Mathematisches Forschungsinstitut Oberwolfach

Report No. 15/2007

Semiparametric and Nonparametric Methods
in Econometrics

Organised by
Yacine Ait-Sahalia (Princeton)
Joel Horowitz (Evanston)
Oliver Linton (London)
Enno Mammen (Mannheim)

March 18th – March 24th, 2007

ABSTRACT. 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. Moreover, the development of non- and semiparametric methods that are suitable to the needs of economics presents a variety of mathematical challenges.

Topics to be addressed in the workshop included nonparametric methods in finance, identification and estimation of nonseparable models, nonparametric estimation under the constraints of economic theory, statistical inverse problems, long-memory time-series, and nonparametric cointegration.

Introduction by the Organisers

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.

Semiparametric and Nonparametric Methods in Econometrics

Table of Contents

Abstracts

Identification with Discrete Outcomes
Andrew Chesher

This paper studies models for discrete outcomes which permit explanatory variables to be endogenous. Outcomes can be binary, or integer valued such as arise when considering counts, or ordered as might be obtained when there is interval censoring of a latent continuous outcome.

Y and U are scalar random variables and X and Z are vector random variables. Y, X and Z are observed but U is not. There is the following model: $\mathcal{D}$.

D1. $Y = h(X, U)$ with U continuously distributed and h is weakly monotonic (normalized caglad, non-decreasing) in its last argument with as codomain the ascending sequence $\{y_m\}_{m=1}^M$ which is independent of X. M may be unbounded.

D2. For some $\tau \in (0, 1)$ there exists Z such that for all z: $\mathrm{pr}[U \leq \tau | Z = z]$ is free of z and normalised to equal τ.

The scalar discrete outcome is determined by a structural function $h(X, U)$. There may be endogeneity in the sense that U and X may not be independently distributed. Condition D2 is a local-to-τ independence condition involving instrumental variables Z.

This paper considers identification of the function $h(x, \tau)$. If h were strictly increasing in U then Y would be continuously distributed and the model is the basis for the identifying models developed in [1] and [2]. Since a discrete outcome can be very "close" to continuous if it has many densely packed points of support it seems plausible that there is an identification result for the discrete outcome case. The contribution of this paper is the development of identification results for this case, a case excluded from consideration in the papers just cited.

Under weak nonparametric restrictions there is only *partial* identification of the structural function h when the outcome it delivers is discrete. As points of support of Y become more dense the sets within which a structural function is identified shrink, approaching point identification results for the continuous Y case under suitable conditions.

The key to analysis of the continuous outcome case is, as shown in [1], the following condition implied by the model set out above.

$$\text{for all } z: \quad \mathrm{pr}[Y \leq h(X, \tau) | Z = z] = \tau$$

Under some additional conditions this leads to point identification of the function $h(\cdot, \tau)$.

When Y is *discrete* the model comprising D1 and D2 implies that $h(\cdot, \tau)$ simultaneously satisfies *two* sets of *inequalities*, as follows.

(1) $\qquad\qquad\qquad \text{for all } z: \quad \mathrm{pr}[Y \leq h(X, \tau) | Z = z] \geq \tau$

(2) $\qquad\qquad\qquad \text{for all } z: \quad \mathrm{pr}[Y < h(X, \tau) | Z = z] < \tau$

This leads to set identification of the structural function $h(\cdot, \tau)$ and can place tight bounds on admissible structural functions when Y has many densely packed points of support.

For any joint distribution of Y and X conditional on $Z = z$, $F^*_{YX|Z}$, and for some set of values $\mathcal{Z}$ within which values of Z can be observed there is a set, $\mathcal{H}^*_\tau(\mathcal{Z})$ of structural functions $h(\cdot, \tau)$ which do not violate the inequalities (1) and (2) for any $z \in Z$. These are the structural functions set identified by the model $\mathcal{D}$. It is interesting to study the properties of this set for particular types of structure. Examples studied in the paper include structures admitted by certain ordered probit and covariate dependent Poisson, binomial and binary logit models with endogeneity. It will be necessary to introduce additional restrictions such as monotonicity with respect to variation in x if the identified set is to be informative. In practice, given an estimate $\hat{F}^*_{YX|Z}$ one can calculate an estimate of $\mathcal{H}^*_\tau(\mathcal{Z})$ and examine hypotheses about features of the structural function. There are challenging inferential problems to be solved here.

The paper complements the analysis of triangular models in [3] in which there is set identification when endogenous variables are discrete as set out in [4]. Detailed results are given in [5]. The results shed light on the impact of endogeneity in situations where outcomes are by their nature discrete, for example where they are counts of events. Classical instrumental variables (IV) attacks fail because the restrictions of the IV model do not lead to point identification when outcomes are discrete. The results are informative about the effect of interval censoring and grouping on the identifying power of models. Calculations show that quite small amounts of discretization due to interval censoring can result in significant degradation in the identifying power of models. This is useful information for designers of survey instruments who have control over the amount of interval censoring banded responses induce.

References

[1] V. Chernozhukov and C. Hansen, *An IV Model of Quantile Treatment Effects*, Econometrica, **73** (2005), 245–261.

[2] V. Chernozhukov, G.W. Imbens and W.K. Newey, *Instrumental Variable Identification and Estimation of Nonseparable Models via Quantile Conditions*, forthcoming in Journal of Econometrics, (2007).

[3] A.D. Chesher, *Identification in Nonseparable Models*, Econometrica, **71** (2003), 1405–1441.

[4] A.D. Chesher, *Nonparametric Identification under Discrete Variation*, Econometrica, **73** (2005), 1525–1550.

[5] A.D. Chesher, *Endogeneity and Discrete Outcomes*, Centre for Microdata Methods and Practice Working Paper CWP 05/07, (2007).

Policy Discontinuity and Duration Outcomes
Gerard J. van den Berg

Regression discontinuity (or discontinuity design) is often used to evaluate policy effectiveness. In case of a policy change at a point of time τ^*, the idea is that a comparison of observed outcomes just before and just after τ^* may provide an estimate of the causal effect of the policy change on individual outcomes. It is often thought that this methodology is not useful if the outcome of interest is a duration variable, like unemployment duration. The main reason is that spells that start before the policy change (and that should provide information on the outcome distribution before the policy change) do not always end before the policy change. The corresponding duration outcomes are affected by both policy regimes.

If one would consider a single cohort of individuals flowing into the state of interest at say $\tau_0 < \tau^*$ then the effect of the policy change cannot be distinguished from the duration dependence of the hazard at $\tau^* - \tau_0$. Suppose one also has data from another cohort, which flows into the state of interest after τ^*. One may restrict attention to exits before duration outcome $\tau^* - \tau_0$, because then in the first cohort there are no outcome durations that are affected by both policy regimes. (Typically a positive fraction of spells in the first cohort will be right-censored at duration $\tau^* - \tau_0$.) One can then restrict attention to the duration distribution on $(0, \tau^* - \tau_0)$ as the outcome of interest (so the outcome is the probability that the duration is smaller than $\tau^* - \tau_0$. However, the smaller $\tau^* - \tau_0$, the less interesting this outcome is, whereas the larger $\tau^* - \tau_0$, the longer one has to wait before the post-policy-change data become available. If one is interested in the effect on the hazard rate after two years of unemployment duration then one would have to wait for two years after the policy change before an estimate can be made. An additional problem is that by comparing pre- and post-policy-change data one can only estimate average effects on the individual hazard if one makes semiparametric assumptions, notably proportionality of the duration dependence effect and the effect of the explanatory variables on the hazard rate, in combination with independence between observed and unobserved individual characteristics. This is problematic because we are primarily interested in features of *individual* hazard rates and because such semi-parametric assumptions may be unappealing.

In this paper we argue that in fact the ongoing spells at the moment of the policy change can be fruitfully used to estimate causal parameters of interest. Specifically, one can estimate an average causal treatment effect on the hazard rate of the duration distribution in the presence of unobserved heterogeneity, without imposing a proportional hazard model structure. The basic insight is that the policy change is an exogenous time-varying binary explanatory variable which jumps only once and whose discontinuity point varies independently across spells that started before τ^*.

Missing and Mismeasured Data in Semiparametric Models
Raymond J. Carroll

The Partially Linear Model With Missing Covariates

Perhaps the most common model used in analyzing observational studies of the causal effect of a possibly multivariate treatment or exposure $X^{T} = (X_1, ..., X_p)$ on a continuous response Y when data are available on one or more continuous pretreatment confounding variables Z is the partial linear model

$$(1) \qquad\qquad Y = X^{T}\beta + \nu(Z) + \epsilon,$$

where β is an unknown parameter, $\nu(\cdot)$ is a smooth unknown function of Z, $E(\epsilon|X, Z) = 0$, and the joint distribution of the regressors (X, Z) is left completely unspecified. Robins, Mark and Newey (1992) prove that this model arises whenever we assume (i) no unmeasured confounders (i.e., ignorability of treatment X within levels of Z) and (ii) a constant additive effect of treatment X on the mean of Y. In particular, given assumption (i), this model is guaranteed to be correctly specified under the causal null hypothesis of no effect of treatment X on Y, as the causal null hypothesis implies (1) with $\beta = 0$. Thus, under (i), an asymptotically correct $1 - \alpha$ confidence interval for β in model (1) provides an asymptotic distribution free α-level test of the causal null hypothesis of no exposure effect. Tests of $\beta = 0$ based on lower dimensional models that impose parametric functional forms on either $\nu(Z)$ and/or the density of $X|Z$ do not provide asymptotically distribution free tests of the causal null hypothesis under (i). Even when (i) cannot be assumed to hold, model (1) remains useful and robust because a large sample test of $\beta = 0$ under model (1) remains an asymptotic distribution free test of the important associational hypothesis that (i) Y is mean independent of X given Z and that (ii) Y is conditionally independent of X given Z.

For these reasons estimation of β in model (1) has been the subject of considerable study; see Härdle, Liang and Gao (2000) for a summary. Our contribution in this paper is to study model (1) when data on X are not fully observed for some study subjects whether by design (as in two stage studies) or by happenstance. The problem of missing exposure variables in regression has been treated in great detail in Robins, Rotnitzky and Zhao (1994). However, those authors assumed a parametric functional form for $\nu(Z)$. For the aforementioned reasons, it is clearly important to relax, as we do in this paper, the assumption that the functional form of $\nu(Z)$ is known. As in Robins et al. we allow the missingness probabilities to depend on both Y and Z but not on the unobserved value of X. Our results include both the case where the missingness probabilities are known (as in a designed two-stage study) and the case where they are unknown. Our results build on the work of Wang, Wang, Gutierrez and Carroll (1998), who consider the nonparametric problem (no X) with missing data; see also Cheng (1990, 1994) and Cheng and Chu (1996).

Not only do we derive the asymptotic distribution of our estimators of β, but we also describe three extensions. First, we compare our methods to those that

use only the complete data with appropriate Horvitz–Thompson weighting, and show that our methods are asymptotically more efficient. Second, besides deriving analytic standard error estimates, we also justify the use of the nonparametric bootstrap in this context. Finally, we show that our methods can be extended to longitudinal and clustered data when working independence is used as the method of estimation, thus extending the work on nonparametric regression for correlated data using working independence (Zeger and Diggle, 1994; Hoover, Rice, Wu and Yang, 1998; Fan and Zhang, 2000; Lin and Ying, 2001) to the missing data context.

We also study asymptotic efficiency for estimation of β in model (1). Here we derive the semiparametric efficient score function and the semiparametric information bound. The semiparametric efficient score function is a solution to a complex integral equation, but in a special case we are able to derive the score function explicitly and compare the result to our methods. Our asymptotic work uses the general asymptotic theory for semiparametric models developed by Newey (1994) and Robins et al. (1994).

References

[1] R.J. Carroll, R.K. Knickerbocker and C.Y. Wang, *Dimension Reduction in Semiparametric Measurement Error Models*, Annals of Statistics, **23** (1995), 161–181.

[2] P.E. Cheng, *Applications of Kernel Regression Estimation: a Survey*, Communications in Statistics, Theory & Methods, **19** (1990), 4103–4134.

[3] P.E. Cheng, *Nonparametric Estimation of Mean Functionals with Data Missing at Random*, Journal of the American Statistical Association, **98** (1994), 81–87.

[4] P.E. Cheng and C.K. Chu, *Kernel Estimation of Distribution Functions and Quantiles with Missing Data*, SS, **6** (1996), 63–78.

[5] W. Härdle, H. Liang and J. Gao, *Partially Linear Models*, Springer Physica-Verlag, Heidelberg, (2000).

[6] W.K. Newey, *The Asymptotic Variance of Semiparametric Estimators*, Econometrica, **62** (1994), 1349–1382.

[7] J.M. Robins, *Correcting for Non-Compliance in Randomized Trials Using Structural Nested Mean Models*, Communications in Statistics, **23** (1994), 2379–2412.

[8] J.M. Robins, S.D. Mark and W.K. Newey, *Estimating exposure effects by modelling the expectation of exposure conditional on confounders*, Biometrics, **48** (1992), 479–495.

[9] J.M. Robins, A. Rotnitzky and L.P. Zhao, *Estimation of Regression Coefficients When Some Regressors are not Always Observed*, Journal of the American Statistical Association, **89** (1994), 846–866.

[10] C.Y. Wang, S. Wang, R.G. Gutierrez and R.J. Carroll, *Local Linear Regression for Generalized Linear Models with Missing Data*, Annals of Statistics, **26** (1998), 1028–1050.

General Semiparametric Measurement Error Modeling

A common practice in facilitating increased level of model flexibility is through nonparametric modeling, resulting in widely used semiparametric models including partially linear models, generalized partially linear models or semiparametric models that contain a single index component. Measurement error problems in such a context are less well studied than their parametric counterpart, probably due to the difficulty of handling multiple infinite dimensional parameters. In this paper, we consider a class of such semiparametric measurement error models. We

will construct estimators for the parametric part of the model that are root-n consistent and asymptotically normally distributed, and for the nonparametric part of the model that enjoys the usual bias and variance properties of the nonparametric estimation. We assume a parametric specification for the measurement error part of the model. The methods are based upon a further parametric specification for the latent variable: if this model specification holds then our methods are semiparametric efficient, while if the latent variable model is misspecified we still obtain root-n consistent and asymptotically normal estimators. As far as we know, this is the first paper on semiparametric measurement error models that proposes general methodology of consistent estimation of parametric and nonparametric parts without resorting to deconvolution method or having to correctly specify a distributional model for the variable measured with error.

An example of such problem is the following. In the Framingham Heart Study data (Kannel et al., 1986), consider a logistic regression of coronary heart disease Y on true systolic blood pressure X and age Z among the nonsmokers. The main interest is in the effect of systolic blood pressure on coronary heart disease. A model that allows for a flexible shape in age is

$$(2) \qquad \text{pr}(Y = 1 | X, Z) = H\left\{\mathcal{B}X + \theta(Z)\right\},$$

where $H(\cdot)$ is the logistic distribution function. Of course, true systolic blood pressure is not observable, and instead we observe W, measured blood pressure. As described by Carroll et al. (2006, Chapter 5), a reasonable model relating W and X is

$$(3) \qquad \log(W - 50) = \log(X - 50) + U,$$

where U is normally distributed with mean zero and variance σ_u^2: Carroll et al. estimate $\sigma_u^2 = 0.0126$ based on $1,615$ degrees of freedom, so that for the purposes of illustration we will consider σ_u^2 as known. Although earlier analysis (Carroll et al., 2006) has assumed a linear model for the age effect, our analysis will show that the linearity assumption is somewhat violated, hence the inclusion of an arbitrary function $\theta(Z)$ is needed.

There are various strategies for analyzing the model (2)-(3). An obvious and reasonable approach in this particular data set is to make the further assumption that $\log(X - 50)$ is normally distributed independently of Z, although a moments analysis suggests that the distribution of this transformed variable is heavier tailed than the normal, with a kurtosis of approximately 9.0. We then have a fully-specified semiparametric model, and we would thus typically apply standard semiparametric methods such as profile likelihood (Severini and Staniswalis, 1994) or backfitting (Chen et al., 2003).

The main point of our paper is illustrated by the following considerations. Assuming that $\log(X - 50)$ is normally distributed implies the assumption that X is a shifted lognormal random variable. However, there is inevitable concern that the analysis will be sensitive to this assumption: the fact that $\log(X - 50)$ has a kurtosis greater than 8.0 indicates a t-distribution with approximately 5 degrees of freedom. A good discussion of this issue is in Gustafson (2004, Chapter 4.6). This

concern is one of the major motivation for the class of functional measurement error methods, including the SIMEX estimator of Cook and Stefanski (1995), which is an approximately consistent estimator. In contrast, we seek methods that are fully consistent and semiparametric efficient when the shifted lognormal assumption is true, and remain fully consistent when the assumption is false.

