Semiparametric and Nonparametric Methods in Econometrics

Joel Horowitz
Northwestern University

Oliver Linton
London School of Economics

Enno Mammen
University of Mannheim

April 05 - April 10

Contents

1 Introduction by The Organizers

The main objective of this workshop was to bring together mathematical statisticians and econometricians who work in the field of nonparametric and semiparametric statistical methods. Nonparametric and semiparametric methods are active fields of research in econometric theory and are becoming increasingly important in applied econometrics. This is because the flexibility of non- and semiparametric modelling provides important new ways to investigate problems in substantive economics. Many of the most important developments in semi- and nonparametric statistical theory now take place in econometrics. Moreover, the development of non- and semiparametric methods that are suitable to the needs of economics presents a variety of mathematical challenges. Econometric research aims at achieving an understanding of the economic processes that generate observed data. This is different from fitting data that may be useful for prediction but that do not capture underlying causes. A large part of economic theory consists of models of equilibria of competing processes. Statistical data are a snapshot of the equilibrium but, by themselves, do not reveal the processes that led to the equilibrium. Consequently a reduced form model (e.g.\ a conditional mean function) does not suffice for much economic research. Achieving an understanding of the economic processes requires a careful combining of economic theory and statistical considerations. This often requires the development of statistical tools that are specific to the problems that arise in economics and are unfamiliar in other statistical specialties. For example, econometric research has focused on developing methods to deal with endogenous covariates (that is, covariates that are correlated with a model´s error terms), time series models that fit equilibria as stationary submodels (cointegration), and time series models for volatility processes (conditional variances) in finance. Semi- and nonparametric methods are being used increasingly frequently in applied econometrics. The models are not necessarily of the simple form of classical regression, "response = signal plus independent noise," where the signal can be recovered by nonparametric smoothing of the responses. Rather, the nonparametric functions enter the model in a much more complicated way. Mathematically this has led to challenging problems. Identifiability of a model is much more involved in nonparametric model specifications. In particular, this is the case for nonseparable models where the error terms do not enter additively into the model. Some nonparametric inference problems with endogenous covariates lead to statistical inverse problems and require the study of estimates and solutions of noisy integral equations. The mathematical analysis of nonparametric time-series models and of nonparametric diffusion models is strongly related to research in stochastic processes, Markov processes, stochastic analysis and financial mathematics. Empirical process theory is an essential tool for the understanding of uniform performance and of convergence rates of nonparametric estimates and for efficiency considerations in semiparametric models. All these problems were topics of talks and discussions at the workshop. The mathematical development in econometrics is complimentary to recent statistical applications in biology. There, the focus tends to be on dimension reduction for the statistical analysis of high-dimensional data. The intellectual charm of mathematical research in modern econometrics comes from the interplay between statistical and economic theory.

2 Abstracts

2.1 Identifying the Returns to Lying When the Truth is Unobserved, Arthur Lewbel, Boston College

Consider an outcome Y, an observed binary regressor D, and an unobserved binary D*. This paper considers nonparametric identification and estimation of the effect of D on Y, conditioning on D*. Suppose Y is wages, unobserved D* indicates college experience, and D indicates claiming to have been to college. This paper identifies the 'returns to lying' difference in wages, about 6% to 20%, between those who falsely claim college versus those who tell the truth about not having college.

Identification is obtained either by observing a variable V roughly analogous to an instrument, or by imposing restrictions on model error moments.

2.2 Testing Conditional Factor Models, Dennis Kristensen, Columbia University

We develop a new methodology for estimating time-varying factor loadings and conditional alphas based on nonparametric techniques. We test whether long-run alphas, or averages of conditional alphas over the sample, are equal to zero and derive test statistics for the constancy of factor loadings. The tests can be performed for a single asset or jointly across portfolios. The traditional Gibbons, Ross and Shanken (1989) test arises as a special case when there is no time variation in the factor loadings. As applications of the methodology, we estimate conditional CAPM and Fama and French (1993) models. We reject the null that long-run alphas on book-to-market and momentum decile portfolios are equal to zero even though there is substantial variation in the conditional factor loadings of these portfolios.

