

Journées de STATistique à Rennes 2008

Statistique, Econométrie



4 - 5 décembre 2008

ENSAI

Campus de Ker Lann, Rue Blaise Pascal - BP 37203, 35172 Bruz Cedex

Programme

Jeudi 4 décembre 2008

09h30-10h00. Accueil, Hall de l'ENSAI

10h00-12h00. Première session

M. A. Delgado. *Bootstrap assisted specification tests for the ARFIMA model.*
E. Guerre. *Adaptive rate-optimal detection of small autocorrelation functions.*
S. Darolles. *L-Performance with an Application to Hedge Funds.*

12h00-14h00. Pause déjeuner

14h00-17h00. Deuxième session

A. Philippe. *Random sampling of long-memory stationary processes.*
F. Lavancier. *A two-sample test for comparison of long memory parameters.*

15h20-15h40. Pause café / thé

M. Hoffmann. *Reconstruction de volatilité à travers les échelles.*
A. Mas. *Quelques résultats récents pour l'inférence locale sur données fonctionnelles.*

Vendredi 5 décembre 2008

09h30-12h30. Troisième session

S. Grégoir. *Testing in vector autoregressions with possibly seasonally and non-seasonally (co-)integrated processes.*
C. Francq. *Testing the nullity of GARCH coefficients : correction of the standard tests and relative efficiency comparisons.*

10h50-11h10. Pause café / thé

L. Horváth. *Topics in autoregressive time series with random coefficients.*
F. Le Gland. *Comportement asymptotique du filtre de Kalman d'ensemble pour des grands échantillons.*

12h30-14h00. Pause déjeuner

14h00-16h00. Quatrième session

E. Gautier. *Nonparametric Estimation in Random Coefficients Binary Choice Models.*
C. Villa. *Representative Yield Curve Shocks and Stress Testing.*
S. Srisumaz. *Semiparametric Estimation of Markov Decision Processes with Continuous State Space.*

16h00-16h15. Discussions de clôture

Résumés des exposés

Bootstrap assisted specification tests for the ARFIMA model

Miguel A. Delgado, Javier Hidalgo and Carlos Velasco.

This article proposes omnibus bootstrap assisted tests for time series model specification using test statistics based on Bartlett's T_p -process, which is a standardized estimator of the residual spectral distribution. We focus the attention on the ARFIMA model, but extensions to more general models are fairly straightforward. Unlike the U_p -process, which is the standardized spectral distribution of the raw data, the T_p -process is distribution free under the null hypothesis with known parameters, assuming that innovations satisfy certain martingale conditions. However, when parameters are estimated the limiting distribution of the T_p -process under the null is non-pivotal, depending on the unknown parameters of the specified model, as well as the specific estimation method employed. We propose to approximate the unknown distribution of the test statistics using bootstrap procedures under alternative resampling methods. We consider naïve residual resampling, periodogram resampling imposing the specification on the null and a wild resample of the first order asymptotic expansion of the T_p -process with estimated parameters. The bootstrap approximation applied to asymptotically pivotal transformations of the T_p -process, as proposed by Delgado, Hidalgo and Velasco (2005), results in improvements of the level accuracy with respect to tests based on the asymptotic approximation. We also consider bootstrap versions of the transformed process, which results in level accuracy gains of the corresponding tests. The validity of the alternative bootstrap tests is formally justified and we compare their finite sample performance by means of a Monte Carlo experiment.

Adaptive rate-optimal detection of small autocorrelation functions

Alain Guay, Emmanuel Guerre and Stepana Lazarova

A new test is proposed for the null of absence of serial correlation. The test uses a data-driven smoothing parameter. The resulting test statistic has a standard limit distribution under the null. The data-driven choice of the smoothing parameter is calibrated to achieve various rate optimality properties. The test can detect sparse autocorrelation functions which converge to the null with an optimal adaptive rate than can be faster than the parametric rate. The adaptive rate-optimality against smooth alternatives of the new test is also established. The test can also detect ARMA and Local Pitman alternatives which converges to the null with a rate close or equal to the parametric one. A simulation experiment and an application to monthly financial squared returns illustrate the usefulness of the proposed approach.