Our method is based upon a computationally convenient backfitting method for estimating $\theta(\bullet)$, in conjunction with kernel based local polynomial methods. Denote the response variable Y, the predictor measured with error X, and predictors measured without error (S, Z). The likelihood function for Y given (X, S, Z) is

$$[Y|X, S, Z] = p\{y|x, s, z, \mathcal{B}, \theta(z)\} \tag{4}$$

for some unknown function $\theta(z)$ and parameter $\mathcal{B}$. However, instead of observing X, we observe W: which is conditionally independent of Y given (X, S, Z). The likelihood function of W given (X, S, Z) is $p(w|x, s, z, \gamma_{\mathrm{mem}})$ depending on a parameter γ_{mem}. Often, γ_{mem} can be estimated using additional information.

Here, we separate the covariates measured without error into S and Z in order to allow both parametric and nonparametric entry of these covariates. Throughout the paper, S can be ignored without hampering understanding of the methodology.

In order to complete a parametric likelihood specification, one needs a model for the unobservable X given (S, Z), one that we denote by $p_{\mathrm{c}}(x|s, z, \xi_{\mathrm{latent}})$ depending on a parameter ξ_{latent}, where the subscript means conjectured. We assume that ξ_{latent} can be estimated at root-n rate by an estimator $\widehat{\xi}_{\mathrm{latent}}$. We show how to construct estimators of $\mathcal{B}$ such that:

- Whether or not $p_{\mathrm{c}}(x|s, z, \xi_{\mathrm{latent}})$ is correct, the estimator is consistent and asymptotically normally distributed, with limiting distribution independent of the method for estimating ξ_{latent}. If $p_{\mathrm{c}}(x|s, z, \xi_{\mathrm{latent}})$ is correct, the estimator is semiparametric efficient.
- For any chosen distributional model of X given (S, Z), the estimator achieves the minimal estimation variance under such model. That is, no further improvement for estimating $\mathcal{B}$ can be achieved through an improved estimation of $\theta(Z)$.

One interesting example is the partially linear model with measurement error $Y = X\beta + \theta(Z) + \epsilon$, and $W = X + U$, where both ϵ and U are assumed to be normal. When $\theta(Z)$ is replaced by a constant θ in this model, Stefanski and Carroll (1987, Equation 3.5) derived an efficient estimator. We generalize this work to the partially linear model, deriving the semiparametric efficient estimator. When the latent variable X is also assumed normal, the resulting estimator is explicit, and enjoys the robustness property that it is consistent and asymptotically normal even if all the normality assumptions are violated. We also show that this estimator is the same as one proposed in Liang et al. (1999), thus characterizing their estimator in terms of the optimality/sub-optimality property under different conditions.

References

[1] R.J. Carroll, D. Ruppert, T. Tosteson, C. Crainiceanu and M. Karagas, *Nonlinear and Nonparametric Regression and Instrumental Variables*, Journal of the American Statistical Association, **99** (2004), 736–750.

[2] R.J. Carroll, D. Ruppert, C. Crainiceanu and L.A. Stefanski, *Measurement Error in Nonlinear Models: A Modern Perspective*, Second Edition, London (2006): CRC Press.

[3] X. Chen, O. Linton and I. van Keilegom. *Estimation of Semiparametric Models When the Criterion Function Is Not Smooth*, Econometrica, **71** (2003), 1591-1608.

[4] J. Cook and L.A. Stefanski, *A Simulation Extrapolation Method for Parametric Measurement Error Models*, Journal of the American Statistical Association, **89**, (1995), 1314–1328.

[5] W.B. Kannel, J.D. Neaton, D. Wentworth, H.E. Thomas, J. Stamler, S.B. Hulley and M.O. Kjelsberg, *Overall and Coronary Heart Disease Mortality Rates in Relation to Major Risk Factors in 325,348 men Screened for MRFIT*, American Heart Journal, **112** (1986), 825–836.

[6] H. Liang, W. Härdle and R.J. Carroll, *Estimation in a Semiparametric Partially Linear Errors-in-variables Model* Annals of Statistics, **27** (1999), 1519–1535.

[7] H. Liang, S. Wang, J.M. Robins and R.J. Carroll, *Estimation in Partially Linear Models with Missing Covariates* Journal of the American Statistical Association, **99** (2004), 357–367.

[8] W.K. Newey, *Semiparametric Efficiency Bounds* Journal of Applied Econometrics, **5**, (1990), 99–135.

[9] L.A. Stefanski and R.J. Carroll, *Conditional Scores and Optimal Scores for Generalized Linear Measurement-error Models* Biometrika, **74** (1987), 703–716.

[10] A.A. Tsiatis and Y. Ma, *Locally Efficient Semiparametric Estimators for Functional Measurement Error Models* Biometrika, **91** (2004) 835–848.

Semiparametric models with data missing by design and inverse probability weighted empirical processes

Jon A. Wellner

(joint work with Norman E. Breslow)

Weighted likelihood, based on solving Horvitz-Thompson or inverse probability weighted (IPW) versions of the likelihood equations, offers a simple and robust method for fitting models to two phase stratified samples. We consider semiparametric models for which solution of infinite dimensional estimating equations leads to $\sqrt{N}$ consistent and asymptotically Gaussian estimators of both Euclidean and nonparametric parameters. If the phase two sample is selected via Bernoulli (i.i.d.) sampling with known sampling probabilities, standard estimating equation theory shows that the influence function for the weighted likelihood estimator of the Euclidean parameter is the IPW version of the ordinary influence function. We establish weak convergence of the IPW empirical process by borrowing results on weighted bootstrap empirical processes. By using the resulting of the IPW empirical processes, we derive a parallel asymptotic expansion for finite population stratified sampling. The asymptotic variance for Bernoulli sampling involves the within strata second moments of the influence function, while for finite population stratified sampling it involves only the within strata variances. We also show that the latter asymptotic variance also arises when the observed sampling fractions

are used as estimates of those known *a priori*. We propose a general procedure for fitting semiparametric models with estimated weights to two phase data. Several of our key results have already been derived for the special case of Cox regression with stratified case-cohort studies, other complex survey designs and missing data problems more generally. This paper is intended to help place this previous work in appropriate context and to pave the way for applications to other models.

Structural Econometrics
Rosa Matzkin

We discussed nonparametric identification in structural econometric models. These are models that use behavioral and equilibrium assumptions to map the unobservable functions and distributions in the model into the distribution of the observable variables. Nonparametric identification studies conditions under which the unobservable functions and distributions can be recovered from the distribution of the observable variables. We considered models that can be described by a system of nonparametric equations of the type $Y = m(X, e)$, where X and Y are vectors of observable variables, and e is a vector of unobservable variables, such that for some function r, $e = r(Y, X)$. Within these, we consider the identification of the function r and the distribution of e.

Threshold Crossing Models and Bounds on Treatment Effects: A Nonparametric Analysis
Edward Vytlacil
(joint work with Azeem M. Shaikh)

This paper considers the evaluation of the average treatment effect of a binary endogenous regressor on a binary outcome when one imposes a threshold crossing model on both the endogenous regressor and the outcome variable but without imposing parametric functional form or distributional assumptions. Without parametric restrictions, the average effect of the binary endogenous variable is not generally point identified. This paper constructs sharp bounds on the average effect of the endogenous variable that exploit the structure of the threshold crossing models and any exclusion restrictions. We also develop methods for inference on the resulting bounds.

Direct Semiparametric Estimation of the Binary Choice Model with Endogenous Regressors under Varying Identification Conditions

Stefan Hoderlein

In this paper we consider the case of endogenous regressors in the binary choice model without specifying the distribution of the unobserved latent error term to be from a parametric family. We show that the use of instruments in a control function fashion opens up the way to a rich class of direct semiparametric estimators for the slope coefficient.

All estimators within this class have the same building principle in common, namely they are the ratio of the derivatives of two functions of the instruments. These ratio may contain both mean or quantile regression as well as nonseparable functions, depending on the respective assumptions that define the control function residuals. We discuss identification under varying assumptions, as well as the large sample behavior of estimators based on sample counterparts. Simulation and application conclude this paper.

Semiparametric Estimation of Binary Response Models with Endogenous Regressors

Christoph Rothe

In this paper, we investigate the identification and semiparametric estimation of single-index binary choice models with endogenous regressors. The model is usually written in the latent variable form

$$
Y \quad = \quad \begin{cases} 1 & \text{if } Y^* = X\beta_o - U > 0 \\ 0 & \text{else,} \end{cases}
$$

where Y is an indicator of the sign of an unobserved variable Y^* generated through a linear model with regressors X, vector of parameters β_o and error term U. If one is willing to assume that X and U are independent and that the distribution of the error term follows some parametric law, it would be straightforward to estimate β_o by well-established likelihood methods such as Logit or Probit.

Our aim in this paper is to propose an estimator that relys on neither of these two assumptions. This is of considerable practical importance since both might be inappropriate for many empirical applications. First, economic theory usually provides little to no guidance about the functional form of the distribution of the error term, but misspecifications will generally result in inconsistent estimates for likelihood-based approaches. A number of semiparametric estimators have therefore been proposed which do not impose parametric restrictions on the distribution of U (see Horowitz [2] for a review).

Second, when the binary choice model arises in the context of a system of triangular or fully simultaneous equations, or certain measurement error models, some components of X will typically be endogenous and thus correlated with U, violating the independence assumption. Although neglecting this problem will

again render the usual estimates inconsistent, the corresponding literature is much less extensive.

In this paper, we adopt a framework similar to Blundell and Powell [1]. They use a control function approach to identify the vector of parameters, which introduces residuals from a reduced form of the regressors as covariates into the outcome equation to account for endogeneity. This idea is well established in parametric econometrics and has recently been used in the identification and estimation of various non- and semiparametric models with endogenous regressors. It has the drawback that it requires the endogenous regressor to be continuously distributed, but all other variables, including the instruments, can well be discrete.

In the context of a binary choice model, the control function approach for identification can be combined with different estimation procedures. Following one of the suggestions of Blundell and Powell [1], we propose a two-step semiparametric maximum likelihood (SML) estimator for the index coefficients that can be seen as an extension of the Klein and Spady [3] estimator, which achieves the semiparametric efficiency bound in the exogenous case. The first step consists of estimating a reduced form equation for the endogenous regressors and extracting the corresponding residuals. In the second step, the latter are added nonparametrically as control variates to the outcome equation, which is in turn estimated by semiparametric maximum likelihood. The resulting estimator is computationally somewhat more involved than the other competing procedures but still tractable. It possesses the classic desirable asymptotic properties of $\sqrt{n}$-consistency and asymptotic normality, and valid standard errors and tests statistics can be obtained through consistent estimates of the asymptotic covariance matrix. Furthermore, a simulation study we conduct shows that our SML estimator's final sample properties compare very favourably to those of its existing competitors. It should thus be appealing to applied researchers.

References

[1] Blundell, R.W. and Powell, J.L., *Endogeneity in Semiparametric Binary Response Models*, Review of Economic Studies, **71** (2004), 655–679.
[2] Horowitz, J.L., *Semiparametric Methods in Econometrics*, Springer-Verlag (1998).
[3] Klein, R.W. and Spady, R.H., *An Efficient Semiparametric Estimator for Binary Response Models*, Econometrica, **61** (1993), 387–421.

On the Bootstrap of Rank Correlation Estimators

Viktor Subbotin

Estimators based on monotone relations are important tools in econometrics. A classical example is the Maximum Rank Correlation Estimator (MRC) of Han [3]. It applies to the generalized regression model $Y = D \circ F(X'\beta_0, \varepsilon)$, where X (observed) and ε (unobserved) are independent random variables, function D is nondecreasing, F is strictly increasing in both arguments, and β_0 is a finite-dimensional vector of parameters. The binary choice model and the censored regression model are popular examples of this model. In general, D, F, and the

distribution function of ε need not be specified. Given a sample $\{(X_i, Y_i)\}_{i=1}^n$ of i.i.d. observations, β_0 can be consistently estimated, up to scale, by maximizing the objective function

$$\sum_{i \neq j} \mathbf{1}\{Y_i < Y_j\} \mathbf{1}\{X_i'\beta < X_j'\beta\}.$$

with a scale normalization restriction. Estimators with similar structure, with applications to transformation models, censored regressions and panel data have also been proposed by Cavanagh and Sherman [2], Abrevaya [1], Lee [5] and Khan and Tamer [4], among others.

One practical advantage of the rank correlation estimators is that their computation does not depend on any tuning parameters (e.g. bandwidths). Han [3] and Sherman [7] provided general theory for $n^{1/2}$-consistency and asymptotic normality for this type of estimators. The asymptotic variance, however, was previously estimated by either nonparametric or numerical derivative methods, both depending on tuning parameters. The purpose of this paper is to show that the nonparametric bootstrap, which does not involve tuning parameters, consistently estimates the asymptotic distribution function and the asymptotic variance. We also characterize the rates of convergence of the finite-sample distribution and the bootstrap distribution of the estimators to the asymptotic limit.

Setup and Results. Let $\mathcal{H} = \{h_\theta : \mathcal{Z}^m \to \mathbb{R}, \ \theta \in \Theta \subset \mathbb{R}^d\}$ be a family of real symmetric functions defined on $\mathcal{Z}^m = \mathcal{Z} \times ... \times \mathcal{Z}$ (m times). Let P be a data-generating measure on $\mathcal{Z}$. In the general model we assume that the estimated vector of parameters, θ_0, is identified by the relation $\theta_0 = \arg\max_{\theta \in \Theta} P^m h_\theta$, where P^m denotes integration over the measure $P \times ... \times P$ (m times). Given an i.i.d. sample of data $\{Z_i\}_{i=1}^n$, the estimator θ_n is defined as a maximizer of a U-process of order m:

$$(1) \qquad \theta_n = \arg\max_{\theta \in \Theta} \sum_{i_1, ..., i_m \in 1, ... n, \ distinct} h_\theta\left(Z_{i_1}, ..., Z_{i_m}\right)$$

For the bootstrap, form samples $\left\{\hat{Z}_i\right\}_{i=1}^n$ by drawing randomly with replacement from the data set $\{Z_i\}_{i=1}^n$. The bootstrapped estimator, $\hat{\theta}_n$, maximizes the criterion function formed as in (1) with $\hat{Z}_i$ replacing Z_i.

Assumptions made by Han and Sherman for asymptotic normality of MRC are sufficient for consistency of the bootstrap. Assumptions 1-3 below are their stylized version. Additional assumption 4 is usually trivially satisfied for this type of estimators.

Assumption 1. Θ *is a compact set;* $P^m h_\theta$ *is continuous on* Θ *and* θ_0 *is its unique maximum on* Θ.

Assumption 2. $\mathcal{H}$ *is a Euclidean class of functions for a* P^m-*square-integrable envelope* H *in the sense of Nolan and Pollard [6]. As* $\theta \to \theta_0$, $P^2\left[(h_\theta - h_{\theta_0})^2\right] \to 0$.

Assumption 3. *Parameter θ_0 is an interior point of Θ. Function $\tau(z, \theta) = P^{m-1}h_\theta(z, \cdot, ..., \cdot)$ is twice continuously differentiable at θ_0, and there is a P-integrable function $M(z)$ such that for all z and all θ in a neighborhood of θ_0,*

$$|\nabla_2 \tau(z, \theta) - \nabla_2 \tau(z, \theta_0)| \leq M(z) |\theta - \theta_0|,$$

where $\nabla_2 \tau$ is the Hessian matrix of τ with respect to θ. The gradient of $\tau(\cdot, \theta)$ with respect to θ at θ_0, $\nabla \tau(\cdot, \theta_0)$, has finite variance matrix B. Matrix $A = P\nabla_2 \tau(\cdot, \theta_0)$ is finite and negative definite.

Assumption 4. *For every $h_\theta \in \mathcal{H}$, $h_\theta(z, z, z_3, ..., z_m) \equiv 0$.*

Theorem 1. *(a) Under Assumptions 1-3,*

$$(2) \qquad \sup_{q \in \mathbb{R}^d} \left| P\left\{ n^{1/2}(\theta_n - \theta_0) < q \right\} - \Phi_{m^2 A^{-1} B A^{-1}}(q) \right| = o(1)$$

where $\Phi_{m^2 A^{-1} B A^{-1}}$ is the c.d.f. of the $N\left(0, m^2 A^{-1} B A^{-1}\right)$-distribution.
(b) Under Assumptions 1-4, bootstrap of θ_n is consistent in probability:

$$(3) \qquad \sup_{q \in \mathbb{R}^d} \left| \hat{P}\left\{ n^{1/2}(\widehat{\theta}_n - \theta_n) < q \right\} - \Phi_{m^2 A^{-1} B A^{-1}}(q) \right| = o_p(1).$$

The rates of convergence in Theorem 1 depend on a stronger version of Assumption 2. The next condition is relevant for rank correlation estimators whose objective functions involve, like in MRC, the nonsmooth indicator function $\mathbf{1}\{X_i'\beta < X_j'\beta\}$.

