University of Illinois
Spring 1998

Department of Economics
Roger Koenker

Economics 478

Topics in Semiparametric Econometrics

This course is a research seminar in semiparametric methods in econometrics. The objective of the course from a student viewpoint should be to write a research paper of potentially publishable quality, either as an exploratory venture directed toward finding a thesis topic, or as an opportunity to explore a new area in some depth just for the fun of it.

The course will be organized as an informal seminar. There will be no text *per se*, but there will be a considerable range of readings from the current literature. There will be no exams only a required term paper, and there will be periodic student presentations of course material. The topics covered in the course can be tailored somewhat to student demand. There are a few topics in which I have a current research interest, and for which I think there might be some scope for constructive interaction in a course of this type. I will try to encourage toward these topics in an effort to increase effective collaboration among students in the course. Not surprisingly, many of these topics have some connection to quantile regression, although in some cases this connection is rather speculative.

In the first class I will try to describe and motivate the topics I have identified and provide some basic bibliography.

In the second week I would like to get down to business by discussing a group of recent papers on estimating treatment effects. This is a topic on which there is considerable recent activity by both econometricians and statisticians and I believe that it is worthwhile for the entire class to take a close look at a few of these papers. I have prepared a xeroxed collection of papers which will be distributed through Up Close, one of the campus copy centers.

In the third week I would like to reconsider some classical readings on instrumental variable methods focusing on the connection between LIML canonical correlation and some related eigenvalue problems. The objective will be to begin to forge a link between the treatment effect discussion of the previous week and this older literature which would eventually result in a more refined approach to IV methods for quantile regression.

In the fourth week I will discuss some recent developments on numerical methods for quantile regression, and raise some questions about my ongoing project to revise the software distribution for Splus for quantile regression. This discussion is likely to focus on extensions of existing methods for inference.

In subsequent weeks there will be a sequence of mini-lectures by students to report progress on their research. Since the scope for these presentations will be rather severely constrained by the size of the class, I am investigating schemes which would enable students to "present" research results on lab web pages. This would enable us to share bibliography, software and other materials in an efficient and timely manner. I hope to provide some tutorial material on this in the first weeks of class. to provide some tutorial material on this in the first weeks of class. Some basic material related to the course, including this document, is available on the web at `http://www.econ.uiuc.edu/ roger/courses/478/478.html`. There will also be an expanded version of the bibliography which appears below which will be available on the web. One of the byproducts of the course will be an expanded version of this bibliography.

I would strongly recommend that you to begin to learn LaTeXand Splus as part of the research process. Some (hopefully helpful) hints on this are available in my paper "Reproducible Econometrics" which I will distribute in class. The book by (Venables and Ripley 1997) is highly recommended for Splus. (Lamport 1994) is the definitive introduction to LaTeX.