2.3 Semiparametric modeling and estimation of the dispersion function in regression, Ingrid Van Keilegom, Universite catholique de Louvain

Modeling heteroscedasticity in semiparametric regression can improve the efficiency of the estimator of the parametric component in the regression function, and is important for inference problems such as plug-in bandwidth selection and the construction of confidence intervals. However, the literature on exploring heteroscedasticity in a semiparametric setting is rather limited. Existing work is mostly restricted to the partially linear mean regression model with a fully nonparametric variance structure. The nonparametric modeling of heteroscedasticity is hampered by the curse of dimensionality in practice. Moreover, the approaches used in existing work need to assume smooth objective functions, therefore exclude the emerging important class of semiparametric quantile regression models. To overcome these drawbacks, we propose a general semiparametric location-dispersion regression framework, which enriches the currently available semiparametric regression models. With our general framework, we do not need to impose a special semiparametric form for the location or dispersion function. Rather, we provide easy to check sufficient conditions such that the asymptotic normality theory we establish is valid for many commonly used semiparametric structures, for instance, the partially linear structure and single-index structure. Our theory permits non-smooth location or dispersion functions, thus allows for semiparametric quantile heteroscedastic regression. We demonstrate the proposed method via simulations and the analysis of a real data set. (This is joint work with Lan Wang).

2.4 Asymptotic Theory for Nonparametric and Semiparametric Estimation with Spatial Data, Peter Robinson, London School of Economics

We develop conditions for asymptotic statistical theory for estimates of nonparametric and semiparametric models when the data are spatial, or spatio-temporal. We attempt to cover data that are regularly or irregularly-spaced, as well as ones where only pairwise distances are available, and cross-sectional data in which even this information is lacking but dependence is feared. The stress is on allowing for a broad range of spatial dependence, including long-range dependence, and heterogeneity, including conditional and unconditional heteroscedasticitya

2.5 On the Regularization Power of the Prior Distribution in Linear ill-Posed Inverse Problems, Anna Simoni, Toulouse School of Economics

We consider a functional equation of type ^Y = Kx + U in an Hilbert space. We wish to recover the functional parameter of interest x after observing ^Y . This problem is ill-posed because the operator K is assumed to be compact. We consider a class of models where the prior distribution on x is able to correct the ill-posedness even for an infinite dimensional problem. The prior distribution must be of the g-prior type and depends on the regularization parameter and on the degree of penalization. We prove that, under some conditions, the posterior distribution is consistent in the sampling sense. In particular, the prior-to-posterior transformation can be interpreted as a Tikhonov regularization in the Hilbert scale induced by the prior covariance operator. Finally, the regularization parameter may be treated as an hyperparameter and may be estimated using its posterior distribution or integrated out.

2.6 Efficient Estimation in ICA and ICA-Like Models, P.J. Bickel, UC Berkeley

The ICA (Independent Component Analysis) generalization of the multivariable Gaussian model corresponds to observing n iid observations of the form:

$$X = AZ,$$

where X, Z are p=1 and the components of Z are independent. It is well known that if Z has at most one Gaussian component A is identifiable up to permutation and scaling of the rows. Chen and Bickel (2006) Ann.Statist. constructed efficient estimates of A, ie ones which achieved the information bound if the components of Z had distributions known up to scale.

We'll discuss estimation in a generalization and a related model. The generalization is to the case where:

$$X = (X_1, \ldots, X_n) \qquad\qquad X_i = AZ_i$$

the X_i are p dimensional, the Z have dimension q<p, p is unknown, the Zi components are independent but differ in i through scale only and the rows of A are sparse. A second model discussed is due to Blanchard et al (2006), JMLR. We indicate its identity to the "sliced inverse regression" model of D.Cook(2007)Statistical Science and discuss efficient estimation in that model as well.

2.7 Inference Based on Conditional Moment Inequalities, Donald W. K. Andrews, Yale University

In this paper, we propose an instrumental variable approach to construct confidence sets for the true parameter in models defined by conditional moment inequalities/ equalities.

We show that by properly choosing instrument functions, one can transform conditional moment inequalities/equalities into unconditional ones without losing identification power. Based on the unconditional moment inequalities/equalities, we construct confidence sets by inverting Cramér-von Mises-type tests. Critical values are chosen using generalized moment selection (GMS), plug-in asymptotic, and subsampling procedures. We show that the proposed confidence sets have correct uniform asymptotic coverage probabilities. New methods are required to establish these results because an infinite-dimensional nuisance parameter affects the asymptotic distributions. We show that the tests considered are consistent against all fixed alternatives and have power against a broad array of n-1/2-local alternatives, though not all such alternatives. We extend the results to allow for an infinite number of conditional or unconditional moment inequalities/equalities.

2.8 Identification and Estimation of Marginal Effects in Nonlinear Panel Models, Whitney Newey, MIT

This paper gives identification and estimation results for marginal effects in nonlinear panel models. We find that linear fixed effects estimators are not consistent, due in part to marginal effects not being identified. We derive bounds for marginal effects and show that they can tighten rapidly as the number of time series observations grows. We also show in numerical calculations that the bounds may be very tight for small numbers of observations, suggesting they may be useful in practice. We propose two novel inference methods for parameters defined as solutions to linear and nonlinear programs such as marginal effects in multinomial choice models. We show that these methods produce uniformly valid confidence regions in large samples. We give an empirical illustration.

