infinity.
For super smooth errors, Fan and Truong (1993) obtained the optimal rates for their estimator, namely rates of order $\ln(n)^{k/\alpha}$. Their results have been extended to regression models in an α-mixing framework by Masry (1993) without loss of rate. This is thus no surprise to keep the same rate in the autoregressive case. For ordinary smooth errors, we still keep the optimal rates obtained by Fan and Truong (1993) for an independent regressive model: $n^{-k/[(2k+\alpha)+1]}$. But this time, our result is better than Masry's (1993) who had a loss of $\ln(n)$ because he considered an $\mathbb{L}_\infty$-norm instead of an $\mathbb{L}_2$-integrated norm as we do. It appears so that it is wishable to work in an ordinary smooth error context. This is the reason why we study, in the stochastic volatility model defined by $h_t = \exp(X_t/2)\xi_t$ and (33) (which leads to the equation (34) by setting $Y_t = \ln(h_t^2) - \mathbb{E}(\ln(\xi_1^2))$ and $\varepsilon_t = \ln(\xi_t^2) - \mathbb{E}(\ln(\xi_1^2))$), which kind of distribution for ξ_t gives ordinary smooth ε_t's.

Another way of improving the rates, in the context of super smooth errors in particular (for instance in the case of gaussian errors), is to work in a framework where the function

m is somehow better specified. This is the reason why [14], written with M.L. Taupin, aims at estimating $m = m_\theta$ in a semiparametric context. Two kinds of result already exist in an anlogous framework: those of Taupin (1998) who considers independent variables and gaussian measurement errors, and those of Chanda (1995) who more specifically considers a dependent framework defined by an ARMA(p,q) process with errors-in-variables. Obviously, our work is an extension of Taupin's (1998) results.
If the distribution of the X_i's was known, we would study the following criterion:

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

where $W(.)$ is a compactly supported weight function with support denoted by S_W. Indeed the knowledge of f_X and f_ε allows to use the formula:

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

whre $f_Y = f_\varepsilon * f_X$ is the density of Y, $*$ denoting the convolution product. The criterion (36) is associated to the estimator:

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

We check in a preliminary step the consistency and asymptotic normality of this estimator under the conditions

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

In the general case, $\mathbb{E}(m_\theta(X_i)|Y_{i-1})$ can not be computed, and we replace it by an estimator deduced from the formula (37).

First, we care about the numerator. We denote by $g^\star$ the Fourier transform of g, $g(x) = \int g(u)e^{ixu}du$, and

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

Since

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

we set

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

where $Re(z)$ denotes the real part of z and K_n is defined by $K_n(x) = C_nK(C_nx)$. K is a kernel fulfilling some standard assumptions. For instance, the naive kernel $K(x) = \sin(x)/(\pi x)$ satisfies these assumptions and

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

as well.
For the denominator we standardly choose

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

wher $V_n(x) = h_n^{-1} V(x/h_n)$, $h_n^{-1} \to +\infty$ remains to be chosen and $V(x)$ is defined by (41). Thus the global estimate is $\widehat{\theta}$:

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

where

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

with $\widehat{\Gamma}_\theta$ defined by (40) and $\widehat{f}_Y$ defined by (42).

A precise study of the numerator and the denominator allows us to prove that $\widehat{\theta}$ is consistent for relevant choices of the bandwidths, for ordinary or super smooth errors, whether the mixing is geometrical or arithmetical $\beta_k \leq M k^{-\rho}$ with $\rho > 2$.