1. **Endogoneity, Sample Selection, and Treatment Effects**
   There is a strongly revived interest in the statistical fundementals of estimating treatment effect models in observational and experimental studies exemplified by recent work by (Angrist, Imbens, and Rubin 1996, Abadie, Angrist, and Imbens 1997, Heckman 1996, Imbens and Angrist 1994, Manski 1997). This will be our first topic and I think that it will provide a useful fresh look at the basic idea of instrumental variables which is one of the few, really original, ideas in econometrics. After reading some of these new papers, I would like to take a week to revisit some of the classical literature on IV's, LIML and related topics from the 1950's. This literature will, I hope, lead us into a deeper consideration of how to apply these ideas to the next topic.
2. **Open Problems in Quantile Regression**
   I would like to devote some time to what I regard as important open research problems in quantile regression. These might include some of the following:
   - Computational methods for quantile regression (Portnoy and Koenker 1997, Koenker 1997, Portnoy 1997, Fitzenberger 1997, Koenker and Park 1996)
   - Panel data and random effects models (Chay 1995, Arias, Hallock, and Sosa 1997, Rocke 1983, Markatou and Horowitz 1995)
   - Instrumental variable methods and l1 canonical correlation (Amemiya 1982, Powell 1983, Ammann 1993, Arias, Hallock, and Sosa 1997)
   - Influence robust quantile regression (Rousseeuw and Hubert 1996, He and Portnoy 1997, de Jongh, de Wet, and Welsh 1988)
   - Recursive estimation in time-series, ARMA, GARCH, etc. (Koenker and Zhao 1996, Holst 1985, Holst 1987, Englund, Holst, and Ruppert 1988, Ruppert 1985)
   - Binary Response Models with Heteroscedasticity and Manski's Maximum Score Estimator (Manski 1975, Manski 1985, Manski and Thompson 1986, Horowitz 1992)
   - Multivariate quantiles and principal curves (Oja 1983, Ninimaa, Oja, and Nyblom 1992, Oja and Niinimaa 1985, Niinimaa, Oja, and Tableman 1990, Niinimaa and Oja 1995, Chakraborty and Chaudhuri 1995, Chakraborty 1996, Chaudhuri 1996, Liu 1990, Ramsay and Silverman 1997, Small 1990, Small 1997, Koltchinskii 1997, Hastie and Stuetzle 1989)
   - Rank based inference and related bootstrapping methods (Koenker 1994, Koenker 1996, Buchinsky 1994, Buchinsky 1995, Hahn 1995, Gutenbrunner and Jurečková 1992, Gutenbrunner, Jurečková , Koenker, and Portnoy 1993, Horowitz 1996, Parzen, Wei, and Ying 1994, Fitzenberger 1996, Gutenbrunner 1994, Koenker and Machado 1998)
3. **Comparison of Density and Nonparametric Regression Methods**
   The time seems right for another look at the pros and cons of various nonparametric methods currently available in the literature. In particular, I would like to compare current kernel methods with the sieve-like methods of the (Fenton and Gallant 1996, Gallant and Tauchen 1997) SNP approach and the log-spline methods of (Stone, Hansen, Kooperberg, and Truong 1997) and others. There are some interesting current applications to options pricing which might be an interesting case study here.
4. **Ecological Regression and Spatial Statistics**
   There is an interesting new book by a political scientist (King 1997) on the ecological regression fallacy which draws interesting parallels to current work on the statistical theory of image processing and tomography, and is also connected to current advances in spatial statistics, see for example (Small 1997). There are interesting connections to density estimation, nonparametric quantile regression and other topics which are suggested by these developments.
5. **Average Derivative Estimation, Single Indices, and the Transformation Model**
   There is still considerable scope for drawing together methods for estimating transformation models using a variety of techniques closely associated with some of the topics listed above.

Here there are also deeper theoretical issues relating to adaptive estimation. Single index models have a large literature in econometrics, see e.g. (Stoker 1992, Chaudhuri, Doksum, and Samarov 1997) Survival analysis is an important application in this context, as is the connection to discrete choice models. There are also interesting connections to rank statistics, see (Han 1987, Gutenbrunner 1997).

References

Abadie, A., J. Angrist, and G. Imbens (1997): "Instrumental Variables Estimation of Quantile Treatment Effects," preprint.

Amemiya, T. (1982): "Two Stage Least Absolute Deviations Estimators," *Econometrica*, 50, 689–712, Quantile regression for simultaneous equation models.

Ammann, L. P. (1993): "Robust Singular Value Decompositions: A New Approach to Projection Pursuit," *J. of Am. Stat. Assoc.*, 88, 505–514.

Angrist, J., G. Imbens, and D. B. Rubin (1996): "Identification of Casual Effects Using Instrumental Variables," *J. of Am. Stat. Assoc.*, 91, 444–472.

Arias, O., K. Hallock, and W. Sosa (1997): "Indidividual heterogeneity in the returns to schooling: IV-QR using Twins data," preprint.

Buchinsky, M. (1994): "Changes in US Wage Structure 1963-87: An Application of Quantile Regression," *Econometrica*, 62, 405–458.

——— (1995): "Quantile Regression, the Box-Cox Transformation Model and U.S. Wage Structure 1963-1987," *J. of Econometrics*, 65, 109–154.

Chakraborty, B. (1996): "On multivariate median regression," preprint.

Chakraborty, B., and P. Chaudhuri (1995): "On an adaptive transformation-retransformation estimate of multivariate," *J. of the Royal Statistical Society (B)*, forthcoming.