2.9 Gaussian process priors in nonparametric and semiparametric estimation, Aad van der Vaart VU University Amsterdam

We discuss the use of Gaussian processes as priors for an unknown function in a Bayesian analysis. We study the posterior distribution of the function or a parameter in the "frequentist" set-up, where the data are generated according to a fixed "true distribution". We explain how the rate of contraction of the posterior to the true parameter depends on the properties of the Gaussian process (reproducing kernel Hilbert space and small ball probability), how (random) scaling influences this rate, and how this changes if we are interested in a (semiparametric) functional. The first part of the talk is based on joint work with Harry van Zanten; the second part is mainly based on work of Ismael Castillo, both from VU University Amsterdam.

2.10 Nonparametric partial-frontier estimation: robustness and efficiency, Irene Gijbels, Katholieke Universiteit Leuven, Belgium,

One of the major aims in recent nonparametric frontier modeling is to estimate a partial frontier well inside the sample of production units but near the optimal boundary. Two concepts of partial boundaries of the production set have been proposed: an expected maximum output frontier of order m = 1, 2, . . . and a conditional quantile-type frontier of order 2]0, 1]. In this talk, we answer the important question of how the two families of partial production frontiers are linked. For each order m, we specify the order for which both partial boundaries can be compared. A discussion on the breakdown as well as on the efficiency of the nonparametric order-m frontiers and the order- frontiers is provided. Some asymptotic results are discussed. The theoretical findings are illustrated through some simulations and data analysis. This talk is based on joint work with Abdelaati Daouia.

2.11 Estimation of nonparametric models with simultaneity, Rosa L. Matzkin, University of California, Los Angeles

We introduce new estimators for nonparametric, nonadditive models with simultaneity. The estimators are computed as simple functionals of nonparametric estimators of the distribution of the observable variables, using constructive methods for identification. They are shown to be consistent and asymptotically normal. It is shown that when each structural equation possesses an exclusive observable exogenous variable, then under some restrictions on the derivatives with respect to those variables, one can estimate all the remaining derivatives by matrix inversion and multiplication, analogous to Least Squares. The paper analyzes in detail the identification and estimation of a single equation model when simultaneity is present and an instrument is available.

2.12 Identification in Accelerated Failure Time Competing Risks Models, Sokbae Lee University College London

We provide new conditions for identification in accelerated failure time competing risks models. In our model, we specify unknown regression functions and the joint survivor function of latent disturbance terms nonparametrically. We show that the model can be identified with covariates that are independent of latent errors, provided that certain rank conditions are satisfied. We present a simple example in which our rank conditions for identification are verified. Our identification strategy does not have the problem of identification at near zero".

2.13 On Robust Estimation of Moment Condition Models with Dependent Data, Y. Kitamura, Yale University

Moment condition models are frequently used in dynamic econometric analysis. They are particularly useful when one wishes to avoid fully parameterizing the dynamics in the data. Even with such flexibility, however, it is often highly desirable to use an estimation method that is robust against small deviations from the model assumptions. For example, measurement errors, which cause vexing problems in time series analysis, can contaminate observations and thereby leading to such deviations.

Though GMM is generally considered to be a robust estimator, this paper demonstrates that an alternative estimator, which is termed the blockwise minimum Hellinger distance estimator (the blockwise MHDE), possesses desirable optimal properties in terms of robustness. Simulations confirm these theoretical results, and GMM is found to be quite sensitive to deviations of data from the model specification. Previous results obtained by Kitamura, Otsu, and Evdokimov (2008) are extended in several aspects.

2.14 Semiparametric estimation of markov decision processes with continuous state space, S.T. Srisuma, London School of Economics

We provide two-step root T consistent estimators for the structural parameters for a class of semiparametric Markovian discrete choice models. Such models are popular in applied work, in particular with labor and industrial organization. We extend the simple methodology of Pesendorfer and Schmidt-Dengler (2008) to allow for continuous observable state space. This extension is non-trivial as the value functions, to be estimated nonparametrically in the first stage, are defined recursively in a non-linear functional equation. Utilizing structural assumptions, we show how to consistently estimate these infinite dimensional parameters as a solution to an integral equation of type 2, see Linton and Mammen (2005), the solving of which is a well-posed problem. We employ the method of kernel smoothing in the first stage and also provide the distribution theory for the value functions.