L-Performance with an Application to Hedge Funds

Serge Darolles, Christian Gourieroux, Joann Jasiak

This paper introduces a new fund performance measure, called the L-performance. It is proposed as an alternative to the Sharpe performance measure that is commonly used for fund performance valuation despite its inability to account for the skewness and thick tails of fund return distributions. The L-performance improves upon the Sharpe measure in this respect. Technically, the L-performance is based on sample statistics, called L-moments, which are conceptually close to the conventional power moments, but provide more detailed information about the extremes. For this reason, the L-moments

are used for prediction and assessment of extreme events, such as floods and earthquakes. In this paper, the new L-performance measure is calculated for a variety of hedge funds and is used to derive a fund ranking.

Random sampling of long-memory stationary processes

Anne Philippe and Marie-Claude Viano

The effects of random sampling on the second order characteristics and more specially the memory of stationary discrete time process are the subject of this talk.

We start from

- $\mathbf{X} = (X_n)_{n \geq 0}$, a stationary discrete time second order process with covariance sequence $\sigma_X(h)$
 - $(T_n)_{n \geq 0}$ a random walk independent of \mathbf{X}
- and we consider the sampled process \mathbf{Y} defined by

$$Y_n = X_{T_n}, \quad n \in \mathbb{N}$$

We present results on the L^p -convergence of $\sigma_Y(h)$ and the existence of the spectral density of the sampled process. We show in particular that short memory is always preserved as well as absolute continuity of the spectral measure.

The main results concern changes of memory by sampling processes with regularly varying covariances. We show that the intensity of memory of such processes is preserved if $\mathbb{E}(T_1) < \infty$, while this intensity decreases when $\mathbb{E}(T_1) = \infty$.

In the last part, we consider processes having long-memory with seasonal effects. We investigate the particular effects of deterministic sampling. We show that in some cases the seasonal effects can totally disappear after sampling.

A two-sample test for comparison of long memory parameters

Frédéric Lavancier, Anne Philippe and D. Surgailis

Given two stationary time series, we propose to test the equality of their long memory parameters (possibly null). These series may be correlated. The test statistic is based on ratios of the two V/S statistics of the series.

We prove that our test is consistent as soon as the partial sums of the two series converge jointly to a bivariate Gaussian process satisfying certain scaling and stationary increment properties. We prove that such a process is necessarily the general bivariate fractional Brownian motion, that we define and study.

From a practical point of view, the test procedure demands to fix a tuning parameter. We propose an adaptive choice of it, depending on the long memory part, but also on the short memory part of the series. We finally present some simulations. They show that this choice guaranties a good size of the test, even when the estimation of the long memory parameter is disturbed by a short memory part. Moreover, they allow to assess the power of the test under different alternatives.

Reconstruction de volatilité à travers les échelles

Marc Hoffmann

On suppose que l'on peut observer le prix d'un actif financier presque continûment (l'échelle de quelques secondes, pour fixer les idées) sur un intervalle de temps grand (de l'ordre de plusieurs mois). De telles hypothèses deviennent de plus en plus réalistes en pratique sur les marchés organisés électroniques. L'hypothèse fondamentale est que le (logarithme) du prix "ressemble" dans une gamme d'échelles à une semi-martingale d'Itô (éventuellement avec des sauts). On cherche à estimer la volatilité intégrée (normalisée)

$$\Lambda_t(\tau) := \tau^{-1} \int_{t-\tau}^t \sigma_s^2 ds, \quad t \geq 0$$

simultanément pour plusieurs échelles temporelles $\tau > 0$ lorsque l'on laisse t glisser au cours du temps, et où σ_s est la partie (éventuellement aléatoire) continue de la volatilité du prix. Cette question est posée par certaines techniques d'arbitrages statistiques ou de problèmes de trading sur variance en finance statistique.

Si l'échelle temporelle τ n'est pas petite, le problème est essentiellement de nature paramétrique, alors que dans les petites échelles de temps, il devient génériquement non-paramétrique, et dépend alors fortement de la structure sous-jacente (généralement inconnue) de la régularité du processus de volatilité. A ces échelles, les effets de bruit de microstructure doivent être incorporés. Nous développerons à cet effet une méthode rudimentaire adaptative basée sur du seuillage par ondelettes et reconstruirons $\Lambda_t(\tau)$ simultanément à toutes les échelles τ avec une forme d'optimalité que nous décrirons.