Chaudhuri, P. (1996): "On a geometric notion of quantiles for multivariate data," *J. of the American Statistical Association*, 91, 862–872.

Chaudhuri, P., K. Doksum, and A. Samarov (1997): "On average derivitive quantile regression," *Annals of Stat.*, 25, 715–744.

Chay, K. (1995): "Evaluating the impact of the 1994 Civil Rights Act on the Economic Status of Black Men Using Censored Longitudinal, Earnings Data," preprint.

de Jongh, P. J., T. de Wet, and A. H. Welsh (1988): "Mallows-type Bounded-influence-regression Trimmed Means," *J. of Am. Stat. Assoc.*, 83, 805–810.

Englund, J.-E., U. Holst, and D. Ruppert (1988): "Recursive M-estimators of Location and Scale for Dependent Sequences," *Scand. J. of Stat.*, 15, 147–159.

Fenton, V., and A. Gallant (1996): "Qualitative and Asymptotic Performance of SNP Density Estimators," *J. Econometrics*, 74, 77 –118.

Fitzenberger, B. (1996): "A Guide to Censored Quantile Regressions," in *Handbook of Statistics*, ed. by C. Rao, and G. Maddala. North-Holland: New York.

——— (1997): "Computational aspects of censorted quantile regression," in *L_1-Statistical Procedures and Related Topics*, ed. by Y. Dodge, IMS Lecture Notes-Monograph Series.

Gallant, A., and G. Tauchen (1997): "The Relative Efficiency of Method of Moments Estimators," preprint.

Gutenbrunner, C. (1994): "Tests for heteroscedasticity based on regression quantiles and regression rank scores," in *Asymptotic Statistics: Proceedings of the Fifth Prague Symposium*, ed. by P. Mandl, and Hušková . Physica-Verlag: Heidelberg.

——— (1997): "Regression Rank Statistics," in *L_1-Statistical Procedures and Related Topics*, ed. by Y. Dodge, IMS Lecture Notes-Monograph Series.

Gutenbrunner, C., and J. Jurečková (1992): "Regression quantile and regression rank score process in the linear model and derived statistics," *Ann. Statist.*, 20, 305–330.

Gutenbrunner, C., J. Jurečková , R. Koenker, and S. Portnoy (1993): "Tests of linear hypotheses based on regression rank scores," *J. of Nonparametric Statistics*, 2, 307–33.

Hahn, J. (1995): "Bootstrapping quantile regression models," *Econometric Theory*, 11, 105–121.

Han, A. (1987): "Nonparametric analysis of a generalized regression model," *J. Econometrics*, 35, 303–318.

Hastie, T., and W. Stuetzle (1989): "Principal Curves," *J. of Am. Stat. Assoc.*, 84, 502–516.

He, X., and S. Portnoy (1997): "Asymptotics of the Deepest Line," preprint.

Heckman, J. (1996): "Randomization as an Instrumental Variable," *The Review of Economics and Statistics*, 78, 336–340.

Holst, U. (1985): "Recursive Estimation of Quantiles," in *Contributions to Prob. and Stat. in Honour of Gunar Blom*, ed. by G. Lindgren, pp. 179– 188.

——— (1987): "Recursive Estimation of Quantiles Using Recursive Kernel Density Estimators," *Sequential Analysis*, 6, 219–237.

HOROWITZ, J. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505–531.

——— (1996): "Bootstrap methods for median regression models," *Econometrica*, forthcoming.

IMBENS, G., AND J. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

KING, G. (1997): *A Solution to the Econlogical Inference Problem*. Harvard: Cambridge.

KOENKER, R. (1994): "Confidence intervals for regression quantiles," in *Asymptotic Statistics: Proceedings of the 5th Prague Symposium on Asymptotic Statistics*, ed. by M. P., and M. Hušková . Physica-Verlag: Heidleberg.

——— (1996): "Rank Tests for Linear Models," in *Handbook of Statistics*, ed. by C. Rao, and G. Maddala. North-Holland: New York.

——— (1997): "ℓ_1 computation: an interior monologue," in *L_1-Statistical Procedures and Related Topics*, ed. by Y. Dodge, IMS Lecture Notes-Monograph Series.

KOENKER, R., AND J. MACHADO (1998): "Goodness of fit and likelihood ratio processes for quantile regression," preprint.