Assumption 5. *There exists $C > 0$ such that for all θ_1, θ_2 in a neighborhood of θ_0, $P^2\left[(h_{\theta_1} - h_{\theta_2})^2\right] \leq C|\theta_1 - \theta_2|$.*

Theorem 2. *Let Assumptions 1-3 and 5 hold, $PM^2 < \infty$ and $P\|\nabla\tau(\theta_0, z)\|^4 < \infty$. If H is a constant, put $b_n = n^{-1/6}(\log n)^2$. If $P^m H^k < \infty$ for $k \geq 6$, put $b_n = \left(n^{-1/6}(\log n)^{2/3}\right)^{1/(1+2/3k)}$. Then the left-hand side in (2) is $O(b_n)$. Under additional Assumption 4, the left-hand side in (3) is $O_p(b_n)$.*

References

[1] J. Abrevaya, *Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable*, Journal of Econometrics, **93** (1999), 203–228.

[2] C. Cavanagh and R.P. Sherman, *Rank Estimators for Monotonic Index Models*, Journal of Econometrics, **84** (1998), 351–381.

[3] A.K. Han, *Non-Parametric Analysis of a General Regression Model. The Maximum Rank Correlation Estimator*, Journal of Econometrics, **35** (1987), 303–316.

[4] S. Khan and E. Tamer, *Partial Rank Estimation of Duration Models with General Forms of Censoring*, Journal of Econometrics, **136** (2007), 251–280.

[5] M. Lee, *A root-N consistent semiparametric estimator for related-effect binary response panel data*, Econometrica, **67** (1999), 427–433.

[6] D. Nolan and D. Pollard, *U-Processes: Rates of Convergence*, The Annals of Statistics, **15** (1987), 780–799.

[7] R.P. Sherman, *The Limiting Distribution of the Maximum Rank Correlation Estimator*, Econometrica, **61** (1993), 123–137.

Bootstrap Inference in Partially Identified Models
Federico A. Bugni

Chernozhukov, Hong and Tamer [2] propose a very general approach to perform inference for partially identified models, referred to as the criterion function approach. Their procedure is implemented through subsampling. The goal of this paper is to show that for a class of partially identified models, we can perform the same type of inference using the bootstrap instead of subsampling.

The class of models considered in this paper encompasses relevant economic applications, such as missing data problems and static entry games with multiplicity of equilibria.

The bootstrap procedure we propose differs qualitatively from a bootstrap analogue of Chernozhukov, Hong and Tamer [2]. We show that, in general, replacing subsampling with the bootstrap in their procedure will not result in consistent inference. One of the reasons behind this issue is the well known inconsistency of the bootstrap when the parameter of interest lies in the boundary of the parameter space, as discussed by Andrews [1].

The main contribution of this paper is to provide an alternative bootstrap procedure that avoids these problems. Under our assumptions, we can provide rates of convergence of the error in the coverage probability of the bootstrap approximation. Moreover, we show that the bootstrap has a faster rate of convergence than subsampling, resulting in errors of coverage probability that are orders of magnitude smaller.

Using Monte Carlo experiments, we explore the finite sample behavior of our bootstrap procedure. The simulations show that our bootstrap has satisfactory finite sample performance and a superior performance when compared to subsampling.

References

[1] D.K.W. Andrews, *Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space*, Econometrica, **68** No. 2, (2000), 399–405.
[2] V. Chernozhukov, H. Hong and E. Tamer, *Inference on parameter sets in econometric models. Mimeo: M.I.T., Duke and Northwestern*, (2002).

Dilation bootstrap for inference with incomplete models
Marc Henry
(joint work with Alfred Galichon)

Our basic econometric question is how to reject a given structure based on its observable components. The general methodology proposed here is to test the existence of a match between the observations and the structure. Sampling uncertainty is taken into account through a suitable *dilation* of the structure. The test is inverted to provide confidence regions for partially identified parameters. Economic theory provides us with a structure of relations between observed variables

Y, and latent variables U (bids vs valuations/information, entry vs productivity shocks, chosen level of insurance vs risk level/risk attitude). Without loss of generality, the structure can be formulated as a $U \in \Gamma_\theta(Y)$ and $P_U \in \mathcal{V}_\theta$ where the multi-valued correspondence (many-to-many mapping) Γ_θ, the set of latent variable distributions $\mathcal{V}_\theta$ and the parameter $\theta \in \Theta$ define the structure. Identification obtains if there is a one-to-one correspondence between θ and the data generating process P. When the structure involves censoring, preference heterogeneity, interaction effects, etc..., identification often requires the original structure to be refined with arbitrary equilibrium selection mechanisms or restrictions on unobserved heterogeneity. Otherwise, there is a many-to-many correspondence between θ and P, denoted $P \mapsto \Theta(P)$, and the latter is called identified set. For instance, let $U \in \mathbb{R}$ be the value of houses in a neighbourhood, with median θ. Let $Y \in \mathbb{R}$ be level of insurance coverage chosen for the house. With heterogeneity of risk attitudes, all we know is that $Y \leq U$, so that the structure is summarized by $\Gamma_\theta(Y) = [Y, +\infty)$, $\mathcal{V}_\theta$ is the set of distributions with median θ, and the identified set is $\Theta(P) = [\text{med}(P), +\infty)$. As a second example, consider an m-player game, with strategies $Y = (Y_1, \ldots, Y_m)$, unobserved shocks $U = (U_1, \ldots, U_m)$ to the payoff functions and payoff profiles $\Pi_j = \Pi(Y_j, Y_{-j}, U_j; \theta)$, $\theta \in \Theta$. Maximizing behaviour yields

$$\Gamma_\theta(Y) = \{U \text{ such that for all } j, \ \Pi(Y_j, Y_{-j}, U_j; \theta) \geq \Pi(Y^*, Y_{-j}, U_j; \theta)\}$$

which, combined with an assumption on the distribution of U, defines the structure. Equilibria for this game will generically not be unique, so Γ_θ is generally multi-valued.

The motivation for considering partially identified structures is the following: there is a hierarchy of structural restrictions, some relatively uncontroversial (such as maximizing behaviour), and some more so (such as equilibrium selection mechanisms). One wishes to test a controversial (hence salient) assumption without maintaining a host of equally controversial assumptions necessary for identification, or one wishes to see how the identified set shrinks when one rises in the hierarchy of assumptions. We propose to cover elements of the identified set with confidence regions such that

$$\liminf_n \mathbb{P}\{\theta \in \mathcal{C}_n\} \geq 1 - \alpha \text{ for all } \theta \in \Theta(P)$$

To that end, we find a test $\tau_n^\alpha(\theta)$ such that

- $\limsup_n \mathbb{P}(\tau_n^\alpha(\theta) \neq 0) \leq \alpha$ if the data is compatible with the structure
- $\liminf_n \mathbb{P}(\tau_n^\alpha(\theta) \neq 0) = 1$ if the data and the structure are not compatible.

The region $\mathcal{C}_n = \{\theta : \tau_n^\alpha(\theta) = 0\}$ has the desired pointwise coverage.

The compatibility hypothesis is that one can embed the observable Y into the structure. Formally, there exists a joint distribution π over (Y, U) such that $\pi_Y = P$, ie. the marginal for Y is the true DGP, $\pi_U \in \mathcal{V}_\theta$, ie. the marginal for U satisfies a set of (parametric or nonparametric) restrictions and $\pi(\{U \notin \Gamma_\theta(Y)\}) = 0$, ie. the binary relation defined by Γ_θ is satisfied $\pi-$almost surely. We can rewrite the compatibility condition as $T(P, \theta) := \min_{\pi \in \mathcal{M}(P, \theta)} \pi[U \notin \Gamma_\theta(Y)] = 0$, where

$\mathcal{M}(P, \theta)$ is the set of probability measures which satisfy $\pi_Y = P$ and $\nu \in \mathcal{V}_\theta$. This problem is an optimization problem with a dual program which allows feasible computation. A natural approach would be to use a suitable normalization of $T(P_n, \theta)$ as test statistic. However, there are difficulties in practice, as the limiting distribution of $T(P_n, \theta)$ is non-pivotal, complicated, and needs to be approximated adaptively (i.e. with a user-chosen sequence), and a single evaluation $T(P_n, \theta)$ may be computationally intense, which is a problem when running bootstrap or subsampling procedures.

So we propose an alternative inference methodology to lift this critical burden. The idea is to dilate each point of the sample space into a set of neighbouring points, in order to control for the difference between the true DGP and the empirical distribution. Then we are looking for a match between our latent variable and any point in the neighbourhood. We construct sets $y \to J_n^\alpha(y)$ such that with probability tending to no less than $(1 - \alpha)$ one has a representation (Y, Y^*) with joint distribution ρ such that $Y \sim P$, $Y^* \sim P_n$ and $\rho(Y \in J_n^\alpha(Y^*)) = 1$. As we have representations such that $U \in \Gamma(Y)$, and $Y \in J_n^\alpha(Y^*)$, we can chain these relations (on a common probability space). Thus, with probability tending to no less than $1 - \alpha$ there is a representation (U, Y^*) with joint distribution μ such that $\mu(U \in \Gamma_\theta(J_n^\alpha(Y^*))) = 1$, and checking the latter does not involve sampling uncertainty any more. This provides us a ready-made test statistic, namely

$$\tau_n^\alpha(\theta) = \inf_{\pi \in \mathcal{M}(P_n, \theta)} \pi[\{U \notin \Gamma_\theta(J_n^\alpha(Y^*))\}]$$

and the rejection region is given by $\{\tau_n^\alpha(\theta) \neq 0\}$.

We build the dilation J_n^α with an appeal to the empirical bootstrap principle, i.e. such that $\liminf_n \mathbb{P}^*(\exists \rho^* \in \mathcal{M}(P_n, P^*) : \rho^*(Y^* \in J_n^\alpha(Y^{**}) = 1)) \geq 1 - \alpha$, where $Y^* \sim P_n$ and $Y^{**} \sim P^*$ (bootstrap distribution). The dilation J_n^α can be approximated by simulation: Draw B bootstrap samples $(Y_1^b, \ldots, Y_n^b)$, $b = 1, \ldots B$. For each b, find a 1-to-1 matching (i.e. permutation π^b) of $(Y_1, \ldots, Y_n)$ with $(Y_1^b, \ldots, Y_n^b)$ that minimizes the number of nearest neighbours l^b involved in the matches. Call l_α the $(1 - \alpha)$ quantile of the bootstrap distribution of neighbours, and define $J_n^\alpha(Y_i) = \{$ Y_i and its l_α nearest neighbours $\}$. As an illustration, suppose we want a $1 - \alpha$ confidence interval for the γ quantile $q_Y(\gamma)$ of the true DGP P. We observe P_n, and use the dilation bootstrap to construct $J_n^\alpha(y) = [y - \delta_n^-(y), y + \delta_n^+(y)]$ such that with probability tending to no less than $1 - \alpha$, there exists a joint distribution ρ satisfying $\rho(Y^* - \delta_n^-(Y^*) \leq Y \leq Y^* + \delta_n^+(Y^*)) = 1$. Then $\rho(Y^* - \delta_n^-(Y^*) \leq q_Y(\gamma)) \geq \gamma$ and $\rho(Y^* + \delta_n^+(Y^*) \geq q_Y(\gamma)) \geq 1 - \gamma$. Hence, with probability tending to no less than $1 - \alpha$, we have $q_{Y^* - \delta_n^-(Y^*)}(\gamma) \leq q_Y(\gamma) \leq q_{Y^* + \delta_n^+(Y^*)}(\gamma)$.

However, by the Law of the Iterated Logarithm for the quantile process, the bootstrap dilation procedure proposed above yields a dilation which is asymptotically equivalent to $\delta_n^-(y) = \delta_n^+(y) = (2 \ln \ln n / n f(y))^{1/2}$ where f is the density associated with P. Hence, the true asymptotic coverage is 1 rather than $1 - \alpha$: the insistence on minimax matching yields a dilation that is too conservative. Similar results obtain in higher dimensions (with different rates). So we propose a

modified procedure. We construct sets $y \to J_n^{\alpha,\beta}(y)$ such that with probability tending to no less than $(1 - \alpha)$ one has a representation (Y, Y^*) with joint distribution ρ such that $Y \sim P$, $Y^* \sim P_n$ and $\rho(Y \in J_n^{\alpha,\beta}(Y^*)) \geq 1 - \beta$. $J_n^{\alpha,\beta}$ can be approximated by simulation with a slight modification of the procedure described above. Finally, we revisit the example of the quantile. With probability tending to no less than $1 - \alpha$, there exists a joint distribution ρ satisfying $\rho(Y^* - \delta_n^-(Y^*) \leq Y \leq Y^* + \delta_n^+(Y^*)) \geq 1 - \beta$. Then $\rho(Y^* - \delta_n^-(Y^*) \leq q_Y(\gamma)) \geq \gamma - \beta$ and $\rho(Y^* + \delta_n^+(Y^*) \geq q_Y(\gamma)) \geq 1 - \gamma - \beta$. Hence, with probability tending to no less than $1 - \alpha$, we have $q_{Y^* - \delta_n^-(Y^*)}(\gamma - \beta) \leq q_Y(\gamma) \leq q_{Y^* + \delta_n^+(Y^*)}(\gamma + \beta)$.

Empirical Analysis of Static and Dynamic Models of Strategic Interactions

Han Hong

(joint work with Patrick Bajari)

Our current research project consists of three studies of empirical methods for analyzing static and dynamic models of strategic interactions under different information assumptions. We develop a sequence of estimation methods for strategic interaction models that are both flexible and computationally attractive. These methods are being applied to the market of stock analyst recommendations and the entry behavior in California highway procurement auctions. The first paper is joint work with John Krainer who is at the Federal Reserve Bank of San Francisco and Denis Nekipelov who is at Duke University. The second paper is joint with Victor Chernozhukov who is at MIT. The third paper is joint with Stephen Ryan who is at MIT.

Game theory has had a profound effect on microeconomic theory and theoretical industrial organization in particular. Also, game theory has had an important impact on economic policy, especially in antitrust and regulation. It is therefore desirable to have empirical methods that are applicable when

agents are behaving strategically as predicted by game theory. Following Bresnahan and Reiss, in these papers we study econometric models of gaming where players choose between a finite number of mutually exclusive actions. As in a standard discrete choice model, utility depend on exogenous covariates, preference parameters and random preference shocks. However, these models generalize standard discrete choice models by allowing an agent's utility to depend on the actions of other agents.

In the **first paper**, we study identification and estimation of static versions of these game theoretic models under the assumption that the error terms are private information to each agent. We first demonstrate that exclusion restrictions can generate nonparametric identification of the latent mean utility functions. Secondly, we study a flexible two step semiparametric estimator that is easy to compute and characterize its asymptotic sampling properties. Third, we develop an algorithm that computes the entire set of equilibria in these models. The estimation and computation methods are then applied to the market of stock

analyst recommendations, where we find strong evidence of peer influence and substantial impact of multiple equilibria.

In the **second paper**, we extend these static models to a dynamic setting where agents interact repeatedly in a markov-perfect equilibrium. We first present an identification result for both discrete and continuous state variables by breaking the analysis into two stages. The first stage resembles a single agent dynamic discrete model. In this stage, we show that the expected static mean utility functions can be nonparametrically identified from the data through a single value function iteration as long as the per-period mean utility of one action, e.g. staying out of the market, is normalized to zero. In the second stage, we show how the results from the first paper can be used to recover the structural utility functions from their respective expectations. Our identification analysis naturally leads to a flexible nonparametric estimator and a practical semiparametric model for dynamic oligopolistic models with general continuous or discrete state variables. We briefly describe ongoing work to apply this method to oil exploration using a unique data set.

The first two papers focus on a private information setting, where firms and agents only observe their own private shocks. This assumption can potentially impose restrictions on the model when unobserved heterogeneity is an important element of the market. In the **third paper**, we relax this assumption for static games and allow for a complete information setting where the latent shocks are observable to all the firms and agents. The identification and estimation results that we develop for this paper allow for both multiple and mixed strategy equilibria. By exploiting two recent algorithmic developments, the simulated method of moment estimator that we define has significant computational advantages over existing methods. Not only does it compute all the equilibria of the model, including mixed strategy ones, it also makes use of an importance sampling scheme to allow for speedy optimization of the model parameters. We apply our method to analyze entry behavior in California highway procurement auctions. The empirical analysis recovers significant entry costs by large bidders, and also finds that both multiple and mixed strategy equilibria are important determinants of entry behavior. We also describe a planned project where we wish to use experimental data to flexibly model equilibrium selection in normal form games.

Non-Parametric Estimation of Demand Functions and Bounds Under Revealed Preferences Restrictions
Dennis Kristensen

We consider a simultaneous demand system for individual's consumption of nondurable goods where unobserved heterogeneity enters in a nonadditive manner. We first give sufficient conditions in terms of the individual's preferences for the simultaneous system to be rewritten as a triangular set of equations. We then propose two sets of nonparametric sieve estimators of the demand functions: The first ones are unrestricted estimators and are computed under no constraints on the

consumer's behaviour. The second one are restricted estimators where we impose revealed preferences constraints in the estimation. In both cases, the estimators are essentially nonparametric quantile estimators. We establish the convergence rates of the estimators under regularity conditions. Next, we consider the estimation of demand bounds for the consumer given the arrival of a new set of prices. We demonstrate how these bounds can be estimated based on our estimators of the demand functions obtained in the first step. The estimation problem can be stated within the framework of Chernuzhukov, Hong and Tamer (2007), and we derive the asymptotics of the estimated demand bounds by verifying their general conditions for our specific problem.

