Université catholique de Louvain – Institut de statistique, biostatistique et sciences actuarielles

Young Researchers' Day

February 4th, 2011

9^{00}	**Majda Talamakrouni**	Guided censored regression
9^{30}	**Rachida El Mehdi**	Stochastic Frontier Analysis With Copulas
10^{00}	**Rudolf Schenk**	Adaptive local functional regression

Coffee Break

11^{00}	**Mathieu Pigeon**	Individual stochastic loss reserving: model and preliminary results
11^{30}	**Mohammed Rida Soumali**	Detecting influential data in partially linear models
12^{00}	**Catherine Timmermans**	Bases Giving Distances. A new semimetric and its use for nonparametric functional data analysis

The seminar is followed by the annual lunch of the ISBA.

The Young Researchers' Day is held in room c.115 of the Institut de statistique, biostatistique et sciences actuarielles, Voie du Roman Pays 20, Louvain-la-Neuve.

Guided censored regression

MAJDA TALAMAKROUNI (Majda.Talamakrouni@uclouvain.be)

Parametrically guided nonparametric estimation is one of the most promising approaches that improves the bias of traditional nonparametric regression estimators without an increase in the variance. In the context of completely observed i.i.d. data, many techniques are available in the literature. These include Glad (1998), Fan and Ullah (1999), Gozalo and Linton (2000), Mays et al. (2001) and Martins-Filho et al. (2008).

However, a common problem in practice is the presence of censoring. We are studying the guided nonparametric estimator of the regression function when the dependent variable is subject to censoring. To deal with censoring we will simply transform the observed data in unbiased way, a technique that is largely used in the literature, see for example Delecroix et al. (2008). We will first study the case when the dependent and the censoring variable are independent and then we will move on to the more difficult and realistic case : the two variables are only conditionally independent given a covariate. We will study the properties of the new approach like the asymptotic normality, the efficiency and the robustness. The asymptotic results will also be illustrated with finite sample simulations.

References

Glad, I. K. (1998), Parametrically Guided Nonparametric Regression, *The Scandinavian Journal of Statistics. Theory and Applications, 25, 649-668.*

Fan, Y., and Ullah, A. (1999). Asymptotic Normality of a Combined Regression Estimator. *Journal of Multivariate Analysis, 71, 191-240.*

Gozalo, P., and Linton, O. (2000). Local Nonlinear Least Squares: Using Parametric Information in Nonparametric Regression. *Journal of Econometrics, 99, 63-106.*

Mays, J. E., Birch, J. B., and Starnes, B. A. (2001). Model Robust Regression: Combining Parametric, Nonparametric, and Semiparametric Methods. *Journal of Nonparametric Statistics, 13, 245-277.*

Martins-Filho, C., Mishra, S. and Ullah, A. (2008). A class of improved parametrically guided nonparametric regression estimators. *Econometric Reviews, 27, 542-573.*

Delecroix, M., Lopez, O. and Patilea, V. (2008). Nonlinear censored regression using synthetic data. *Scandinavian Journal of Statistics, 35, 248-265.*

Stochastic Frontier Analysis With Copulas

Rachida El Mehdi (Rachida.Elmehdi@student.uclouvain.be)

The basic idea of the efficiency analysis is the comparison between the Decision Making Units (firms, for example) in order to know how the inputs are used to produce outputs. In this study, Parametric Stochastic Frontier Analysis (SFA) is adopted to estimate Technical Efficiency (TE_i) of Moroccan municipalities using the production frontier. So, the error term is divided in to two dependent components: the normal noise v and the inefficiency u, then $\varepsilon = v - u$. This dependence is expressed by a copula function which is included in the joint density of (u, v) to determine the likelihood function.

The Corrected Ordinary Least Square (COLS) and Maximum Likelihood (ML) methods are used to estimate the model $y_i = f(x_i, \beta) + \varepsilon_i$, $i = 1, \cdots, n$. So, several models are considered with the Cobb-Douglas function, the Normal distribution for v, the Half or the Truncated Normal distribution for u and using several copulas. The expression of the log-likelihood function being complex, the use of numerical optimization was necessary. Estimation has revealed that models with copulas are better than those of the independence case and generally Technical Efficiency (TE) scores $TE_{dep.} < TE_{indep.}$ but the rank is the same.

Moreover, to make statistical inference, the parametric bootstrap procedure described by the algorithm#3 in Simar and Wilson (2010) is adapted to the copula case and adopted to estimate the efficiency percentile confidence interval.