KOENKER, R., AND B. PARK (1996): "An interior point algorithm for nonlinear quantile regression," *Journal of Econometrics*, 71, 265–283.

KOENKER, R., AND Q. ZHAO (1996): "Conditional quantile estimation and interence for ARCH models," *Econometric Theory*, 12, 793–813.

KOLTCHINSKII, V. (1997): "M-estimation, convexity, and quantiles," *Annals of Stat.*, 25, 435–477, Multivariate Quantiles.

LAMPORT, L. (1994): *LaTeX*. Addison-Wesley.

LIU, R. (1990): "On a notion of data depth based on random simplices," *Annals of Statistics*, 36, 405–411.

MANSKI, C. (1975): "Maximum score estimation of the stochastic utility model of choice," *J. of Eonometrics*, 3, 205–228.

——— (1985): "Semiparametric analysis of discrete response asymptotic properties of the maximum score estimator," *J. of Eonometrics*, 27, 205–228.

——— (1997): "Monotone Treatment Response," *Econometrica*, 65, 1311–1334.

MANSKI, C. F., AND T. S. THOMPSON (1986): "Operational Characteristics of Maximum Score Estimation," *J. Econometrics*, 32, 85–108.

MARKATOU, M., AND J. L. HOROWITZ (1995): "Robust Scale Estimation in the Error-components Model Using the Empirical Characteristic Function," *Canadian J. of Stat.*, 23, 369–381.

NIINIMAA, A., AND H. OJA (1995): "On the Influence Functions of Certain Bivariate Medians," *JRSS-B*, 57, 565–574.

NIINIMAA, A., H. OJA, AND M. TABLEMAN (1990): "The Finite-sample Breakdown Point of the Oja Bivariate Median and of the Corresponding Half-samples Version," *StPrLet*, 10, 325–328.

NINIMAA, A., H. OJA, AND J. NYBLOM (1992): "The Oja Bivariate Median," *Applied Statistics*, 41, 611–17.

OJA, H. (1983): "Descriptive statistics for multivariate distributions," *Statistics and Probability Letters*, 1, 327–332.

OJA, H., AND A. NIINIMAA (1985): "Asymptotic Properties of the Generalized Median in the Case of Multivariate Normality," *JRSS-B*, 47, 372–377.

PARZEN, M. I., L. J. WEI, AND Z. YING (1994): "A Resampling Method Based on Pivotal Estimating Functions," *Biometrika*, 81, 341–350.

PORTNOY, S. (1997): "On computation of regression quantiles," in *L_1-Statistical Procedures and Related Topics*, ed. by Y. Dodge, IMS Lecture Notes-Monograph Series.

PORTNOY, S., AND R. KOENKER (1997): "The Gaussian Hare and the Laplacian Tortoise: Computability of squared-error versus absolute-error estimators, with discusssion," *Stat. Science*, 12, 279–300.

POWELL, J. L. (1983): "The Asymptotic Normality of Two-stage Least Absolute Deviations Estimators," *Econometrica*, 51, 1569–1576.

RAMSAY, J., AND B. SILVERMAN (1997): *Functional Data Analysis*. Springer.

ROCKE, D. M. (1983): "Robust Statistical Analysis of Interlaboratory Studies," *Biometrika*, 70, 421–431.

ROUSSEEUW, P., AND M. HUBERT (1996): "Regression Depth," preprint.

RUPPERT, D. (1985): "A Newton-Raphson Version of the Multivariate Robbins-Monro Procedure," *Annals of Stat.*, 13, 236–245.

SMALL, C. (1990): "A survey of multidimensional medians," *International Statistical Review*, 58, 263–277.

——— (1997): *The Statistical Theory of Shape*. Springer.

STOKER, T. (1992): *Lectures on Semiparametric Econometrics*. CORE Foundation: Louvain.

STONE, C., M. HANSEN, C. KOOPERBERG, AND Y. TRUONG (1997): "Polynomial Splines and their Tensor Products in Extended Linear Modeling, with discussion," *Annals of Statistics*, 25, 1371–1470.

VENABLES, W., AND B. RIPLEY (1997): *Modern Applied Statistics with Splus*. Springer.