Quelques résultats récents pour l'inférence locale sur données fonctionnelles

André Mas

Les méthodes d'estimation locale en statistique sont bien connues dans le cas d'observations réelles ou vectorielles. Par estimation locale on entend ici estimation en un point fixé d'une densité ou d'une fonction de régression par exemple. On sous-entend aussi l'emploi de méthodes à noyau ou de type 'plus proches voisins'.

Ces techniques -quand on tente de les généraliser à l'inférence sur les courbes- viennent toutes buter sur l'absence d'un équivalent à la mesure de Lebesgue sur des espaces de dimension infinie. Les dérivées seconde et d'ordre supérieur sont aussi nettement plus délicates à manipuler et à estimer dans ce contexte.

Je présenterai un modèle de régression non linéaire sur données fonctionnelle de la forme :

$$y = r(X) + \varepsilon, \quad \mathbb{E}(\varepsilon|X) = 0$$

où X est un élément aléatoire à valeur dans un espace fonctionnel \mathcal{F} , y est réelle et r est une fonction inconnue de \mathcal{F} vers \mathbb{R} . On propose une méthode d'estimation de $r(x_0)$ où $x_0 \in \mathcal{F}$ est fixé par approximation linéaire locale de r . Nous montrerons que cette problématique de départ, simple mais encore peu étudiée dans le paradigme des données fonctionnelles nous amène vers des développements nombreux. Nous évoquerons ainsi les problèmes inverses, la théorie des petites boules ou celle des valeurs extrêmes, nous introduirons l'ACP fonctionnelle locale, tenterons de montrer en quoi la notion de distribution à contour elliptique peut apporter des solutions à notre problème initial.

Les travaux exposés sont publiés pour certains, soumis pour d'autres, mais une bonne partie reste encore en cours d'investigation.

**Testing in vector autoregressions with possibly seasonally
and non-seasonally (co-)integrated processes**

Stéphane Grégoir

To avoid possible distortions introduced by seasonal adjustment, econometricians may be interested in working with non seasonally adjusted data. This involves numerous seasonal and non seasonal unit root and cointegration rank tests that may imply pretest biases. Drawing on Toda and Yamamoto (1995) lag-augmented approach, we show how we can estimate VAR's formulated in levels and test for general restrictions on the matrix parameters when the processes may be integrated or cointegrated at various frequencies and of various orders. When introducing in a levels VAR for a given variable at least as many additional lags as the number of unit roots present in its individual data generating process, we get that the Wald test statistic associated to nonlinear constraints on the initial VAR parameters is asymptotically chi-squared distributed.

**Testing the nullity of GARCH coefficients : correction of the standard tests and relative
efficiency comparisons**

Christian Francq

This talk is concerned by testing the nullity of coefficients in GARCH models. The problem is non standard because the quasi-maximum likelihood estimator is subject to positivity constraints. The asymptotic null and local alternative distributions of Wald, score, and quasi-likelihood ratio tests are established. Efficiency comparisons under fixed alternatives are also considered. Two cases of special interest are : (i) tests of the null hypothesis of one coefficient equal to zero and (ii) tests of the null hypothesis of no conditional heteroscedasticity. The proposed approach is used in the analysis of a set of financial data and leads to reconsider the preeminence of GARCH(1,1) among GARCH models.

Topics in autoregressive time series with random coefficients

Lajos Horvath

Autoregressive processes with random coefficients (RCA processes) have long been studied in the nonlinear time series literature and have found applications in econometrics and biology, among others. In the first part of this talk, we explore the connection to ordinary AR processes and review results describing the structure of RCA processes such as conditions for stationary solutions and finiteness of moments. In a second part, we discuss statistical inference techniques to estimate the unknown parameters of an RCA process. Particular emphasis will be given to a unified quasi-likelihood estimation theory that, perhaps surprisingly, works in both the stationary and nonstationary case.