Adaptive local functional regression

Rudolf Schenk (Rudolf.Schenk@uclouvain.be)

We consider the estimation of the value of a linear functional of the slope parameter in functional linear regression, where scalar responses are modeled in dependence of random functions. The theory in this presentation covers in particular point-wise estimation as well as the estimation of averages of the slope parameter. The proposed plug-in estimator is based on dimension reduction and additional thresholding. It is shown in Johannes and Schenk (2010) that this estimator can attain the minimax optimal rate of convergence up to a constant. However, this requires an optimal choice of the dimension parameter depending on certain characteristics of the unknown slope function and the covariance operator of the regressor, which are not known in practice. The main issue addressed in this talk is a fully data-driven choice of the dimension parameter using a model selection approach. We develop a penalized minimum contrast estimator with randomized penalty and collection of models and show that the adaptive estimator can attain the lower minimax risk bound up to a logarithmic factor over a wide range of classes of slope functions and covariance operators.

References

Johannes, J., and Schenk, R. (2010). *On rate optimal local estimation in functional linear model.* (submitted). Université catholique de Louvain. Available from http://arxiv.org/abs/0902.0645v2

Johannes, J., and Schenk, R. (2011). *Adaptive local functional regression.* (in preparation). Université catholique de Louvain.

Individual stochastic loss reserving: model and preliminary results

MATHIEU PIGEON (Mathieu.Pigeon@uclouvain.be)

Loss reserving is one of most difficult and exciting tasks facing actuaries. A loss reserve is a provision for an insurer's liability for claims and loss reserving is the term used to describe the actuarial process of estimating the amount of an insurance company's liabilities for loss. Naturally, the financial condition of an insurance company can not be assessed without valid loss reserve estimates. Numerous approaches have been developed to give reasonable estimates, but most of these models are based on grouped information (loss development triangles). One of the most used model in practice is the chain ladder model (Mack's model). However, aggregated structures are only a convenient summary of a more complete data set and recently, some individual models using information with regard to the actual claims processes have been developed.

The intent of this presentation is to introduce a new approach of loss reserving in the framework of the chain ladder model but based on individual information. The general structure of the model and distributional assumptions will be presented. A numerical example will be performed and preliminary results will be exposed.

Detecting influential data in partially linear models

MOHAMMED RIDA SOUMALI (Mohammed.Soumali@uclouvain.be)

Suppose that $(X_1, Y_1, T_1)', \ldots, (X_n, Y_n, T_n)'$ is a sequence of independent random vectors, identically distributed as a $d + 2$-dimensional random vector $Z = (X, Y, T)' \sim F$.We consider partially linear models of the form $Y = \beta_0 X + g_0(T) + \varepsilon$ with $\beta_0 \in \Re$ an unknown parameter, g_0 an unknown function and $E(\varepsilon|X, T) = 0$.

Let $\phi_0(T)$ and $\phi(T)$ denote, respectively, $E(Y|T)$ and $E(X|T)$.The parameter of interest β_0 can be estimated by:

$$\widehat{\beta}_{LS} = \text{argmin}_{\beta\in\Re} n^{-1} \sum_{i=1}^{n} \left(Y_i - \widehat{\phi}_{0,LS}(T_i) - \beta\left(X_i - \widehat{\phi}_{LS}(T_i)\right)\right)^2$$

where $\widehat{\phi}_{0,LS}(t)$ and $\widehat{\phi}_{LS}(t)$ are, respectively, Nadaraya−Watson estimarors of $\phi_0(t)$ and $\phi(t)$.We are interested in the effect that may have a slight departure from the F on our estimator(infinitesimal contamination),we provide in this paper, using a Robust bootstrap approach, the asymptotic distribution of a robust version of the empirical influence function of $\widehat{\beta}_{LS}$, a cutoff is then applied to identify highly influential points.

Bases Giving Distances. A new semimetric and its use for nonparametric functional data analysis

CATHERINE TIMMERMANS (Catherine.Timmermans@uclouvain.be)

This communication aims firstly at highlighting a new semimetric for measuring dissimilarities between regularly discretized curves, typically time series or spectra. Its main originality is that it is based upon the expansion of each series of a dataset into a different wavelet basis, one that is particularly suited for its description. Measuring dissimilarities in such a way implies comparing not only the projections of the series onto the bases, as usual, but also the bases themselves. As a consequence of this feature, our semimetric has the ability to capture the variations of patterns occurring in series along both the vertical and the horizontal axis. This property makes the semimetric particularly powerful when dealing with curves with sharp local features that might be affected simultaneously by horizontal shifts and vertical amplification.

Secondly, this communication aims at illustrating how we can advantageously make use of our semimetric in the framework of nonparametric functional data analysis (as developped in *Ferraty and Vieu* (2006)) when the curves we are dealing with are characterized by some horizontally- and vertically-varying sharp patterns. Simulated examples are shown as well as a real data example.