Estimation of a Semiparametric Transformation Model
INGRID VAN KEILEGOM
(joint work with Oliver Linton and Stefan Sperlich)

This paper proposes consistent estimators for transformation parameters in semiparametric models. The problem is to find the optimal transformation into the space of models with a predetermined regression structure like additive or multiplicative separability. We give results for the estimation of the transformation when the rest of the model is estimated non- or semi-parametrically and fulfills some consistency conditions. We propose two methods for the estimation of the transformation parameter: maximizing a profile likelihood function or minimizing the mean squared distance from independence. First the problem of identification of such models is discussed. We then state asymptotic results for a general class of nonparametric estimators. Finally, we give some particular examples of nonparametric estimators of transformed separable models. The small sample performance is studied in several simulation exercises.

REFERENCES

[1] X. Chen, O.B. Linton and I. Van Keilegom, *Estimation of semiparametric models when the criterion function is not smooth*, Econometrica, **71** (2003), 1591–1608.
[2] S.C. Cheng, L.J. Wei and Z. Ying, *Analysis of transformation models with censored data*, Biometrika, **82** (1995), 835–845.
[3] J. Horowitz, *Nonparametric Estimation of a Generalized Additive Model With An Unknown Link Function*, Econometrica, **69** (2001), 499–513.
[4] J. Horowitz and E. Mammen, *Rate-Optimal estimation for a general class of nonparametric regression models with unknown link functions*, Manuscript, University of Mannheim (2005).
[5] O.B. Linton, R. Chen, N. Wang and W. Härdle, *An analysis of transformations for additive nonparametric regression*, Journal of the American Statistical Association, **92** (1997), 1512–1521.
[6] O.B. Linton, S. Sperlich and I. Van Keilegom, *Estimation of a semiparametric transformation model*. Annals of Statistics (2007) (under revision).
[7] E. Mammen, O.B. Linton and J.P. Nielsen, *The existence and asymptotic properties of a backfitting projection algorithm under weak conditions*, Annals of Statistics, **27** (1999), 1443–1490.

[8] .J. van den Berg, *Duration models: specification, identification, and multiple durations*, In *The Handbook of Econometrics*, vol. V, eds. J.J. Heckman and E. Leamer. North Holland (2000).

Testing for Stochastic Monotonicity
SOKBAE LEE

(joint work with Oliver Linton and Yoon-Jae Whang)

Let Y and X denote two random variables whose joint distribution is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^2$. Let $F_{Y|X}(\cdot|x)$ denote the distribution of Y conditional on $X = x$. This paper is concerned with testing the stochastic monotonicity of $F_{Y|X}$. Specifically, we consider the hypothesis
(1)

   H_0 : For each $y \in \mathcal{Y}$, $F_{Y|X}(y|x) \leq F_{Y|X}(y|x')$ whenever $x \geq x'$ for $x, x' \in \mathcal{X}$,

where $\mathcal{Y}$ and $\mathcal{X}$, respectively, are the supports of Y and X. We propose a test statistic and obtain asymptotically valid critical values. To our best knowledge, we are not aware of any existing test for (1) in the literature.

This hypothesis can be of interest in a number of applied settings. If X is some policy, dosage, or other input variable, one might be interested in testing whether its effect on the distribution of Y is increasing in this sense. Also, one can test whether the stochastic monotonicity exists in well-known economic relationships such as expenditures (Y) vs. incomes (X) at household levels, wages (Y) vs. cognitive skills (X) using individual data, outputs (Y) vs. the stock of capital (X) at the country level, sons' incomes (Y) vs. fathers' incomes (X) using family data, and so on.

We now describe our test statistic. Let $\{(Y_i, X_i) : i = 1, \ldots, n\}$ denote a random sample from (Y, X). We suppose throughout that the data are i.i.d., but the main result also holds for the Markov time series case where $Y_i = Y_{t+1}$ and $X_i = Y_t$. We actually suppose that X_i is not observed but an estimate $\widehat{X}_i = \psi(W_i, \widehat{\theta})$ is available, where $X_i = \psi(W_i, \theta_0)$ is a known function of observable W_i for some true parameter value θ_0 and $\widehat{\theta}$ is a root-n consistent estimator thereof. Let $1(\cdot)$ denote the usual indicator function and let $K(\cdot)$ denote a one-dimensional kernel function with a bandwidth h_n. Consider the following U-process:

$$\widehat{U}_n(y, x) = \frac{2}{n(n-1)}$$
$$\cdot \sum_{1 \leq i < j \leq n} [1(Y_i \leq y) - 1(Y_j \leq y)] \mathrm{sgn}(\widehat{X}_i - \widehat{X}_j) K_{h_n}(\widehat{X}_i - x) K_{h_n}(\widehat{X}_j - x),$$

where $K_{h_n}(\cdot) = h_n^{-1} K(\cdot/h_n)$ and $\mathrm{sgn}(x) = 1(x > 0) - 1(x < 0)$. Note that the U-process $\widehat{U}_n(y, x)$ can be viewed as a locally weighted version of Kendall's tau statistic, applied to $1(Y \leq y)$ and that $\widehat{U}_n(y, x)$ is related to the U-process considered in Ghosal, Sen, and van der Vaart (2000, equation (2.1)).

Let $U_n(y, x)$ denote $\widehat{U}_n(y, x)$ computed using X_i instead of $\widehat{X}_i$. First, notice that under regularity conditions including smoothness of $F_{Y|X}(y|x)$, as $n \to \infty$,

$$h_n^{-1} E U_n(y, x) \to F_x(y|x) \left(\int \int |u_1 - u_2| K(u_1) K(u_2) du_1 du_2 \right) [f_X(x)]^2,$$

where $F_x(y|x)$ is a partial derivative of $F_{Y|X}(y|x)$ with respect to x. Therefore, since $\widehat{\theta}$ is a consistent estimator, under the null hypothesis such that $F_x(y|x) \leq 0$ for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$, $\widehat{U}_n(y, x)$ is less than or equal to zero on average for large n. Under the alternative hypothesis such that $F_x(y|x) > 0$ for some $(y, x) \in \mathcal{Y} \times \mathcal{X}$, a suitably normalized version of $\widehat{U}_n(y, x)$ can be very large. In view of this, we define our test statistic as a supremum statistic

$$(2) \qquad S_n = \sup_{(y,x) \in \mathcal{Y} \times \mathcal{X}} \frac{\widehat{U}_n(y, x)}{c_n(x)}$$

with some suitably defined $c_n(x)$, which may depend on $(X_1, \ldots, X_n)$ but not on $(Y_1, \ldots, Y_n)$. The U-statistic structure suggests that we use the scaling factor $c_n(x) = \widehat{\sigma}_n(x)/\sqrt{n}$, where

$$\widehat{\sigma}_n^2(x) = \frac{4}{n(n-1)(n-2)} \sum_{i \leq i \neq j \neq k \leq n} \text{sgn}(\widehat{X}_i - \widehat{X}_j) \text{sgn}(\widehat{X}_i - \widehat{X}_k)$$
$$\times K_{h_n}(\widehat{X}_j - x) K_{h_n}(\widehat{X}_k - x) [K_{h_n}(\widehat{X}_i - x)]^2.$$

Our statistic is based on the supremum of a rescaled second order U-process indexed by two parameters x and y, Nolan and Pollard (1987). It generalizes the corresponding statistic introduced by Ghosal, Sen and van der Vaart (2000) for testing the related hypothesis of monotonicity of a regression function.

Our first contribution is to prove that the asymptotic distribution of our test statistic is a Gumbel with certain nonstandard norming constants, thereby facilitating inference using critical values obtained from the limiting distribution. We also show that the test is consistent against all alternatives.

The proof technique is quite complicated and novel because the approximating Gaussian stochastic process contains both a stationary part (corresponding to x) and a nonstationary part (corresponding to y) and so we have to extend existing results that only apply to either one or the other case.

One issue with the extreme value limiting distributions is known to be the poor quality of the asymptotic approximation in the sense that the error declines only at a logarithmic (in sample size) rate. The usual approach to this has been to use the bootstrap, which provides an asymptotic refinement by removing the logarithmic error term and giving an error of polynomial order, Hall (1993). In a special case of ours (of a stationary Gaussian process), Piterbarg (1996) provides a higher order analytic approximation to the limiting distribution that involves including the (known) logarithmic factor in the first order error. His Theorem G1 shows that this corrected distribution is closer to the actual distribution and indeed has an error of polynomial (in sample size) magnitude. We apply this analysis to our

more complicated setting and compute the corresponding "correction" term. Our simulation study shows that this approach gives a dramatic improvement in size.

References

[1] S. Ghosal, A. Sen and A. W. van der Vaart, *Testing monotonicity of regression.* Ann. Statist, **4** (2000), 1054–1082.
[2] P. Hall, *On Edgeworth Expansion and Bootstrap Confidence Bands in Nonparametric Curve Estimation.* J. Royal Statist. Soc., Ser. B. **55** (1993), 291–304.
[3] D. Nolan and D. Pollard, *U-processes: Rates of convergence.* Ann. Statist., **15** (1987), 780–799.
[4] V.I. Piterbarg, *Asymptotic Methods in the Theory of Gaussian Processes and Fields.* Translation of Mathematical Monographs vol 148. American Mathematical Society, Providence, Rhode Island (1996).

Improving Tests with Many Weak Instruments
Yukitoshi Matsushita

This talk is about properties of t-ratios associated with the limited information maximum likelihood (LIML) estimators and likelihood ratio (LR) statistic in a structural form estimation when the number of instrumental variables is large. An asymptotic expansion of the null distribution of a large K t-ratio statistic and an asymptotic null distribution of the LR statistic are derived under large-K_n asymptotics. From these asymptotic approximations, size-improved tests of the t-ratio test and LR test are proposed, respectively.

References

[1] Y. Matsushita, *t-Tests in A Structural Equation with Many Instruments," Discussion Paper CIRJE-F-467*, Graduate School of Economics, University of Tokyo (2006a).
    http://www.e.u-tokyo.ac.jp/cirje/research/dp/2007/2007cf467.pdf
[2] Y. Matsushita, *Approximate Distributions of the Likelihood Ratio Statistic in a Structural Equation with Many instruments*, Discussion Paper CIRJE-F-466, Graduate School of Economics, University of Tokyo (2006b).
    http://www.e.u-tokyo.ac.jp/cirje/research/dp/2007/2007cf466.pdf

Smoothed Lasso for many high-dimensional regressions
Peter Bühlmann

In many application areas, the number of covariates is very large, e.g. in the thousands, while the sample size is quite small, e.g. in the dozens. In such high-dimensional settings, standard exhaustive search methods for variable selection are computationally infeasible and forward selection methods are typically very unstable yielding poor results.

We will show that ℓ^1-penalty methods, i.e. the Lasso [4], can be very useful as a first stage: with high probability, the (mathematically) true model is a subset of the estimated model [1, 2]. Moreover, the adaptive Lasso [5] corrects Lasso's

overestimation behavior yielding a consistent variable selection schemes whose exhaustive computation can be done very efficiently.

Further improvements are possible when having multiple datasets, e.g. over different time points. The prime example is a time course of high-dimensional linear models

$$Y(t) = X(t)\beta(t) + \varepsilon(t)$$

with $n(t) \times 1$ response vector $Y(t)$ and noise term $\varepsilon(t)$ and $n(t) \times p$ design matrix $X(t)$ where typically $p \gg n(t)$. In addition, the high-dimensional $p \times 1$ parameter vector $\beta(t)$ is assumed to change slowly with respect to t. We propose the new Smoothed Lasso [3] which employs a weighted ℓ^1-penalized likelihood (using a kernel function). The smoothed Lasso can lead to markedly improved prediction and variable selection in such time course (or "panel-type") data-structures.

References

[1] N. Meinshausen and P. Bühlmann, *High-dimensional graphs and variable selection with the Lasso*, Ann. Statist. **34** (2006), 1436–1462.
[2] N. Meinshausen and B. Yu, *Lasso-type recovery of sparse representations for high-dimensional data* (2006), Preprint.
[3] L. Meier and P. Bühlmann, *The smoothed Lasso* (2007), Preprint.
[4] R. Tibshirani, *Regression shrinkage and selection via the Lasso*, J. Roy. Statist. Soc. Ser. B **58** (1996), 267–288.
[5] H. Zou, *The adaptive Lasso and its oracle properties*, J. Amer. Statist. Assoc., **101** (2006), 1418–1429.

Inference and Testing for Jumps in Financial Data

Yacine Ait-Sahalia

(joint work with Jean Jacod)

We present a new test to determine whether jumps are present in asset returns or other discretely sampled processes. As the sampling interval tends to 0, our test statistic converges to 1 if there are jumps, and to another deterministic and known value (such as 2) if there are no jumps. The test is valid for all semimartingales, depends neither on the law of the process nor on the coefficients of the equation which it solves, does not require a preliminary estimation of these coefficients, and when there are jumps the test is applicable whether jumps have finite or infinite activity. We then implement the test on simulations and asset returns data.

Next, we discuss estimating the behavior of the jump measure near 0: first if it does not explode near 0, meaning that the number of jumps is finite; second, when this number is infinite, we can say something about the concentration of small jumps. For this purpose, we propose a generalization of the activity index to semimartingales and construct consistent estimators of this index. These estimators are applicable despite the fact that the semimartingale has a continuous part, which makes it more challenging to learn about the small jumps. We can

then test for instance the null hypotheses that jumps have any given fixed degree of activity, or activity greater or smaller than a fixed degree.

Sieve Estimates for Conditional Quantiles of Financial Time Series
JÜRGEN FRANKE
(joint work with Jean-Pierre Stockis and Joseph Tadjuidje-Kamgaing)

We consider a stationary time series $(Y_t, X_t), \infty < t < \infty$, with $Y_t \in \mathbb{R}, X_t \in \mathbb{R}^d$. Our goal is to estimate nonparametrically the conditional α-quantile $q(x)$ of Y_t given $X_t = x$ defined by

$$\mathrm{pr}\{Y_t \leq q(x)|X_t = x\} = \alpha.$$

An immediate application in financial time series analysis is the estimation of conditional $(1 - \alpha)$-Value-at-Risk (VaR) which, if Y_t is a time series of asset returns, is the absolute value of the conditional α-quantile given past values $Y_s, s < t$, as well as past volatilities or data from other financial time series. All of them are combined to form the multivariate time series X_t which is assumed to be observable at time $t - 1$.

Usually, VaR-estimates are calculated from estimates of volatility, i.e. conditional standard deviation of Y_t given X_t. For nonparametric estimates, this approach has been discussed in (Franke et al., 2004). There are some economic reasons against relating extreme losses only to volatility which strongly depends on th frequent small fluctuations of the return time series (Engle and Manganelli, 2002). The main disadvantage, however, is the necessity to specify the distribution of innovations which is a difficult problem in practice.

We therefore follow the direct regression quantile approach of (Koenker and Bassett, 1978) and note that the conditional quantile function $q(x)$ is characterized as the solution of

$$(1) \qquad \mathrm{E}\{|Y_t - q(x)|_\alpha \,|\, X_t = x\} = \min_{f \in L^1(\mu)} \mathrm{E}\{|Y_t - f(x)|_\alpha \,|\, X_t = x\},$$

where μ denotes the stationary law of X_t and $|u|_\alpha = (1-\alpha)u^- + \alpha u^+$. We consider a sieve of function classes $\mathcal{F}_n \subset L^1(\mu)$ increasing with sample size n and with a union which is dense in $L^1(\mu)$. A sieve estimate of $q(x)$ is then given as solution of the sample version of (1):

$$q_n = \mathrm{argmin}_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{t=1}^n \big|Y_t - f(X_t)\big|_\alpha.$$

Finally, we truncate $q_n(x)$ at $\pm\Delta_n$ and get the bounded sieve estimate $\hat{q}_n(x)$. Under appropriate, rather weak assumptions we show that this estimate is non-parametrically consistent in the mean, i.e.

$$\mathrm{E} \int |\hat{q}_n(z) - q(z)|\mu(dz) \to 0,$$

as well as a.s., provided that the *size* of $\mathcal{F}_n$, measured by means of appropriate covering numbers, and Δ_n go to ∞ with the right rate depending on n. The proof is based on a Vapnik-Cervonenkis (1971) inequality holding for stationary processes (Franke and Diagne, 2006) and on a modification of the L^2-regression techniques of (Györfy et al., 2002).