**Comportement asymptotique du filtre de Kalman d'ensemble
pour des grands échantillons**

François Le Gland, Vu-Duc Tran et Valérie Monbet

Le filtre de Kalman d'ensemble (EnKF) est largement utilisé en assimilation de données séquentielle, où des vecteurs d'état de très grande dimension (résultant par exemple de la discrétisation spatiale de champs de vitesse et de pression à l'échelle d'un continent, en météorologie) doivent être estimés récursivement à partir d'observations (recueillies par exemple à des stations réparties in-situ). Même

dans le cas simple où les équations d'état et d'observation sont linéaires avec des bruits blancs additifs gaussiens, calculer et conserver en mémoire les matrices de covariance d'erreur mises en jeu dans le filtre de Kalman est pratiquement impossible, et l'idée du filtre de Kalman d'ensemble consiste à représenter la distribution conditionnelle de filtrage à l'aide d'un échantillon (ensemble) constitué d'un petit nombre d'éléments et à voir la matrice de covariance empirique associée à cet échantillon comme une approximation de la matrice de covariance d'erreur. Des extensions au cas d'équations d'état non-linéaires ont également été proposées.

De manière surprenante, peu de résultats sont disponibles sur le comportement asymptotique du filtre de Kalman d'ensemble, c'est-à-dire sur la qualité numérique de cette approximation, alors que d'un autre côté le comportement asymptotique des filtres particuliers et de leurs nombreuses variantes est bien compris et documenté, quand le nombre de particules tend vers l'infini. En interprétant les éléments d'ensemble comme une population de particules en interaction de type champ moyen (et pas seulement comme un procédé instrumental proposant la valeur moyenne de l'ensemble comme estimateur de l'état caché), nous démontrons la convergence de l'EnKF, avec la vitesse classique $1/\sqrt{N}$, quand le nombre N d'éléments de l'ensemble tend vers l'infini. Dans le cas linéaire, on retrouve que la limite de la distribution empirique des éléments de l'ensemble coïncide avec la distribution gaussienne associée au filtre de Kalman usuel, mais dans le cas plus général d'une équation d'état non-linéaire et d'observations linéaires, cette limite diffère du filtre bayésien usuel, et nous en donnons une description. Pour obtenir la limite attendue dans ce cas, Nicolas Papadakis propose d'interpréter le mécanisme qui génère les éléments dans le filtre de Kalman d'ensemble comme une proposition de distribution d'importance, et d'affecter les poids d'importance correspondants aux différents éléments, comme dans un filtre particulier. Nous terminons par une comparaison du filtre de Kalman d'ensemble (EnKF) et du filtre de Kalman d'ensemble pondéré (WEnKF) avec le filtre particulier utilisant la distribution d'importance optimale.

Nonparametric Estimation in Random Coefficients Binary Choice Models

Eric Gautier and Yuichi Kitamura

We consider the estimation of the joint density of the random in a random coefficient model

$$Y = \mathbb{I}\{X'\beta \geq 0\} \quad (1)$$

where \mathbb{I} denotes the indicator function and X is a d -vector of covariates such that the first element is 1. The first element of β in this formulation absorbs the usual scalar stochastic shock term as well as a constant in standard binary choice with non-random coefficients. X and β correspond respectively to observed and unobserved heterogeneity across agents. This is an ill-posed inverse problem. The operator that relates the density of the random coefficient with the choice probability is called the hemispherical transform. It corresponds to a convolution on the $d - 1$ dimensional sphere. Fourier-Laplace series allow to have a clear insight on the identification problem and what useful restrictions allow to have identification. We propose a general procedure to estimate the density of the random coefficient. Characterizing the degree of ill-posedness in the Sobolev spaces based on L^2 we are able to relate the rate of convergence of an estimate of the density with that of an estimate of the choice probability. Finally, given a particular estimate of the choice probability relying on a plug-in of an estimate of the density of the design and smoothed orthogonal projection kernels, the estimate of the density of the random coefficient takes a simple closed form which is easy to implement in empirical applications and for which we obtain rates of consistency in all L^p spaces for $p \geq 1$ and prove asymptotic normality. If time will permit, extensions including estimation of marginals, treatment of nonrandom coefficients and models with endogeneity could be presented.

