

Séminaire de statistique ENSAI

Année 2010-2011

Responsable Laurent Rouvière

Vendredi 13 mai 2011

Hamdi Raissi (Insa, Rennes) : *Adaptive Specification of the Dynamics of Vector Autoregressive Processes with Time Varying Variance.*

Résumé : Linear Vector AutoRegressive (VAR) models where the innovations could be unconditionally heteroscedastic are considered. The volatility structure is deterministic but time-varying and allows for changes that are commonly observed in economic or financial multivariate series such as breaks or smooth transitions. In this framework we propose Ordinary Least Squares (OLS), Generalized Least Squares (GLS) and Adaptive Least Squares (ALS) procedures for the statistical inference of VAR models. The GLS approach requires the knowledge of the time-varying variance structure while in the ALS approach the unknown variance is estimated by kernel smoothing with the outer product of the OLS residuals vectors. The estimation of the VAR models is investigated. It is shown that the standard asymptotic results for the OLS estimators can be quite misleading in our framework. Therefore we derive the asymptotic distribution of the estimators for the VAR model coefficients obtained using the OLS, GLS and ALS methods and compare their properties. Using these results the problem of goodness-of-fit of VAR processes with unconditionally heteroscedastic errors is also studied. The unreliability of the standard portmanteau tests in our framework is highlighted. The correct critical values of the standard portmanteau tests based on the OLS residuals are derived.

Moreover, modified portmanteau statistics based on residual autocorrelations obtained from the GLS and ALS estimation of the VAR coefficients are introduced and their asymptotic critical values are obtained. It is shown through theoretical results and numerical examples that the more elaborated ALS approach achieves a substantial gain of efficiency for the statistical analysis of VAR models when compared to the OLS approach in presence of heteroscedasticity. In particular we establish the asymptotic equivalence between the ALS and GLS methods. The implementation of the studied methods is illustrated using U.S. economic data sets.

Vendredi 1 avril 2011

Philippe Saint-Pierre, (Université Paris 6) : *Régression pour des données bivariées censurées à droite.*

Résumé : on étudie une classe de M-estimateur dans un contexte de régression où la variable à expliquer et la variable explicative sont toutes les deux censurées à droite. Notre approche consiste à introduire des pondérations dans les estimateurs du cas non censuré afin de compenser la censure.

Les estimateurs traditionnels de la fonction de survie bivariée n'étant pas adaptés dans ce contexte, nous introduisons un nouvel estimateur de la distribution jointe des temps de survie. Une approche par les copules est utilisée pour modéliser la dépendance entre les temps de censure. On obtient une représentation i.i.d. de l'estimateur de la distribution des temps de survie qui permet de déduire les propriétés asymptotiques de l'estimateur de la régression. Des simulations sont proposées pour évaluer le comportement de cette méthode pour des petits échantillons.

Vendredi 18 février 2011

[Anne Ruiz-Gazen](#), (Université Toulouse 1) : *EPP-Lab : une nouvelle interface pour l'exploration de données par projections révélatrices.*

Résumé : Les méthodes d'exploration de données par projections révélatrices ont été développées dans les années 1980 - 1990. Ces méthodes permettent d'obtenir des projections de données multivariées dans des sous-espaces de petites dimensions susceptibles de révéler des structures intéressantes au sein des données. L'intérêt d'une projection est mesuré par un indice qui doit être optimisé. De nombreux indices ainsi que différentes méthodes et stratégies d'optimisation ont été proposées et étudiées. Malgré cela, les méthodes de projections révélatrices ne sont pas ou peu utilisées par les statisticiens car peu présentes dans les principaux logiciels statistiques. Dans cet exposé, nous faisons un tour d'horizon des principales méthodes de projections révélatrices exploratoires et proposons de nouveaux indices, algorithmes et stratégies d'optimisation avec une mise en oeuvre par le biais d'une interface en JAVA.

Vendredi 14 janvier 2011

[Magalie Fromont](#) (ENSAI), *Tests adaptatifs d'homogénéité pour un processus de Poisson*

Résumé : Après un rappel de quelques notions sur les processus de Poisson et les tests statistiques non paramétriques, nous nous intéressons au problème de test d'homogénéité d'un processus de Poisson observé sur un intervalle borné. Nous établissons dans ce cadre des bornes inférieures pour les vitesses de séparation uniformes relativement à la norme L2 sur des Besov bodies classiques ou faibles. De façon surprenante, nous obtenons sur les Besov bodies faibles des bornes inférieures qui peuvent coïncider avec les vitesses minimax d'estimation. Nous construisons ensuite des procédures de tests non asymptotiques et non paramétriques qui sont adaptatives, au sens où elles atteignent, à un facteur logarithmique près dans certains cas, les vitesses de séparation optimales sur plusieurs classes d'alternatives simultanément. Ces procédures sont basées sur des méthodes de sélection de modèles et/ou de seuillage. Enfin, nous complétons cette étude théorique par des simulations afin d'estimer par la méthode de Monte Carlo la puissance de nos tests sous diverses alternatives.

Vendredi 3 décembre 2010

[Andrea Foulkes](#), (University of Massachusetts), *Mixture modeling for discovering genotype-trait associations.*

Résumé : Characterizing the underpinnings of complex diseases will inevitably require

consideration of multiple single nucleotide polymorphisms (SNPs) within and across genes, presenting several analytic challenges due to the large number of potentially informative SNPs and the largely uncharacterized relationships among them. We propose a mixture modeling framework for identifying and characterizing associations between multiple SNPs and a measured trait. Important advances offered by this paradigm include that the mixture model: (1) addresses the degrees-of-freedom challenge inherent in application of a fixed effects analysis of covariance; (2) relaxes the restrictive single normal assumption of a classical mixed effects model; and (3) offers an exploratory framework for discovery of underlying structure across multiple genetic loci. An application to data arising from a study of anti-retroviral associated dyslipidemia in HIV-infected individuals is presented.

Vendredi 5 novembre 2010

[Maud Delattre](#) (Université Paris 11), *Inférence statistique dans les modèles de Markov cachés à effets mixtes.*

Résumé : Les modèles de Markov cachés à effets mixtes étendent les modèles de Markov cachés aux études de population. L'estimation des paramètres dans ces modèles est une problématique complexe. Du fait de leur structure hautement non linéaire et de la présence de données non observées, la vraisemblance des modèles de Markov cachés à effets mixtes n'a pas d'expression analytique simple, et l'estimation par maximum de vraisemblance n'est pas directe. Dans cet exposé, une méthodologie complète est proposée pour l'estimation des paramètres dans les modèles de Markov cachés à effets mixtes. En particulier, un algorithme EM stochastique est mis en œuvre pour estimer les paramètres de population. Les propriétés des estimateurs obtenus sont étudiées sur la base de simulations Monte-Carlo. La méthodologie développée sera ensuite illustrée sur des données concrètes.

Vendredi 24 septembre 2010

[Valérie Monbet](#), (Université Rennes 1), *Space-time models for moving fields with an application to significant wave height fields,*

Résumé : The surface of the ocean, and so such quantities as the significant wave height, H_s , can be thought of as a random surface that develops over time. In this paper, we explore certain types of random fields in space and time, with and without dynamics that may or may not be driven by a physical law, as models for the significant wave height. Reanalysis data is used to estimate the sea-state motion which is modeled as a hidden Markov chain in a state space framework by means of an AR(1) process or in the presence of the dispersion relation. Parametric covariance models with and without dynamics are fitted to reanalysis and satellite data and compared to the empirical covariance functions. The derived models have been validated against satellite and buoy data.