As two specific examples for sieves, we consider feedforward neural networks with one hidden layer, already investigated by (Chen and White, 1999) in the context of quantile estimation, and piecewise constant functions motivated by the qualitative threshold ARCH-model of (Gouriéroux and Montfort, 1992). For those two function classes, we derive the specific rate conditions from the general theorem, resulting in

- $\Delta_n, H_n \to \infty$, $\quad \Delta_n H_n \log(\Delta_n H_n)/\sqrt{n} \to 0$ for neural networks, H_n being the number of neurons in the hidden layer,
- $\Delta_n, H_n \to \infty$, $\quad \Delta_n H_n \log(\Delta_n)/\sqrt{n} \to 0$ for piecewise constant functions, H_n being the number of subsets of the input space on which the function is constant.

For the qualitative threshold quantile estimates, we also discuss how to choose the underlying partition data-adaptively via a CART-like algorithm similar to (Audrino and Bühlmann, 2001). We illustrate the performance of those conditional quantile estimates with some simulations and an application to stock price returns.

REFERENCES

[1] F. Audrino and P. Bühlmann. *Tree-structured GARCH Models.* J. Royal Statist. Soc., Ser. B, **63** (2001), 727–744.

[2] X. Chen and H. White. *Improved Rates and Asymptotic Normality for Nonparametric Neural Network Estimators.* IEEE Trans. Inform. Theory, **45** (1999), 682–691.

[3] R.F. Engle and S. Manganelli. *CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles.* To appear in Journal of Business and Economic Statistics (2002).

[4] J. Franke, M. Diagne and P. Mwita. *Nonparametric Value-at-Risk Estimates.* Oberwolfach Reports **1** (2004), 133–134.

[5] J. Franke and M. Diagne. *Estimating Market Risk with Neural Networks.* Statistics and Decisions **24** (2006), 233–253.

[6] J. Franke, J.P. Stockis and J. Tadjuidje. *Quantile Sieve Estimates for Time Series.* Report in Wirtschaftsmathematik (2007), TU Kaiserslautern.

[7] C. Gouriéroux and A. Montfort. *Qualitative Threshold ARCH Models.* Journal of Econometrics **52** (1992), 159–200.

[8] L. Györfy, M. Kohler, A. Krzyzak and H. Walk. *A Distribution-Free Theory of Nonparametric Regression.* Springer-Verlag (2002), Heidelberg.

[9] R. Koenker und G. Bassett. *Regression quantiles.* Econometrica **46** (1978), 33–50.

[10] V.N. Vapnik and A.Y. Chervonenkis. On the Uniform Convergence of Relative Frequencies
     of Events to their Probabilities. Theory of Probability and its Applications **16** (1971), 264–
     280.

Empirical Pricing Kernels and Investor

WOLFGANG HÄRDLE

Numerous attempts have been undertaken to describe basic principles on which the behaviour of individuals are based. Expected utility theory was originally proposed by J. Bernoulli in 1738. In his work J. Bernoulli used such terms as risk aversion and risk premium and proposed a concave (logarithmic) utility function, see Bernoulli (1956). The utilitarianism theory that emerged in the 18th century considered utility maximization as a principle for the organisation of society. Later the expected utility idea was applied to game theory and formalized by von Neumann and Morgenstern (1944). A utility function relates some observable variable, in most cases consumption, and an unobservable utility level that this consumption delivers. It was suggested that individualsí preferences are based on this unobservable utility: such bundles of goods are preferred that are associated with higher utility levels. It was claimed that three types of utility functions ñ concave, convex and linear ñ correspond to three types of individuals ñ risk averse, risk neutral and risk seeking. A typical economic agent was considered to be risk averse and this was quantified by coefficients of relative or absolute risk aversion. Another important step in the development of utility theory was the prospect theory of Kahneman and Tversky (1979). By behavioural experiments they found that people act risk averse above a certain reference point and risk seeking below it. This implies a concave form of the utility function above the reference point and a convex form below it. Besides these individual utility functions, market utility functions have recently been analyzed in empirical studies by Jackwerth (2000), Rosenberg and Engle (2002) and others. Across different markets, the authors observed a common pattern in market utility functions: There is a reference point near the initial wealth and in a region around this reference point the market utility functions are convex. But for big losses or gains they show a concave form ñ risk aversion. Such utility functions disagree with the classical utility functions of von Neumann and Morgenstern (1944) and also with the findings of Kahneman and Tversky (1979). They are however in concordance with the utility function form proposed by Friedman and Savage (1948). In this paper, we analyze how these market utility functions can be explained by aggregating individual investorsí attitudes. To this end, we first determine empirical pricing kernels from DAX data. Our estimation procedure is based on historical and risk neutral densities and these distributions are derived with stochastic volatility models that are widely used in industry. From these pricing kernels we construct the corresponding market utility functions. Then we describe our method of aggregating individual utility functions to a market utility function. This leads to an inverse problem for 1 the density function that describes how many investors have the utility function

of each type. We solve this problem by discrete approximation. In this way, we derive utility functions and their distribution among investors that allow to recover the market utility function. Hence, we explain how (and what) individual utility functions can be used to form the behaviour of the whole market. We describe the theoretical connection between utility functions and pricing kernels. We present a consistent stochastic volatility framework for the estimation of both the historical and the risk neutral density. Moreover, we discuss the empirical pricing kernel implied by the DAX in 2000, 2002 and 2004. We explain the utility aggregation method that relates the market utility function and the utility functions of individual investors. This aggregation mechanism leads to an inverse problem that is analyzed and solved in this section. We conclude and discuss related approaches.

Inference for Realised Volatility using Infill Subsampling
Ilze Kalnina
(joint work with Oliver Linton)

The subsampling method of Politis and Romano (1994) has been shown to be useful in many situations as a way of conducting inference under weak assumptions and without utilizing knowledge of limiting distributions. Recently, the word subsampling has been used in connection with the estimation of quadratic variation of a semimartingale subject to market microstructure noise, see Zhang, Mykland, and Aït-Sahalia (2005) and Barndorff-Nielsen and Shephard (2007). The subsampling scheme in this setting is slightly different from the usual one and is perhaps better called 'infill price subsampling', as subsamples there consist of prices on a lower frequency. Zhang, Mykland, and Aït-Sahalia (2005) use this 'infill price subsampling' to define a bias correction method that achieves consistent estimation. It is the purpose of this paper to explore the use of this infill subsampling as a means to conducting inference. We show that in an infill sampling scheme the usual subsampling method described in Politis and Romano (1994) does not achieve the required consistency. We show that infill price subsampling delivers an asymptotically unbiased estimator of the asymptotic variance of the estimator of interest (that is realised volatility in our paper), but it is still inconsistent (for the variance of the realised volatility). We propose an infill returns subsampling method that delivers consistent estimator of the asymptotic variance under some smoothness assumption on the volatility. We follow the notation of Politis, Romano, and Wolf (1999) for easier comparison with the usual subsampling. We also conduct a simulation study where we simulate a log-price sample paths that follow a Heston model, which clearly reflects the theoretical properties of the different subsampling approaches that we derive.

References

[1] O.E. Barndorff-Nielsen and N. Shephard, *Variation, jumps, market frictions and high frequency data in financial econometrics. Advances in Economics and Econometrics.* Theory

and Applications, Ninth World Congress, (edited by Richard Blundell, Persson Torsten and Whitney K Newey), Econometric Society Monographs, Cambridge University Press (2007).

[2] D.N. Politis and J.P. Romano, *Large sample confidence regions based on subsamples under minimal assumptions.* Annals of Statistics, **22** (1994), 2031–2050.

[3] L. Zhang, P. Mykland and Y.A ït-Sahalia, *A tale of two time scales: determining integrated volatility with noisy high-frequency data.* Journal of the American Statistical Association, **100** (2005), 1394–1411.

Inference for diffusion processes in the simultaneous presence of noise and jumps

Mark Podolskij

We consider a noisy diffusion model of the type

$$Z = X + U + J \ ,$$

where

$$X_t = X_0 + \int_0^t a_u du + \int_0^t \sigma_u dW_u \ , \qquad t \in [0,1]$$

J is a jump process and U is an i.i.d. process (independent of X), observed at time points i/n, $i = 0, \ldots, n$. We propose a new methodology for the estimation of the volatility functionals of the form

$$\int_0^1 |\sigma_u|^p du$$

for some $p \geq 0$. In particular, this approach provides consistent estimates for the integrated volatility ($p = 2$) and integrated quarticity ($p = 4$), which play a crucial role in econometrics. Furthermore, the new method provides tests for jumps in the process Z, estimates for the quadratic variation of $X + J$ and solutions for some related problems.

The main idea of the new approach can be described as follows. In a first step we apply a smoothing procedure to balance the influence of the noise and the continuous part. After that we construct the class of bipower variation statistics based on the transformed data. Under very mild assumptions we show convergence in probability for this class. The limit term contains a volatility part as well as the second moment of the noise. After estimating the second moment of the noise we can bias-correct the limit, so we finally obtain consistent estimates of volatility functionals. Furthermore, by using appropriate powers we obtain estimators which are robust to the jump component J.

Under some stronger assumption we present a stable central limit theorem for the class of bipower variation statistics. The resulting convergence rate is $n^{-1/4}$, which is known to be optimal. By the approximation of the conditional variance we obtain a feasible (standard) central limit theorem, which enables us to construct confidence bands for the estimated quantities.

Issues in Semiparametric Modelling of Multivariate Long Memory Time Series

Peter M. Robinson

Moving from univariate to bivariate jointly dependent long memory time series introduces a phase parameter (γ), at the frequency of principal interest, zero; for short memory series $\gamma = 0$ automatically. The latter case has also been stressed in the long memory case, along with the "fractional differencing" case $\gamma = (\delta_2 - \delta_1)\pi/2$, where δ_1, δ_2 are the memory parameters of the two series. We develop time domain conditions under which these are and are not relevant, and relate the consequent properties of cross-autocovariances to ones of the (possibly bilateral) moving average representation which, with martingale difference innovations of arbitrary dimension, is used in asymptotic theory for local Whittle parameter estimates depending on a single smoothing number. Incorporating also a regression parameter (β) which, when non-zero, indicates cointegration, the consistency proof of these implicitly-defined estimates is nonstandard due to the β estimate converging faster than the others. We also establish joint asymptotic normality of the estimates, and indicate how this outcome can apply in statistical inference on several questions of interest. Issues of implemention are discussed, along with implications of knowing β and of correct or incorrect specification of γ, and possible extensions to higher-dimensional systems and nonstationary series.

Nonstationary Nonparametric Regression

Melanie Schienle

This article studies nonparametric estimation of a regression model for $d \geq 2$ nonstationary regressors. Given n joint observations $(X, Y) \in \mathbb{R}^{d+1}$, I estimate an additive conditional mean function

$$(1) \qquad Y_i = m_0 + \sum_{j=1}^{d} m_j(X_i^j) + \epsilon_i \quad \text{for all } i \in \{1, \dots, N\}$$

under suitable identification conditions for the component functions. Furthermore Y and all univariate X^j and pairs of bivariate marginal components X^{jk} are (potentially nonstationary) β–Harris recurrent processes. Under different types of independence assumptions, results are derived for the general case of a (potentially nonstationary) β–Harris recurrent noise term ε but also for the special case of ε being stationary mixing. The later case deserves special attention since the model might be regarded as an additive type of cointegration model. In contrast to the existing more general approach in [1], the number of cointegrated regressors is not restricted.

In economic time series we often deal or should truly deal with nonstationary components. In reality, neither prices nor exchange rates nor other macro variables thoroughly follow an invariant stationary law over time (See e.g [4] in demand).

Thus practitioners might feel more comfortable avoiding restrictions like stationarity or not testable mixing conditions. However, it might also appear inappropriate to exclusively impose nonstationary behavior in the model specification. So far an econometrician has had to make a decision upfront: Model the situation parametrically and have nonstationary components or have a nonparametric procedure in a stationary environment. But can we not have both nonparametrics and nonstationarity?

In this paper, I want to explore how much generality can be theoretically admitted at which possible results. I present a nonparametric procedure which offers a uniform treatment of certain nonstationary and all stationary cases within a suitably chosen class of processes.

The appropriate framework for nonstationary kernel type inference is Harris recurrence. This assumption allows for a certain type of nonstationarity of the processes. Intuitively, it is the minimal assumption to still ensure consistency of any type of nonparametric Kernel estimator. Within the imperceptibly smaller class of β–recurrent processes, nonstationarity does not change the type of estimation procedure applied. The degree of nonstationarity of the data is captured by a single parameter β – the degree of regular variation of the recurrence time process. It also represents the polynomial degree of the expected stochastic rates of convergence and therefore offers an important way to compare the nonstationary results to the well–known stationary theorems. The method works irrespective of stationarity or not and is the same in both cases.

The idea of Harris recurrence as the key property for Kernel regression with Markov processes was first suggested by Yakowitz [8]. But he only studied the positive recurrent case. Park and Phillips were the first to move towards possibly null recurrent processes in [7]. Their approach, however, was still quite restrictive as it was valid only for one dimensional processes on a Brownian space with a constant link function. Independently Moloche [6] and Karlsen and Tjøstheim [3] have introduced an estimation framework for regression with general recurrent Markov processes. While the first use embedding techniques under quite restrictive assumptions and employs existing results from probability theory literature, the later is more general with different direct techniques.

In general, as in the stationary case, high-dimensional nonparametric regression models suffer from a curse of dimensionality (COD). The more regressors are included the worse the finite sample behavior. In the stationary mixing case, additive models have provided a powerful technique to overcome this problem and to still maintain high flexibility. But in the nonstationary setting an additional even more severe curse of dimensionality complicates nonparametric estimation. For dimensions larger than two, Harris recurrence of the joint regressors is quite unlikely. In fact the more regressors are added, the more unlikely it is to still fit the framework of Harris recurrence. Most prominently, a random walk is null recurrent only up to dimension two and transient for any higher dimension. In such cases, none of the existing procedures in [1] and [6] can be applied any more.

There is no nonparametric method at all.

In this paper, I provide an estimation method which countervails both curses of dimensionality. To overcome the first, ordinary COD an additive model is estimated. In order to tackle the second nonstationary COD, however, an estimation method for the additive model must rely on low dimensional components only - at best only univariate and bivariate ones to include the widest possible class of processes. In a stationary setting, smooth backfitting introduced by Mammen et al. in [5] fulfills both requirements. As the only estimation procedure for additive models, it does not need a full–dimensional estimate in any step of the method, but uses estimates of one– and two dimensional marginals only.

For nonstationary data, smooth backfitting estimates (SBE) are still defined as minimizers of a smoothed least squares criterion. It is sufficient to assume all pairwise bivariate marginal processes to be recurrent. Full dimensional recurrence is not needed, therefore e.g. high dimensional random walks can be fitted, and a broad class of models can be treated for which there has been no method so far.

In a general nonstationary setting, the derived general backfitting iteration operator has an additional factor which does not vanish and which complicates the analysis. Asymptotic properties are derived under weak conditions. Rates of convergence are of univariate type but governed by the most nonstationary univariate component. In order to achieve asymptotic normality, the speed of convergence is stochastic due to the nonstationarity of the data. The variance of a single component, however, is shown to be the corresponding marginal variance type expression. The similar the regressors are in their degree of nonstationarity, the more efficient is the estimator. Oracle efficiency i.e., the same asymptotic bias and variance as the theoretical estimator based on the knowledge of all other components, can only be achieved if the degree of nonstationarity is exactly the same for all components. Finite sample properties are evaluated using a simulation study for a five dimensional random walk.

References

[1] H.A. Karlsen, T. Myklebust and D. Tjøstheim, *Nonparametric estimation in a nonlinear cointegration type model*, working paper (2005), 1–57.

[2] H. Karlsen and D. Tjøstheim, *Nonparametric estimation in null–recurrent time series*, Working paper HU Berlin (1998).

[3] H. Karlsen and D. Tjøstheim, *Nonparametric estimation in null–recurrent time series*, Annals of Statistics **29** (2001), no. 2, 372–416.

[4] A. Lewbel and S. Ng, *Demand systems with nonstationary regressors*, Rev. of Econ. and Stat. **87** (2005), no. 3, 479–494.

[5] E. Mammen, O. Linton, and J. Nielsen, *The existence and asymptotic properties of a backfitting projection algorithm under weak conditions*, Annals of Statistics **27** (1999), no. 5, 1443–1490.

[6] G. Moloche, *Kernel regression for nonstationary harris–recurrent processes*, MIT working paper 2001, 2001.

[7] C.B. Phillips and J.Y. Park, *Nonstationary density estimation and kernel autoregression*, Cowles Foundation Discussion Papers 1181, Cowles Foundation, Yale University, June 1998,

available at
http://ideas.repec.org/p/cwl/cwldpp/1181.html.
[8] S. Yakowitz, *Nonparametric density and regression estimation for markov sequences without mixing assumptions*, Journal of of Multivariate Analysis, **30** (1989), 359–372.