Representative Yield Curve Shocks and Stress TestingFrancis X. Diebold, Canlin Li, Christophe Pérignon, and Christophe Villa

In this paper we propose a systematic procedure to identify a set of representative yield curve shocks and use them for stress-testing purposes. We first fit a factor model to actual bond yields and estimate the main shape factors of the yield curve. We then partition the factors into non-overlapping sets of representative shocks. The key feature of our procedure is that it provides a wide variety of yield curve shocks including typical, uncommon, and extreme ones. We apply our methodology to a variety of bond strategies using actual U.S. yields.

Semiparametric Estimation of Markov Decision Processes with Continuous State SpaceOliver Linton and Sorawoot Srisumaz

We develop a semiparametric method to estimate the structural parameters and their functionals that arise in the study of a class of discrete Markov decision processes. Unlike most existing work in the literature, our method allows for continuous components in the observable state space. The insight is to view the smoothed Bellman Equation as a linear integral equation of type II, the solving of which is a well-posed problem. We also show that our method can be used to estimate a popular class of Markovian games with little modification.

Conférenciers

Darolles Serge, CREST, Paris,
Delgado Miguel A., University Carlos III, Madrid,
Francq Christian, Université Lille 3,
Gautier Eric, ENSAE, Malakoff,
Grégoir Stéphane, EDHEC Paris,
Guerre Emmanuel, Department of Economics, Queen Mary, University of London,
Hoffmann Marc, ENSAE, Malakoff,
Horvath Lajos, University of Utah, Department of Mathematics,
Lavancier Frédéric, Université de Nantes,
Le Gland François, INRIA Rennes – Bretagne Atlantique / IRISA,
Mas André, Université Montpellier 2,
Philippe Anne, Université de Nantes,
Srisumaz Sorawoot, London School of Economics,
Villa Christophe, Audencia, Nantes.

Participants

Azzaoui Nourddine, Université de Technologie de Troyes.
Basseville Michèle, CNRS / IRISA.
Beninel Farid, ENSAI.
Cadoret Isabelle, CREM.
Chauvet Guillaume, ENSAI.
Chen Jiaqi, Université Rennes 1.
Clauss Pierre, ENSAI.
Coquet François, ENSAI.
Dacunha-Castelle Didier, Université Paris Sud 11.
Delyon Bernard, Université Rennes 1.
Deshayes Jean, Université Rennes 1.
Dong Qian, ENS Cachan Antenne de Bretagne.
Gelein Brigitte, ENSAI.
Golhen Yann, MAIF - Direction Marketing.
Guyader Arnaud, IRISA / Université Rennes 2.
Hoang Thi Thu Huong, Université Paris Sud 11.
Hristache Marian, ENSAI.
Huchet-Bourdon Marilyne, Agrocampus Ouest.
Jegou Nicolas, Université Rennes 2.
Josse Julie, Agrocampus Ouest.
Kachour Maher, Université Rennes 1.
Kouakou Kouadio Simplicie, Université Rennes 2.
Larribeau Sophie, CREM - Université Rennes 1.
Le Meur Jean-Baptiste, Université Rennes 2.

Leipus Remis, Université de Vilnius.
Marchand Jean-Louis, Université Rennes 1.
Mom Alain, Université Rennes 2.
Morin Annie, IRISA / Université Rennes 1.
Paroux Katy, INRIA Rennes / Université de Franche-Comté.
Pentecôte Jean-Sébastien, CREM - Université Rennes 1.
Regnault Philippe, Université de Caen.
Rouvière Laurent, Université Rennes 2.
Tanvez Charlotte, Université de Nantes.
Vimond Myriam, ENSAI.
Wuthrich Guillaume, Université Paris Sud 11.
Yao Jian-Feng, Université Rennes 1.

Organisateurs

Cadre Benoît, ENS Cachan Antenne de Bretagne,
Dehay Dominique, Université Rennes 2,
Fromont Magalie, ENSAI / Université Rennes 2,
Lavergne Pascal, ENSAI / SFU,
Patilea Valentin, INSA Rennes / ENSAI,
Pumo Besnick, Agrocampus Ouest,
Raïssi Hamdi, INSA Rennes,

Avec l'aide précieuse des services de la Communication, de la Gestion Administrative et de la Logistique de l'ENSAI...