2.15 Functional Linear Instrumental Regression, Sébastien Van Bellegem, Toulouse School of Economics

This talk is devoted to the functional linear model $Y = \langle Z, \varphi \rangle + U$ where Z is a random element in a Hilbert space $\mathcal{H}$, φ is an unknown function of $\mathcal{H}$, $\langle \cdot, \cdot \rangle$ is the scalar product in $\mathcal{H}$ and U is a random error that is orthogonal to some functional instrumental variable W. A particular case is given when $\mathcal{H} = \mathbb{R}^p$, in which case we recover the standard linear instrumental regression. We show that solving this problem is a linear ill-posed inverse problem, with a known but data-dependent operator. Our goal is to analyse the rate of convergence of the Tikhonov-regularized estimator, when we premultiply the problem by an instrument-dependent operator B. This extends the Generalized Method of Moments to functional GMM. We then discuss the optimal choice of B and propose an extension of the notion of “weak instrument” to this nonparametric framework.

2.16 Nonparametric estimation of Exact consumer surplus with endogeneity in price, Anne Vanhems, Toulouse School of Economics

This paper deals with nonparametric estimation of variation of exact consumer surplus with endogenous prices. The variation of exact consumer surplus is linked with the demand function via a non linear differential equation and the demand is estimated by nonparametric instrumental regression. We analyze two inverse problems: smoothing the data set with endogenous variables and solving a differential equation depending on this data set. We provide some nonparametric estimator, present results on consistency and optimal choice of smoothing parameters, and compare the asymptotic properties to some previous works.

2.17 Identification and Estimation of a Nonparametric Transformation Model, Hidehiko Ichimura, University of Tokyo

A nonparametric version of the Box-Cox transformation model is considered. Namely, the dependent variable, Y, and a vector of explanatory variables, X , are observed, and (Y,X) is being generated by the model $\Lambda(Y) = m(X) + U$, where Λ is a strictly increasing unknown function and m($\cdot$) is an unknown function, U is an unobserved random variable that is independent of X with the cumulative distribution function F.

Sufficient conditions under which identification of Λ, m(.), and F are achieved are discussed, estimators of these parameters developed, and their consistency and asymptotic distribution theory established.

2.18 Nonparametric Identification in Generalized Competing Risks Models with Applications to Second-Price Auctions, T. Komarova, London School of Economics

This paper proposes an approach to proving nonparametric identification for distributions of bidders' values in asymmetric second-price auctions. I consider the case where bidders have independent private values, and the only available data pertain to the winner's identity and to the transaction price. I provide conditions on observable data sufficient to guarantee point identification. My identification proof is constructive and based on establishing the existence and uniqueness of a solution to the system of non-linear differential equations that describes the relationships between unknown distribution functions and observable functions. It comprises two logical steps: proving the existence and uniqueness of a local solution, and then extending that solution to the whole support.

In addition to the main result, I demonstrate how this approach can obtain identification in more general auction settings, such as those with a stochastic number of bidders, or with weaker support conditions. I also show that my results can be extended to generalized competing risks models. Moreover, contrary to classical competing risks (Roy model) results, I describe how generalized models can yield implications that can help check for model misspecification. Finally, I provide a sieve minimum distance estimator and show that it consistently estimates the underlying valuation distribution of interest.

2.19 Sparse non-Gaussain component analysis, V.Spokoiny, Humboldt University

Non-gaussian component analysis (NGCA) introduced in Blanchard et al (2006) offered a method for high dimensional data analysis allowing for identifying a low-dimensional non-Gaussian component of the whole distribution in an iterative and structure adaptive way. An important step of the NGCA procedure is identification of the non-Gaussian subspace using Principle Component Analysis (PCA) method. This article proposes a new approach to NGCA called \emph{sparse NGCA} which replaces the PCA-based procedure with a new the algorithm we refer to as \emph{convex projection}.

2.20 On PSMD Plug-in Estimation of Functionals of Semi/nonparametric Conditional and unconditional Moment Models, Xiaohong Chen, Yale University

In this paper, we consider estimation of functionals of unknown parameters that are identified via the "plug-in" semi/nonparametric conditional and unconditional moment models, in which the generalized residual functions may be non-pointwise smooth with respect to the unknown functions of endogenous variables. We establish the asymptotic normality of the penalized sieve minimum distance estimator (PSMD) of any functionals, which may or may not be root-n estimable. For functionals that are root-n estimable, our PSMD estimator achieves the semiparametric efficiency bound of Ai and Chen (2005). Regardless whether the functionals are root-n estimable or not, we show that the profile optimally weighted criterion function is chi-square distributed. We provide two example applications: (1) root-n efficient estimation of weighted average derivative of nonparametric quantile instrumental variables (IV) regression; (2) pointwise asymptotic normality of nonparametric quantile IV regression.