Probability and moment inequalities for sums of weakly dependent random variables

Michael H. Neumann

(joint work with Paul Doukhan, Efstathios Paparoditis)

1. Weak dependence vs. mixing

For a long time mixing conditions have been the dominating type of conditions for imposing a restriction on the dependence between time series data. They are considered to be useful since they are fulfilled for many classes of processes and since they allow to derive tools similar to those in the independent case. On the other hand, it turns out that certain classes of processes which are of interest in statistics are not mixing although a successive decline of the influence of past states takes place. The simplest example of such a process is an AR(1)-process, $X_t = \theta X_{t-1} + \varepsilon_t$, where the innovations are independent and identically distributed with $P(\varepsilon_t = 1) = P(\varepsilon_t = -1) = 1/2$ and $0 < |\theta| \le 1/2$; see also [11]. It is clear that this process has a stationary distribution supported on $[-2, 2]$, and for a process in the stationary regime, it can be seen from the equality $X_t = \varepsilon_t + \theta \varepsilon_{t-1} + \cdots + \theta^{t-s-1} \varepsilon_{s+1} + \theta^{t-s} X_s$ that a past state X_s can always be recovered from X_t. (Actually, since $|\varepsilon_t| > |\theta||\varepsilon_{t-1}| + \cdots + |\theta|^{t-s-1}|\varepsilon_{s+1}| + |\theta|^{t-s}|X_s|$ it follows that X_t has always the same sign as ε_t which means that we can recover ε_t and therefore X_{t-1} from X_t. Continuing in this way we can finally compute X_s.) This, however, excludes any of the commonly used mixing properties to hold. On the other hand, X_s loses its impact on X_t as $t \to \infty$.

Besides this somehow artificial example, there are many other processes of this type which are of great interest in statistics. For example, for bootstrapping a linear autoregressive process of finite order, it is most natural to estimate first the distribution of the innovations by the empirical distribution of the (possibly re-centered) residuals and to generate then a bootstrap process iteratively by drawing independent bootstrap innovations from this distribution. Now it turns out that commonly used techniques to prove mixing for autoregressive processes fail; because of the discreteness of the bootstrap innovations it is in general impossible to construct a coupling of two processes with different initial values.

Inspired by such problems, [5] and [1] introduced the alternative notions of weak dependence and ν-mixing, respectively, which focus on covariances rather than the total variation norm between the joint distribution and the product of marginal distributions of random variables. A slightly simplified version of Doukhan and Louhichi's definition is given here:

Definition 1.1. A process $(X_t)_{t \in \mathbb{Z}}$ is called *weakly dependent* if there exists a universal null sequence $(\epsilon_r)_{r \in \mathbb{N}}$ such that, for any k-tuple $(s_1, \ldots, s_k)$ and any l-tuple $(t_1, \ldots, t_l)$ with $s_1 \leq \ldots \leq s_k < s_k + r = t_1 \leq \ldots \leq t_l$ and arbitrary measurable functions $g : \mathbb{R}^k \to \mathbb{R}$, $h : \mathbb{R}^l \to \mathbb{R}$ with $\|g\|_\infty \leq 1$ and $\|h\|_\infty \leq 1$, the following inequality is fulfilled:

$$|\mathrm{cov}\,(g(X_{s_1}, \ldots, X_{s_k}), h(X_{t_1}, \ldots, X_{t_l}))| \ \leq \ \psi(k, l, \mathrm{Lip}\,g, \mathrm{Lip}\,h)\,\epsilon_r.$$

Here $\mathrm{Lip}\,h$ denotes the Lipschitz modulus of continuity of h, that is,

$$\mathrm{Lip}\,h \ = \ \sup_{x \neq y} \frac{|h(x) - h(y)|}{\|x - y\|_{l_1}},$$

where $\|z\|_{l_1} = \sum_i |z_i|$, and $\psi : \mathbb{N}^2 \times \mathbb{R}_+^2 \to [0, \infty)$ is an appropriate function.

2. Tools under weak dependence

First central limit theorems for weakly dependent sequences were given by Corollary A in [5] and Theorem 1 in [2]. While the former result is for sequences of stationary random variables, the latter one is tailor-made for triangular arrays of asymptotically sparse random variables as they appear with kernel density estimators. The following central limit theorem for general triangular schemes of weakly dependent random variables was proved in [10]. An interesting aspect of this result is that no moment condition beyond Lindeberg's is required.

Theorem 2.1. *Suppose that $(X_{n,k})_{k=1,\ldots,n}$, $n \in \mathbb{N}$, is a triangular scheme of (row-wise) stationary random variables with $\mathbb{E}X_{n,k} = 0$ and $\mathbb{E}X_{n,k}^2 \leq C < \infty$. Furthermore, we assume that*

$$(1) \qquad \frac{1}{n} \sum_{k=1}^n \mathbb{E}X_{n,k}^2 I(|X_{n,k}|/\sqrt{n} > \epsilon) \underset{n \to \infty}{\longrightarrow} 0$$

holds for all $\epsilon > 0$ and that

$$(2) \qquad \mathrm{var}(X_{n,1} + \cdots + X_{n,n})/n \underset{n \to \infty}{\longrightarrow} \sigma^2 \in [0, \infty).$$

For $n \geq n_0$, there exists a monotonously nonincreasing and summable sequence $(\theta_r)_{r \in \mathbb{N}}$ such that, for all indices $1 \leq s_1 < s_2 < \ldots < s_u < s_u + r = t_1 \leq t_2 \leq n$, the following upper bounds for covariances hold true: for all measurable and quadratic integrable functions $f : \mathbb{R}^u \longrightarrow \mathbb{R}$,

$$(3) \qquad |\mathrm{cov}\,(f(X_{n,s_1}, \ldots, X_{n,s_u}), X_{n,t_1})| \ \leq \ \sqrt{Ef^2(X_{n,s_1}, \ldots, X_{n,s_u})}\,\theta_r,$$

for all measurable and bounded functions $f : \mathbb{R}^u \longrightarrow \mathbb{R}$,

$$(4) \qquad |\mathrm{cov}\,(f(X_{n,s_1}, \ldots, X_{n,s_u}), X_{n,t_1} X_{n,t_2})| \ \leq \ \|f\|_\infty\,\theta_r,$$

where $\|f\|_\infty = \sup_{x \in \mathbb{R}^u} |f(x)|$. Then

$$\frac{1}{\sqrt{n}}\,(X_{n,1} + \cdots + X_{n,n}) \ \xrightarrow{\ d\ } \ \mathcal{N}(0, \sigma^2).$$

The following Bernstein-type inequality which generalizes and improves previous inequalities of [5] and [9] was proved in [6].

Theorem 2.2. *Suppose that $X_1, \ldots, X_n$ are real-valued random variables with zero mean, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $\Psi : \mathbb{N}^2 \to \mathbb{N}$ be one of the following functions:*

  (a) $\Psi(u, v) = 2v$,
  (b) $\Psi(u, v) = u + v$,
  (c) $\Psi(u, v) = uv$,
  (d) $\Psi(u, v) = \alpha(u + v) + (1 - \alpha)uv$,   *for some $\alpha \in (0, 1)$.*

We assume that there exist constants $K, M, L_1, L_2 < \infty$, $\mu, \nu \geq 0$, and a nonincreasing sequence of real coefficients $(\rho(n))_{n \geq 0}$ such that, for all u-tuples $(s_1, \ldots, s_u)$ and all v-tuples $(t_1, \ldots, t_v)$ with $1 \leq s_1 \leq \cdots \leq s_u \leq t_1 \leq \cdots \leq t_v \leq n$ the following inequalities are fulfilled:

$$(5) \quad |\mathrm{cov}\,(X_{s_1} \cdots X_{s_u}, X_{t_1} \cdots X_{t_v})| \;\leq\; K^2\, M^{u+v-2}\, ((u+v)!)^\nu\; \Psi(u, v)\, \rho(t_1 - s_u),$$

where

$$(6) \qquad\qquad \sum_{s=0}^{\infty} (s + 1)^k \rho(s) \;\leq\; L_1\, L_2^k\, (k!)^\mu \qquad \forall k \geq 0,$$

and

$$(7) \qquad\qquad \mathbb{E}|X_t|^k \;\leq\; (k!)^\nu\, M^k \qquad \forall k \geq 0.$$

Then, for all $t \geq 0$,

$$(8) \qquad \mathbb{P}\,(S_n \geq t) \;\leq\; \exp\left(-\frac{t^2/2}{A_n \,+\, B_n^{1/(\mu+\nu+2)} t^{(2\mu+2\nu+3)/(\mu+\nu+2)}} \right),$$

where A_n can be chosen as any number greater than or equal to σ_n^2 and

$$B_n \;=\; 2\, (K \vee M)\, L_2\, \left(\left(\frac{2^{4+\mu+\nu}\, n\, K^2\, L_1}{A_n} \right) \vee 1 \right).$$

Remark 1. *(i)*   Inequality (8) resembles the classical Bernstein inequality for independent random variables. Asymptotically, σ_n^2 is usually of order $O(n)$ and A_n can be chosen equal to σ_n^2 while B_n is usually $O(1)$ and hence negligible. In cases where σ_n^2 is very small or where knowledge of the value of A_n is required for some statistical procedure, it might, however, be better to choose A_n larger than σ_n^2. It follows from (5) and (6) that a rough bound for σ_n^2 is given by

$$(9) \qquad\qquad \sigma_n^2 \;\leq\; 2^{1+\nu}\, n\, K^2\, \Psi(1, 1)\, L_1.$$

Hence, taking $A_n = 2^{1+\nu} n K^2 \Psi(1, 1) L_1$ we obtain from (8) that

$$(10) \qquad\qquad \mathbb{P}\,(S_n \geq t) \;\leq\; \exp\left(-\frac{t^2}{C_1 n \,+\, C_2 t^{(2\mu+2\nu+3)/(\mu+\nu+2)}} \right),$$

where $C_1 = 2^{2+\nu} K^2 \Psi(1,1) L_1$ and $C_2 = 2 B_n^{1/(\mu+\nu+2)}$ with

$$B_n = 2(K \vee M) L_2 \Big(\frac{2^{3+\mu}}{\Psi(1,1)} \vee 1 \Big).$$

Inequality (10) is then more of Hoeffding-type.

(ii) Based on a Rosenthal-type inequality, [5] also proved an exponential inequality for S_n, however, with $\sqrt{t}$ instead of t^2 in the exponent. [3] proved a Bennett-type inequality for weakly dependent random variables. This also implies a Bernstein-type inequality, however, with different constants. In particular, the leading term in the denominator of the exponent differs from σ_n^2. This is a consequence of their method of proof which consists of replacing weakly dependent blocks of random variables by independent ones according to some coupling device (an analogous argument is used in [4] for the case of absolute regularity).

(iii) A Bernstein-type inequality with σ_n^2 as a possible leading term in the denominator of the exponent has been derived in [9] under a weak dependence condition which is tailor-made for causal processes with an exponential decay of the coefficients of weak dependence. The result above is more general and is also applicable to interesting classes of processes where Kallabis and Neumann's inequality does not apply.

A first Rosenthal-type inequality for weakly dependent random variables was derived by [5] via direct expansions of the moments of even order. Unfortunately, the variance of the sum did not explicitly show up in their bound. Instead, a rough bound for this expression based on upper estimates was used. Using cumulant bounds in conjunction with Leonov and Shiryaev's formula we are able to obtain a tighter moment inequality which resembles the Rosenthal inequality in the independent case (see [12] and [8] in the independent case, and Theorem 2.12 in [7] in the case of martingales).

Theorem 2.3. *Suppose that $X_1, \ldots, X_n$ are real-valued random variables with zero mean, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let p be a positive integer. We assume that there exist constants $K, M < \infty$, and a nonincreasing sequence of real coefficients $(\rho(n))_{n \geq 0}$ such that, for all u-tuples $(s_1, \ldots, s_u)$ and all v-tuples $(t_1, \ldots, t_v)$ with $1 \leq s_1 \leq \cdots \leq s_u \leq t_1 \leq \cdots \leq t_v \leq n$ and $u + v \leq p$,*

$$(11) \qquad |\text{cov}(X_{s_1} \cdots X_{s_u}, X_{t_1} \cdots X_{t_v})| \leq K^2 M^{u+v-2} \Psi(u,v) \rho(t_1 - s_u).$$

Furthermore, we assume that

$$\mathbb{E}|X_i|^{p-2} \leq M^{p-2}.$$

Then, with $Z \sim \mathcal{N}(0,1)$,

$$|\mathbb{E} S_n^p - \sigma_n^p \mathbb{E} Z^p| \leq B_{p,n} \sum_{1 \leq u < p/2} A_{u,p} K^{2u} (M \vee K)^{p-2u} n^u,$$

where $B_{p,n} = (p!)^2 2^p \max_{2 \le k \le p} \{\rho_{k,n}^{p/k}\}$, $\rho_{k,n} = \sum_{s=0}^{n-1} (s+1)^{k-2} \rho(s)$ *and*

$$A_{u,p} = \frac{1}{u!} \sum_{k_1 + \cdots + k_u = p, \, k_i \ge 2 \, \forall i} \frac{p!}{k_1! \cdots k_u!}.$$

Remark 2. *(i)* For even p, the above result implies that

$$\mathbb{E} S_n^p \le (p-1)(p-3) \cdots 1 \, \sigma_n^p + B_{p,n} \sum_{1 \le u < p/2} A_{u,p} \, K^{2u} \, (M \vee K)^{p-2u} \, n^u,$$

which resembles the classical Rosenthal inequality from the independent case. If $\sup_n B_{p,n} < \infty$ and $\sigma_n^2 \asymp n$, then the first term on the right-hand side is asymptotically dominating, as $n \to \infty$. This term is equal to the p-th moment of a Gaussian random variable with mean 0 and variance σ_n^2.

(ii) [5] also obtained a Rosenthal-type inequality, however, essentially with $n \cdot \sum_{k=-\infty}^{\infty} |\mathbb{E} X_0 X_k|$ instead of $\mathrm{var}(S_n)$ in the first term.

Remark 3. The inequality from Theorem 2.3 is well suited for proving a central limit theorem via the method of moments. Assume first that the random variables X_t are uniformly bounded, centered and satisfy condition (11) with $\lim_{s \to \infty} \rho(s)/s^p = 0$, for all $p > 0$. Then

$$\lim_{n \to \infty} \frac{\sigma_n^2}{n} = \sigma^2 = \sum_{k=-\infty}^{\infty} \mathbb{E} X_0 X_k$$

is a convergent series, and thus the method of moments implies the central limit theorem,

$$\frac{1}{\sqrt{n}} S_n \xrightarrow{d}_{n \to \infty} \sigma Z.$$

REFERENCES

[1] Bickel, P. J. and Bühlmann, P. (1999). A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli* **5**, 413–446.
[2] Coulon-Prieur, C. and Doukhan, P. (2000). A triangular central limit theorem under a new weak dependence condition. *Statist. Probab. Lett.* **47**, 61–68.
[3] Dedecker, J. and Prieur, C. (2004). Coupling for τ-dependent sequences and applications. *J. Theor. Probab.* **17**, 861–885.
[4] Doukhan, P. (1994). *Mixing: Properties and Examples.* Lecture Notes in Statistics **85**, Springer Verlag.
[5] Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and application to moment inequalities. *Stoch. Proc. Appl.* **84**, 313–342.
[6] Doukhan, P. and Neumann, M. H. (2007). Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stoch. Proc. Appl.* **117**, 878–903.
[7] Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications.* Academic Press, New York.
[8] Johnson, B. W., Schechtman, G. and Zinn, J. (1985). Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Probab.* **13**, 234–253.

[9] Kallabis, R. S. and Neumann, M. H. (2006). An exponential inequality under weak dependence. *Bernoulli* **12**, 333–350.

[10] Neumann, M. H. and Paparoditis, E. (2006). Goodness-of-fit tests for Markovian time series models. Manuscript.

[11] Rosenblatt, M. (1980). Linear processes and bispectra. *J. Appl. Probab.* **17**, 265–270.

[12] ROSENTHAL, H. P. (1970). On the subspaces of $L^p (p > 2)$ spanned by sequences of independent random variables. *Israel J. Math.* **8**, 273–303.

A statistical view on inverse problems

MARKUS REISS

Starting with typical inverse problems arising in econometrics like instrumental variables, functional linear regression or density deconvolution, we introduce the abstract notation of an inverse problem in a Gaussian white noise setting. Here, $f \in L^2(D_1)$ and $g \in L^2(D_2)$ satisfy the linear relation

$$g = Kf$$

with a bounded linear operator $K : L^2(D_1) \to L^2(D_2)$ and we observe

$$g_\varepsilon = g + \varepsilon \dot{W}, \ \varepsilon > 0, \ \dot{W} \text{Gaussian white noise}$$

The statistical problem is to estimate f nonparametrically based on g_ε.