The computation of a bound for the rate is somehow more tedious. Nevertheless, we can give it in terms of some quantities that can be defined as follows. For a given function g, we denote by $\overline{\mathbf{I}}_{C_n}(t) = \mathbf{I}_{|t| \geq C_n}$ and $\mathbf{I}_{C_n}(t) = \mathbf{I}_{|t| \leq C_n}$ and we set:

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

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

Thus, under some technical (but standard) regularity assumptions for m_θ, and assuming the consistency conditions fulfilled, if we add some more constraints: $M_n^2(m_\theta) = o(n)$, $M_n^2(m_\theta^{(1)}) = o(n)$, and for geometrical mixing or arithmetical mixing with $\rho > 2$ and for super smooth errors or ordinary smooth errors with $\alpha > 11/5$, we find:

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

where

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

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

where $(Y_j^\star)_j$ is a sequence of independent random variables such that $Y_j^\star$ has the same distribution as Y_j, where $F_n(Y_2^\star)$ is a conditional mean that sometimes can be evaluated and in all other cases is such that $\mathbb{E}^{1/2}(D_n^2(Y_2^\star)) \leq \sigma_n(m_\theta)$.
In other words, we have the less precise (but more concise) result:

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

To summarize, our main result brings a method for computing a bound for the rate of the estimator, that can be used as soon as the regression function is assumed to belong to some more specified family. The quantities to compute in each case are clearly defined.

This result is illustrated throughout some examples $m_\theta(x) = \sum_{k=0}^p \theta_k x^k$, $\sum_{k=1}^p \theta_k^{(1)} e^{\theta_k^{(2)} x}$ for a regression and $e^{-\theta x^2}$ $0 < \theta < 1/e$ for an autoregression, $m_\theta(x) = \sum_{k=1}^p \theta_k \cos(2\pi k x)$. We find for super smooth errors much better rates than the powers of $\ln(n)$ found for the nonparametric kernel estimators.

APPENDICE
Examples of collections of models

We consider three kinds of collections of models $(S_m)_{m\in\mathcal{M}_n}$.

(T) Trigonometric polynomials : we consider the spaces S_m spanned by the functions $\varphi_0(x) = 1/\sqrt{2\pi}$, $\varphi_{2j}(x) = \sqrt{1/\pi}\cos(jx)$, $\varphi_{2j+1}(x) = \sqrt{1/\pi}\sin(jx)$ for $j = 0, \dots, d_m$, where $D_m = 2d_m + 1$ is the dimension of S_m. The collection $\mathcal{M}_n$ is in that case the set of all possible dimensions: $\mathcal{M}_n = \{1, \dots, n\}$. A space $\mathcal{S}_n$ in which all the others are contained is the space generated by all the φ_j, $2j + 1 \le n$.

(RP) Regular piecewise Polynomials: we consider the partition of $[0, 1]$, $\mathcal{I}_m = \{[j/m, (j+1)/m), j = 0, \dots m-1\}$ and a given integer $r \ge 0$. Then we define S_m as the space of piecewise polynomials of degree less or equal to r on the partition $\mathcal{I}_m$ and we have $D_m = (r + 1)m$. The space associated to the thinest partition is denoted by $\mathcal{S}_n$. We denote also by **(DP)** the spaces of piecewise polynomials of degree less or equal to r on the partitions built on diadic knots (i.e. with knots at points $j/2^m$).

(GP) General piecewise Polynomials (with degree $\le r$). These spaces are spanned by polynomials of degree at most r on any sub-interval of a grid whose set of knots is any subset of the given thinest grid: $\{\pi j/2^{J_n}, j = 0, 1, \dots, J_n\}, (r + 1)2^{J_n} \le n$ i.e. $J_n = [\ln(n/r)/\ln(2)]$ in the diadic case.

(W) Regular compactly supported wavelets. See Daubechies (1992), Chapter 9.3.

(GW) General compactly supported wavelets: for the description of those spaces we refer to Barron, Birgé and Massart (1999), Section 3.2.1.

REFERENCES

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. (1997) Model selection for regression on a fixed design. To appear in *Probab. Theory Relat. Fields.*

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.

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.

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.

Content

	General Content	1
1.	Introduction	2
2.	Causality	4
2.1	Continuous time causality	4
2.2	Second order causality	6
3.	Longue memory and stationarity	10
3.1	Continuous time long memory models	10
3.2	Log-periodogram regression	15
3.3	Adaptative estimation of the spectral density of a gaussian stationary proces	16
4.	Regression in a β-mixing context	19
4.1	Regression models	19
4.2	Regression models with errors-in-variables	22
	Appendix: Examples of collections of models	26
	References	27