To tackle this problem we review and discuss regularization and estimation methods from statistics and numerical analysis, in particular:

- Denoising in the image domain
- Singular value decomposition (SVD)
- Tikhonov method and related iterative methods
- Projection methods like Galerkin and least squares methods

Besides different behaviour in the implementations and possible difficulties in the case of noisy operators we focus on the problem of the function classes for which the methods are well designed. In the case of the SVD and Tikhonov's method these classes should have good smoothness properties with respect to the singular functions of the operator (Hilbert scale approach). An intriguing example is that of circular deconvolution where the eigenfunctions are given by the Fourier basis and a smooth, but non-periodic function does have smothness less than $1/2$ with respect to this basis. On the other hand, denoising in the image domain and projection methods usually require classical smoothness assumptions for the function class and the operator considered. A general conclusion is that the method for solving the inverse problem should be chosen according to expected properties of the function f to be estimated. It might be worthwhile to consider different methods and to apply statistical methods to select the best estimator among all those obtained.

References

[1] X. Chen and M. Reiß, *On rate optimality for ill-posed inverse problems in econometrics*, Preprint (2007).
[2] H. Engl, M. Hanke and A. Neubauer, *Regularisation of inverse problems*, Kluwer Academic Press (2000).

Deconvolution with unknown error distribution

Jan Johannes

Let X and ϵ be independent random variables with unknown density functions f_X and f_ϵ, respectively. The objective is to estimate nonparametrically the density function f_X and its derivatives based on a sample of $Y = X + \epsilon$. In this setting the density f_Y of Y is the convolution of the density of interest f_X and the density f_ϵ of the additive noise, i.e.,

$$(1) \qquad f_Y(y) = f_X \star f_\epsilon(y) := \int_{-\infty}^{\infty} f_X(x) f_\epsilon(y - x) dx.$$

Suppose we observe $Y_1, \ldots, Y_n$ from f_Y and the error density f_ϵ is known. Then the estimation of the convolution density f_X is a classical problem in statistics. The most popular approach is to estimate f_Y by a kernel estimator and then to solve equation (1) using a Fourier transform (c.f. [1] and [2]). It is well-known that solving equation (1) leads to an ill-posed inverse problem and, hence equation (1) has to be 'regularized' in some way in order to obtain a consistent estimator. The rate of convergence of the deconvolution problem is determined by the tail-behavior of the Fourier transforms $\mathcal{F}f_X$ and $\mathcal{F}f_\epsilon$ of f_X and f_ϵ, respectively. [3] derive the minimax rate of convergence when the density f_X lies in the well-known Sobolev space $H_{\mathfrak{p}}$, which describes the level of smoothness of a deconvolution density in terms of its Fourier transform $\mathcal{F}f_X$. They consider the two cases, where the error distribution of ϵ is supersmooth, that is the Fourier transform $\mathcal{F}f_\epsilon$ of f_ϵ has exponential descent, i.e., $|\mathcal{F}f_\epsilon(t)|^2 \sim \exp(-t^{2\mathfrak{a}})$, and the other extreme when the Fourier transform of the error density has polynomial descent, i.e., $|\mathcal{F}f_\epsilon(t)|^2 \sim t^{-2\mathfrak{a}}$. Roughly speaking, in the first case the optimal rate of convergence of the mean integrated squared error (MISE) in a minimax sense is of order $O(\log(n)^{-\mathfrak{p}/\mathfrak{a}})$, while in the second case we have $O(n^{-2\mathfrak{p}/(2(\mathfrak{p}+\mathfrak{a})+1)})$. However, we will show in this paper, that the rate of convergence is not determined by the tail behavior of $\mathcal{F}f_\epsilon$ but by the tail behavior of the ratio of $\mathcal{F}f_X$ and $\mathcal{F}f_\epsilon$.

The present paper deals with the estimation of a deconvolution density f_X if in addition the density f_ϵ of the noise is unknown. In this case without any additional information the density f_X can not be recovered from the density of f_Y through (1), i.e., the density f_X is not identified assuming only a sample $Y_1, \ldots, Y_n$ from f_Y. However, sometimes draws of the error distribution are observed. Thus, we assume that we observe the sample $Y_1, \ldots, Y_n$ from f_Y and additionally the sample $\epsilon_i, i = 1, \ldots, m$ from f_ϵ. In such a situation it is of interest to study the sampling properties, in particular the mean squared error, of the estimator when we use an estimator of f_ϵ, rather than the true density. It is interesting to note that

[4] proposes an estimator of f_X and derives its optimal rate of convergence in a minimax sense for a class of densities with Fourier transform having a polynomial decay. The main purpose of this paper is to propose and study a deconvolution scheme which has enough flexibility to allow estimation when the tails of $\mathcal{F}f_X$ and $\mathcal{F}f_\varepsilon$ have a wide range of behaviors.

The estimators proposed in this paper are based on a regularized inversion of (1) using a spectral cut-off (thresholding of the characteristic function of the errors), where we replace the unknown densities f_Y and f_ε by nonparametric estimators. Given the error density f_ϵ is known we show that the estimators based on a spectral cut-off and a nonparametric estimator of f_Y are asymptotically optimal. It is of interest to compare the MISE rates when the density of f_ϵ is estimated with the optimal rates, where f_ϵ is known. We show that the rate of convergence when f_ϵ is unknown depends on both, the sample size n (of observations Y) and also on the errors m. In fact we show that if m grows with n at a sufficiently fast rate, then the error due to the estimation of f_ϵ is asymptotically negligible. The rate is determined by the smoothness of f_ϵ and f_X. For example, if $f_X \in H_{\mathfrak{p}}$ and the Fourier transform $\mathcal{F}f_\epsilon$ has exponential descent, i.e., $|\mathcal{F}f_\epsilon(t)|^2 \sim \exp(-t^{2\mathfrak{a}})$, and $m \approx n^\nu$ for an arbitrary $\nu > 0$, then the rate of convergence of the MISE is of order $O(\log(n)^{-\mathfrak{p}/\mathfrak{a}})$. Therefore, if the sample size m tends to ∞ as a polynomial growth of n, then estimation of the error density does not influence the rate of convergence of the MISE. This leads to the rather surprising result that for normally distributed errors the MISE is, mainly, unaffected by using an estimator of the density rather than the true density. The situation is different if the Fourier transform $\mathcal{F}f_\epsilon$ has polynomial descent, i.e., $|\mathcal{F}f_\epsilon(t)|^2 \sim t^{-2\mathfrak{a}}$, then the optimal rate of convergence of the MISE is of order $O(n^{-2\mathfrak{p}/(2(\mathfrak{p}+\mathfrak{a})+1)})$ provided that $m \approx n^{2(\mathfrak{p}\vee\mathfrak{a})/(2(\mathfrak{p}+\mathfrak{a})+1)}$. Therefore, the smoother the error density, i.e., the larger the value $\mathfrak{a}$, the smaller the necessary sample size m has to be to imply that the proposed estimator has the optimal rate $O(n^{-2\mathfrak{p}/(2(\mathfrak{p}+\mathfrak{a})+1)})$ (where f_ε is known). In contrast, by studying the optimal rate $O(n^{-2\mathfrak{p}/(2(\mathfrak{p}+\mathfrak{a})+1)})$, we see that the rate decreases as $\mathfrak{a}$ is increasing. Conversely, if f_ε is known, then the rate of convergence of the MISE is fast if $\mathfrak{p}$ is large. However, in the case that f_ε is estimated, the same is only true if the sample m is sufficiently large.

One of the main achievements in this paper is the derivation of the MISE of the proposed estimator for a general class of density functions, which unifies and generalizes many of the previous results for known and unknown error distributions. Roughly speaking, we show that the MISE of the proposed estimator can be decomposed essentially into a function of the MISE of the nonparametric density estimator of f_Y plus an additional bias term which depends on the relationship between the tails of $\mathcal{F}f_X$ and $\mathcal{F}f_\epsilon$. Therefore by balancing the bias and variance we are able to obtain the optimal bandwidth. The relationship between $\mathcal{F}f_X$ and $\mathcal{F}f_\epsilon$ essentially determines the bias of the estimator. Returning to the example above, where the error distribution is supersmooth (e.g. in case of a normal distribution) and $\mathcal{F}f_X$ descents polynomial (e.g. in case of a double exponential distribution),

the bias is a logarithm of the smoothing parameter (the parameter which determines the spectral cut off point). On the other hand, if both, the error distribution and X are supersmooth, the bias is a polynomial of the smoothing parameter (c.f. [6]). We show that the theory behind these rates are unified through the 'link function' κ, which 'links' the tail behavior of $\mathcal{F}f_X$ and $\mathcal{F}f_\epsilon$, that is for large t, $|\mathcal{F}f_X(t)|^2 \sim \kappa(|\mathcal{F}f_\varepsilon(t)|^2)^\beta$. This link function determines the bias. For example, if the error distribution is supersmooth (e.g. in case of a normal distribution) and the $\mathcal{F}f_\epsilon$ descents polynomial (e.g. in case of a double exponential distribution), the 'link function' is $\kappa(t) = |\log(t)|^{-1}$, whereas if both the error distribution and X are supersmooth then the link function is $h(t) = t$. We mention that in this paper we use the classical Rosenblatt-Parzen kernel estimator (c.f. [5]) and the empirical characteristic function to estimate the densities f_Y and f_ε, respectively. Therefore, the kernel function does not need to have a compact support. However, since the MISE of the proposed estimator can be decomposed into the MISE of the density estimator of f_Y, any other nonparametric estimation method (e.g. based on Splines or Wavelets) can be used and the theory still holds.

We note that if there is a-priori knowledge concerning the smoothness of f_X characterized by $f_X \in H_{\mathfrak{p}}$ for some $\mathfrak{a} > 0$, then we may use a similar scheme to estimate the derivatives of f_X. Furthermore, similar MISE can be derived.

<div align="center">REFERENCES</div>

[1] R. J. Carroll and P. Hall. *Optimal rates of convergence for deconvolving a density*, Am. Statist. Assoc., **83** (1988), 1184–1186.
[2] J. Fan *On the optimal rates of convergence for nonparametric deconvolution problems.*, Ann. Stat., **19** (1991), 1257–1272.
[3] B.A. Mair and F.H. Ruymgaart. *Statistical inverse estimation in Hilbert scales.* SIAM J. Appl. Math., **56(5)** (1996), 1424–1444.
[4] M.H. Neumann. *On the effect of estimating the error density in nonparametric deconvolution.*, Nonparametric Statistics, **7** (1997), 307–330.
[5] E. Parzen. *On Estimation of a Probability Density Function and Mode.* Ann. Math. Stat., **33(3)** (1962), 1065–1076.
[6] M. Pensky and B. Vidakovic. *Adaptive wavelet estimator for nonparametric density deconvolution.*, Ann. Stat., **27** (1999), 2033–2053.

<div align="center">**Statistical Inference for Deconvolution**</div>

<div align="center">HAJO HOLZMANN</div>

In the talk we discuss inference for nonparametric density deconvolution, as arising in the general context of statistical inverse problems.

Let $X_1, \ldots, X_n$ be i.i.d. real-valued observations with density g. Nonparametric estimation of g by kernel methods was introduced by Rosenblatt (1956) and Parzen (1962). Often, the observations X_i are only noisy versions of the random variables Z_i of interest, i.e. $X_i = Z_i + \epsilon_i$, where ϵ_i and Z_i are independent, the errors ϵ_i have known density ψ and the Z_i have density f. Note that $g = f * \psi$. Estimating f from the observations X_i is therefore called the deconvolution problem.

Under the assumption $\Phi_\psi(t) \neq 0$ for all $t \in \mathbb{R}$, a standard estimator of f is the kernel deconvolution density estimator

$$\hat{f}_n(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \Phi_K(ht) \frac{\hat{\Phi}_n(t)}{\Phi_\psi(t)} \, dt,$$

where $\Phi_f(t)$ is the Fourier transform of f, K is a kernel function such that Φ_K has compact support, $h > 0$ is a smoothing parameter called bandwidth and $\hat{\Phi}_n(t) = 1/n \sum_k e^{itX_k}$ is the empirical characteristic function of $X_1, \ldots, X_n$.
It is well-known that the deconvolution problem depends sensitively on the Fourier transform Φ_ψ of the error density ψ. If

$$\Phi_\psi(t) \sim C_\psi \, t^{-\beta}, \quad t \to \infty,$$

for some $\beta > 0$ and $C_\psi \in \mathbb{C}$, the error density is called ordinary smooth, whereas if

(1) $$\Phi_\psi(t) \sim C_\psi |t|^{\lambda_0} e^{-|t|^\lambda/\mu}, \qquad |t| \to \infty,$$

for $\lambda > 1$, $\mu > 0$ and $\lambda_0, C_\psi \in \mathbb{R}$, and that $\Phi_\psi(t) \neq 0$ for all t, ψ belongs to the class of supersmooth densities (condition (1) does not cover all supersmooth densities). For ordinary smooth error densities, under regularity conditions listed in [1] we construct asymptotic level-α confidence bands of the form

(2) $$\hat{f}_n(t) - b_n(t,x) \leq f(t) \leq \hat{f}_n(t) + b_n(t,x), \qquad t \in [0,1],$$

where

$$b_n(t,\alpha) = \left(\frac{\hat{g}_n(t) C_{K,1}}{nh^{2(\beta+j)+1}} \right)^{1/2} \left(\frac{x_\alpha}{(2\log(1/h))^{1/2}} + d_n \right),$$

and x_α is the $(1-\alpha)$-quantile of the extreme-value distribution $\exp(-2e^{-x})$ and

$$d_n = \left(2\log(1/h) \right)^{1/2} + \frac{\log\left(\frac{1}{2\pi} C_{K,2}^{1/2} \right)}{\left(2\log(1/h) \right)^{1/2}}.$$

Further, we show convergence of the nonparametric n-n bootstrap, present simulation results and give an astrophysical example. Details can be found in [1].

We also study the asymptotic distribution of the statistic T_n, defined by

$$T_n = \int_{\mathbb{R}} \left(\hat{f}_n - K_h * f \right)^2(x) \, dx,$$

which is closely related to the integrated squared error of $\hat{f}_n$, and which can be used to test hypotheses of the form (extensions to composite hypotheses are possible)

$$H_0 : f = f_0.$$

In [2] it is shown that for the ordinary smooth case, T_n is asymptotically normally distributed as follows

$$nh^{2\beta+1/2} \left(T_n - \frac{C_\psi^2 C_{K,1}}{2\pi \, nh^{2\beta+1}} \right) \xrightarrow{\mathcal{L}} N\left(0, \frac{C_\psi^4 C_{K,2} \|g_0\|^2}{\pi} \right).$$

In contrast, for supersmooth error densities of the form (1) with some weak regularity conditions, we show in [3] that if the Fourier transform Φ_K of the kernel K is real-valued, symmetric and supported on $[-1, 1]$, and if $\Phi_K(0) = 1$ and there exist $A > 0$, $\alpha \geq 0$ such that $\Phi_K(1 - t) = At^\alpha + o(t^\alpha)$, $t \searrow 0$, then

$$(3) \qquad \frac{(2\lambda)^{1+2\alpha}\pi C_\psi^2\, n}{A^2 \mu^{1+2\alpha} h^{\lambda-1+2\lambda\alpha+2\lambda_0} \exp(\frac{2}{\mu h^\lambda})\Gamma(2\alpha+1)} T_n \xrightarrow{\mathcal{L}} (Y_1^2 + Y_2^2)/2,$$

where Y_1 and Y_2 are independent standard normal random variables.

In [2], we also derive the asymptotic distribution of T_n under fixed alternatives to the hypothesis H_0, which has applications to model validation.

REFERENCES

[1] N. Bissantz, L. Dümbgen, H. Holzmann and A. Munk, *Nonparametric confidence bands in deconvolution density estimation.* to appear: J. Royal Statistical Society Series B (2007).
[2] H. Holzmann, N. Bissantz and A. Munk, *Density testing in a contaminated sample.* J. Multivariate Analysis, **98** (2007), 57–75.
[3] H. Holzmann and L. Boysen, *Integrated square error asymptotics for supersmooth deconvolution.* Scandinavian J. Statist, **33** (2006), 849–860.

Smoothing Splines Estimators for Functional Linear Regression
ALOIS KNEIP

We consider a regression problem in which the variation of scalar responses Y_i is explained by functions $X_i(t), t \in I$, square integrable on the compact interval I of $\mathbb{R}$, $i = 1, \ldots, n$. More precisely, we investigate *functional linear regression models* of the form

$$Y_i = \int_I \alpha(t)X_i(t)dt + \epsilon_i, \qquad i = 1, \ldots, n,$$

where ϵ_i's are i.i.d. centered random errors, $\mathbb{E}(\epsilon_i) = 0$, with variance $\mathbb{E}(\epsilon_i^2) = \sigma_\epsilon^2$ and α is a square integrable functional parameter defined on I that must be estimated from the pairs (X_i, Y_i), $i = 1, \ldots, n$. $X_1, \ldots, X_n$ is a sequence of identically distributed random functions with the same distribution as X. The main assumption on X is that it is a second order variable *i.e.* $\mathbb{E}(\int_I X^2(t)dt) < +\infty$ and it is assumed moreover that $\mathbb{E}(X_i(t)\epsilon_i) = 0$ for almost every $t \in I$.

As a consequence of developments of modern technology data that may be described by functional regression models can be found in a lot of fields such as economics, medicine, linguistics, or chemometrics. An example is the application motivating our study: the data consists in repeated measurements over the day of pollutant indicators in the area of Toulouse used to explain the maximum (peak) of pollution for the next day. In practice, the whole curves X_i are usually not available, but instead are observed in p discretization points $t_1 < \ldots < t_p$ belonging to I.

The problem of estimating the functional slope parameter α belongs to a class of ill-posed inverse problems. Any sensible procedure for estimating α (or more

precisely of its identifiable part) has to involve regularization procedures. We propose an estimation procedure that can be seen as a generalization of the well-known smoothing splines method in nonparametric regression.

Based on the observation times $t_1 < \ldots < t_p$, some prespecified $m = 1, 2, \ldots$ and a smoothing parameter $\rho > 0$, our estimate $\hat{\alpha}$ of α is determined by minimizing

$$\frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \frac{1}{p} \sum_{j=1}^{p} a(t_j) X_i(t_j) \right)^2 + \rho \left(\frac{1}{p} \sum_{j=1}^{p} \pi_a(t_j)^2 + \int_I a^{(m)}(t)^2 dt \right),$$

over all functions a in the Sobolev space $W^{m,2}(I) \subset L^2(I)$, where

$$\pi_a(t) = \sum_{l=1}^{m} \beta_{a,l} t^{l-1}$$

with $\sum_{j=1}^{p} (a(t_j) - \pi_a(t_j))^2 = \min_{\beta_1, \ldots, \beta_m} \sum_{j=1}^{p} \left(a(t_j) - \sum_{l=1}^{m} \beta_l t^{l-1} \right)^2$.

A detailed asymptotic theory of the estimator $\hat{\alpha}$ is developed for large values of n and p. It is assumed that p is sufficiently large compared to n so that the discretization error is negligible. Motivated by our application, we focus on the error in predicting $\int_I \alpha(t) X(t) dt$ by $\int_I \hat{\alpha}(t) X(t) dt$. This is formalized by evaluating the distance between $\hat{\alpha}$ and α with respect to an L^2 semi-norm induced by the covariance operator Γ of X, $\|u\|_\Gamma = \langle \Gamma u, u \rangle$ with $\langle u, v \rangle = \int_I u(t) v(t) dt$. Note that

$$\mathbb{E} \left([\int_I \hat{\alpha}(t) X_{n+1}(t) dt - \int_I \alpha(t) X_{n+1}(t) dt]^2 \mid \hat{\alpha} \right) = \|\hat{\alpha} - \alpha\|_\Gamma^2$$

for any random function X_{n+1} possessing the same distribution as X and independent of $X_1, \ldots, X_n$. Moreover, by using these semi-norms we explicitly concentrate on evaluating the estimation error only for the identifiable part of the structure of α.

Rates of convergence of $\|\hat{\alpha} - \alpha\|_\Gamma^2$ then depend on the degree of smoothness of α and on the rate of decrease of the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$ of Γ. More precisely, if α is m times differentiable, $\alpha^{(m)}$ belongs to $L^2(I)$, and if $\lambda_j \leq C \cdot j^{-q}$ for some $0 < C < \infty$ and $q > 0$, then under some additional regularity conditions it can be shown that

$$\|\hat{\alpha} - \alpha\|_\Gamma^2 = O_P(n^{-(2m+q)/(2m+q+1)}).$$

We then prove that these rates are optimal over large classes of distributions for the predictive curves and functions α belonging to suitable Sobolev spaces.

The above results have been obtained by joint work with C. Crambes and P. Sarda, Université Paul Sabatier, Toulouse. A detailed description of the conceptual approach, asymptotic theory and a real data application can be found in Crambes, Kneip and Sarda (2007)

References

[1] C. Crambes, A. Kneip and P. Sarda, *Smoothing Splines Estimators for Functional Linear Regression*, Manuscript (2007).

Estimating linear functionals of nonparametric regression models with endogenous regressors

GAUTAM TRIPATHI

(joint work with Thomas A. Severini)

Models containing unknown functions, typically characterized as conditional expectations, are common in economics and economists are often interested in estimating linear functionals of these unknown functions; e.g., [12] estimates the contrast between functionals of $\mathbb{E}[Y|X]$ using before-and-after policy intervention data; letting Y denote the market demand and X the price, [9] consider estimating $\int_a^b \mathbb{E}[Y|X=x]\,dx$, the approximate change in consumer surplus for a given price change; additional examples can be found in [5] and [2, 2005b].

However, in models where variables are determined endogenously, unknown functions cannot always be interpreted as conditional expectations which complicates the problem of estimating their linear functionals. For instance, market demand functions are not identifiable as conditional expectations because prices are endogenous. Hence, simply integrating an estimator of the conditional expectation of equilibrium quantity given equilibrium price over a certain interval will not lead to a consistent estimator of the change in consumer surplus.

The basic objective of this paper is to investigate whether certain linear functionals of unknown structural functions can be efficiently estimated with parametric rates of convergence even when the underlying structural function itself is not a conditional expectation.

Consider the nonparametric regression model

(1) $Y = \mu^*(X) + \varepsilon, \qquad \mathbb{E}[\varepsilon|W] = 0 \text{ w.p.1},$

where X is a vector of regressors some or all of which are endogenous and W denotes the vector of instrumental variables (IV's); since exogenous explanatory variables act as their own instruments, W and X can have elements in common. The functional form of μ^* is unknown; we only assume that it lies in $L_2(X)$, the set of real-valued functions of X that are square integrable with respect to the distribution of X. Endogeneity of regressors means that μ^* cannot be a conditional expectation function because W does not contain all of X; of course, if $W = X$ so that there are no endogenous regressors, then $\mu^*(X) = \mathbb{E}[Y|X]$.

Even if the structural parameter μ^* in (1) is identified, i.e., uniquely defined, it is said to be "ill-posed" because the function that maps the data to μ^* is not continuous; see Lemma 2.4 of [11] for additional properties of this mapping. Although μ^* may be ill-posed and hence difficult to estimate, we study whether its functionals $\mathbb{E}[\psi(X)\mu^*(X)]$ and $\int_{\mathrm{supp}(X)} \psi(x)\mu^*(x)\,dx$, where ψ is a known weight function and $\mathrm{supp}(X)$ the support of X, can be estimated with parametric or $n^{1/2}$-rates of convergence.[1]

---

[1]For $\int_{\mathrm{supp}(X)} \psi(x)\mu^*(x)\,dx$ to make sense it is implicitly understood that X is continuously distributed; the expectation functional $\mathbb{E}[\psi(X)\mu^*(X)]$ is of course well defined even when some components of X are discrete.

In addition to the papers cited earlier, recent works on nonparametric IV methods include, e.g., [6], [1], [4], [10], [7], [8], and the references therein. Our main contribution to this literature is to derive variational and non-variational, i.e., closed form, expressions for the efficiency bounds for estimating $\mathbb{E}[\psi(X)\mu^*(X)]$ and $\int_{\mathrm{supp}(X)} \psi(x)\mu^*(x)\,dx$ without assuming that μ^* is well-posed or even identified. We also conjecture that plug-in estimators of these functionals may be asymptotically efficient (assuming μ^* is identified), although we haven't been able to derive their asymptotic distribution.

References

[1] C. Ai and X. Chen, *Efficient estimation of models with conditional moment restrictions containing unknown functions,* Econometrica, **71** (2003), 1795–1843.

[2] C. Ai and X. Chen, *Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables,* Manuscript (2005a).

[3] C. Ai and X. Chen, *On efficient sequential estimation of semi-nonparametric moment models,* Manuscript (2005b).

[4] R. Blundell and J.L. Powell, *Endogeneity in nonparametric and semiparametric regression models,* in *Advances in economics and econometrics: Theory and applications*, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky, Cambridge University Press, **2** (2003), 312–357.

[5] B.W. Brown and W.K. Newey, *Efficient semiparametric estimation of expectations,* Econometrica, **66** (1998), 453–464.

[6] S. Darolles, J.-P. Florens and E. Renault, *Nonparametric instrumental regression,* Manuscript (2002).

[7] J.-P. Florens, J. Johannes and S. van Bellegem, *Instrumental regression in partially linear models,* Manuscript (2005).

[8] P. Hall and J. L. Horowitz, *Nonparametric methods for inference in the presence of instrumental variables,* Annals of Statisics, **33** (2005), 2904–2929.

[9] W.K. Newey and D. McFadden, *Large sample estimation and hypothesis testing,* in *Handbook of Econometrics, vol. IV*, ed. by R. Engle and D. McFadden, Elsevier Science B.V. (1994), 2111–2245.

[10] W.K. Newey and J. L. Powell, *Instrumental variables estimation of nonparametric models,* Econometrica, **71** (2003), 1557–1569.

[11] T.A. Severini and G. Tripathi, *Some identification issues in nonparametric linear models with endogenous regressors,* Econometric Theory, **22** (2006), 258–278.

[12] J.H. Stock, *Nonparametric Policy Analysis,* Journal of the American Statistical Association, **84** (1989), 567–575.

Reporter: Enno Mammen

Participants

Prof. Dr. Yacine Ait-Sahalia
Bendheim Center for Finance
Princeton University
26 Prospect Avenue
Princeton NJ 08540-5296
USA

Dante D. Amengual
Bendheim Center for Finance
Princeton University
26 Prospect Avenue
Princeton NJ 08540-5296
USA

Dr. Denis Belomestny
Weierstraß-Institut für
Angewandte Analysis und Stochastik
im Forschungsverbund Berlin e.V.
Mohrenstr. 39
10117 Berlin

Prof. Dr. Gerard J. van den Berg
Department of Econometrics
Vrije University
De Boelelaan 1105
NL-1081 HV Amsterdam

Federico A. Bugni
Department of Economics
Northwestern University
Evanston IL 60208-2600
USA

Dr. Peter Bühlmann
Seminar für Statistik
ETH-Zürich
LEO C 17
CH-8092 Zürich

Julio A. Cacho-Diaz
Bendheim Center for Finance
Princeton University
26 Prospect Avenue
Princeton NJ 08540-5296
USA

Prof. Dr. Raymond Carroll
Department of Statistics
Texas A & M University
College Station , TX 77843-3143
USA

Prof. Dr. Xiaohong Chen
Department of Economics
New York University
269 Mercer Street
New York , NY 10003
USA

Prof. Dr. Andrew Chesher
Department of Economics
University College London
Gower Street
GB-London WC1E 6BT

Prof. Dr. Rainer Dahlhaus
Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 294
69120 Heidelberg

Prof. Dr. Holger Dette
Fakultät für Mathematik
Ruhr-Universität Bochum
Universitätsstr. 150
44801 Bochum

Prof. Dr. Bernd Fitzenberger
FB Wirtschaftswissenschaften
Universität Frankfurt
Mertonstr. 17 - 25
60325 Frankfurt

Prof. Dr. Jean-Pierre Florens
Universite Toulouse 1
Science Sociales
Place Anatole
F-31042 Toulouse Cedex

Prof. Dr. Jürgen Franke
Fachbereich Mathematik
T.U. Kaiserslautern
Erwin-Schrödinger-Straße
67653 Kaiserslautern

Prof. Dr. Sara van de Geer
Seminar für Statistik
ETH-Zentrum Zürich
LEO D2
Leonhardstr. 27
CH-8092 Zürich

Prof. Dr. Irene Gijbels
Department of Mathematics and
Center for Statistics
Katolieke Universiteit Leuven
W. de Croylaan 54
B-3001 Leuven (Heverlee)

Prof. Dr. Wolfgang Härdle
Wirtschaftswissenschaftl. Fakultät
Lehrstuhl für Statistik
Humboldt-Universität Berlin
Spandauer Str. 1
10178 Berlin

Prof. Dr. Hong Han
Department of Economics
Duke University
Social Sciences Building
Box 90097
Durham NC 27708-0097
USA

Prof. Dr. Marc Henry
Economics Department
Columbia University
1026 International Affairs Building
420 West 118th Street
New York , NY 10027
USA

Prof. Dr. Stefan Hoderlein
Abteilung f. Volkswirtschaftslehre
Universität Mannheim
L 7, 3-5
68131 Mannheim

Dr. Hajo Holzmann
Institut f. Mathemat. Stochastik
Georg-August-Universität Göttingen
Maschmühlenweg 8-10
37073 Göttingen

Prof. Dr. Joel Horowitz
Northwestern University
2001 Sheridan Road
Evanston , IL 60208-2600
USA

Prof. Dr. Hidehiko Ichimura
Faculty of Economics
University of Tokyo
Bunkyo-ku
Tokyo 113
JAPAN

Dr. Jan Johannes
Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 294
69120 Heidelberg

Ilze Kalnina
Department of Economics
London School of Economics
Houghton Street
GB-London WC2A 2AE

Dr. Alois R. Kneip
Institut für Gesellschafts-
und Wirtschaftswissenschaften
Universität Bonn
Adenauerallee 24-42
53113 Bonn

Prof. Dr. Jens-Peter Kreiß
Institut für Mathematische
Stochastik der TU Braunschweig
Pockelsstr. 14
38106 Braunschweig

Prof. Dr. Dennis Kristensen
Economics Department
Columbia University
1026 International Affairs Building
420 West 118th Street
New York , NY 10027
USA

Prof. Dr. Sokbae Lee
Department of Economics
University College London
Gower Street
GB-London WC1E 6BT

Prof. Dr. Oliver Linton
Dept. of Economics
London School of Economics
and Political Science
Houghton Street
GB-London WC2A 2AE

Prof. Dr. Enno Mammen
Abteilung f. Volkswirtschaftslehre
Universität Mannheim
L 7, 3-5
68131 Mannheim

Dr. Yukitoshi Matsushita
Faculty of Economics
University of Tokyo
Bunkyo-ku
Tokyo 113
JAPAN

Prof. Dr. Rosa L. Matzkin
Northwestern University
2001 Sheridan Road
Evanston , IL 60208-2600
USA

Prof. Dr. Michael Helmut Neumann
Fakultät für Mathematik und
Informatik
Friedrich-Schiller-Universität
Ernst-Abbe-Platz 1-4
07743 Jena

Prof. Dr. Whitney K. Newey
Dept. of Economics
Massachusetts Institute of
Technology
50 Memorial Drive
Cambridge , MA 02142-1347
USA

Dr. Jens Perch Nielsen
Royal & SunAlliance
Gammel Kongevej 60
DK-1790 Copenhagen

Dr. Mark Podolskij
Fakultät für Mathematik
Ruhr-Universität Bochum
Universitätsstr. 150
44801 Bochum

Prof. Dr. Markus Reiß
Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 294
69120 Heidelberg

Prof. Dr. Peter M. Robinson
Dept. of Economics
London School of Economics
and Political Science
Houghton Street
GB-London WC2A 2AE

Christoph Rothe
Institut für Statistik
Universität Mannheim
68131 Mannheim

Melanie Schienle
Abteilung f. Volkswirtschaftslehre
Universität Mannheim
L 7, 3-5
68131 Mannheim

Prof. Dr. Leopold Simar
Institut de Statistique
Universite Catholique de Louvain
Voie du Roman Pays, 20
B-1348 Louvain-La-Neuve

Prof. Dr. Stefan Sperlich
Institut für Statistik und
Ökonometrie
Universität Göttingen
Platz der Göttinger Sieben 5
37073 Göttingen

Prof. Dr. Vladimir Spokoiny
Weierstrass-Institute for Applied
Analysis and Stochastics
Humboldt-University Berlin
Mohrenstr. 39
10117 Berlin

Viktor Subbotin
Department of Economics
Northwestern University
Evanston IL 60208-2600
USA

Supachoke Thawornkaiwong
Dept. of Economics
London School of Economics
and Political Science
Houghton Street
GB-London WC2A 2AE

Prof. Dr. Gautam Tripathi
Department of Economics
University of Connecticut-Storrs
341 Mansfield Road, Unit 1063
Storrs , CT 06269-1063
USA

Prof. Dr. Aad W. van der Vaart
Faculteit Wiskunde en Informatica
Vrije Universiteit Amsterdam
De Boelelaan 1081 a
NL-1081 HV Amsterdam

Prof. Dr. Ingrid Van Keilegom
Institut de Statistique
Universite Catholique de Louvain
Voie du Roman Pays 20
B-1348 Louvain-la-Neuve

Prof. Dr. Edward Vytlacil
Economics Department
Columbia University
1026 International Affairs Building
420 West 118th Street
New York , NY 10027
USA

Prof. Dr. Yazhen Wang
Department of Statistics
University of Connecticut
Box U-4120
Storrs , CT 06269-3120
USA

Prof. Dr. Jon A. Wellner
Department of Statistics
University of Washington
Box 35 43 22
Seattle , WA 98195-4322
USA