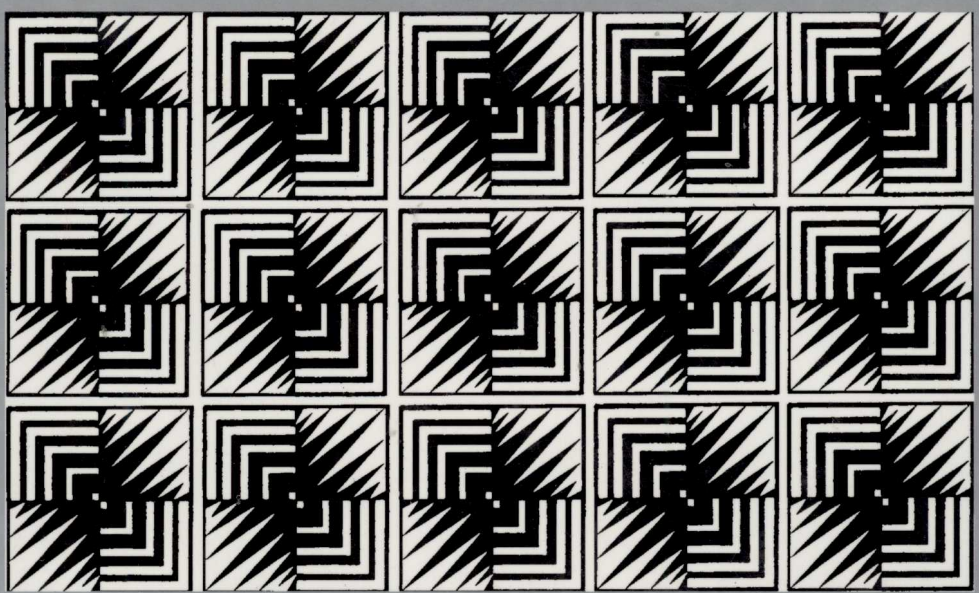
NONLINEAR STATISTICAL MODELING

PROCEEDINGS OF THE THIRTEENTH INTERNATIONAL SYMPOSIUM IN ECONOMIC THEORY AND ECONOMETRICS: ESSAYS IN HONOR OF TAKESHI AMEMIYA



EDITED BY
CHENG HSIAO
KIMIO MORIMUNE
JAMES L. POWELL

INTERNATIONAL SYMPOSIA IN ECONOMIC THEORY AND ECONOMETRICS

Nonlinear statistical modeling

Proceedings of the Thirteenth International Symposium
in Economic Theory and Econometrics:
Essays in Honor of Takeshi Amemiya

Edited by

CHENG HSIAO
University of Southern California

KIMIO MORIMUNE
Kyoto University

JAMES L. POWELL
University of California, Berkeley

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2001

Printed in the United States of America

Typeface Times Roman 10/12 pt. *System* LaTeX 2_ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

International Symposium in Economic Theory and Econometrics (13th)

Nonlinear statistical modeling : proceedings of the thirteenth International Symposium in Economic Theory and Econometrics : essays in honor of Takeshi Amemiya / edited by Cheng Hsiao, Kimio Morimune, James L. Powell.

p. cm. – (International symposia in economic theory and econometrics)

"Bibliography of Takeshi Amemiya."

Includes bibliographical references.

ISBN 0-521-66246-X

1. Econometric models – Congresses. 2. Nonlinear theories – Congresses. 3. Parameter estimation – Congresses. 4. Sampling (Statistics) – Congresses. I. Amemiya, Takeshi. II. Hsiao, Cheng, 1943– III. Morimune, Kimio. IV. Powell, James, 1955– V. Title. VI. Series.

HB141 .I57 2000
519.5 – dc21

00-031289

Series editor's introduction

This volume is the thirteenth in a series, called *International Symposia in Economic Theory and Econometrics*. The series is under the general editorship of William A. Barnett. Individual volumes in the series generally have editors, who differ for each volume, since the topics of the volumes change each year. The editors of this volume are Cheng Hsiao, Kimio Morimune, and James L. Powell.

The primary focus of this book is "Nonlinear Statistical Inference," which is the book's title. But the volume's breadth extends beyond that topic, since the volume was produced in honor of Takeshi Amemiya, whose influence in econometrics extends to the many areas of econometrics and statistics in which he has dramatically increased the level of sophistication, rigor, and depth. The inspiration provided by Amemiya's work to that of other leading econometricians is reflected by the chapters contained in this book, which includes recent advances in (i) the parametric approach to qualitative response and sample selection models, (ii) the nonparametric and semiparametric approaches to qualitative response and sample selection models, and (iii) nonlinear estimation of cross-sectional and time series models.

Many of the prior volumes in this series were sponsored by the IC^2 Institute at the University of Texas at Austin, and some have been cosponsored by the RGK Foundation. The first conference in this Cambridge series was co-organized by William Barnett and Ronald Gallant, who also co-edited the proceedings volume. That volume has appeared as the volume 30, October/November 1985 issue of the *Journal of Econometrics* and has been reprinted as a volume in this Cambridge University Press monograph series. The topic was "New Approaches to Modeling, Specification Selection, and Econometric Inference."

Beginning with the second symposium in the series, the proceedings of the symposia appear exclusively as volumes in this Cambridge University Press monograph series. The co-organizers of the second symposium and co-editors of its proceedings volume were William Barnett and Kenneth Singleton. The

topic was "New Approaches to Monetary Economics." The co-organizers of the third symposium, which was on "Dynamic Econometric Modeling," were William Barnett and Ernst Berndt; and the co-editors of that proceedings volume were William Barnett, Ernst Berndt, and Halbert White. The co-organizers of the fourth symposium and co-editors of its proceedings volume, which was on "Economic Complexity: Chaos, Sunspots, Bubbles, and Nonlinearity," were William Barnett, John Geweke, and Karl Shell. The co-organizers of the fifth symposium and co-editors of its proceedings volume, which was on "Nonparametric and Semiparametric Methods in Econometrics and Statistics," were William Barnett, James Powell, and George Tauchen. The co-organizers and proceedings coeditors of the sixth symposium, which was on "Equilibrium Theory and Applications," were William Barnett, Bernard Cornet, Claude d'Aspremont, Jean Gabszewicz, and Andreu Mas-Colell. The co-organizers of the seventh symposium, which was on "Political Economy," were William Barnett, Melvin Hinich, Douglass North, Howard Rosenthal, and Norman Schofield. The co-editors of that proceedings volume were William Barnett, Melvin Hinich, and Norman Schofield.

The eighth symposium was part of a large scale conference on "Social Choice, Welfare, and Ethics." That conference was held in Caen, France in June 9–12, 1993. The organizers of the conference were Maurice Salles and Herve Moulin. The co-editors of that proceedings volume were William Barnett, Herve Moulin, Maurice Salles, and Norman Schofield. The ninth volume in the series was on "Dynamic Disequilibrium Modeling: Theory and Applications," and was organized by Claude Hillinger at the University of Munich, Giancarlo Gandolfo at the University of Rome "La Sapienza," A. R. Bergstrom at the University of Essex, and P. C. B. Phillips at Yale University. The co-editors of the proceedings volume were William Barnett, Claude Hillinger, and Giancarlo Gandolfo.

Much of the contents of the tenth volume in the series comprises the proceedings of the conference, "Nonlinear Dynamics and Economics," held at the European University Institute in Florence, Italy, on July 6–17, 1992. But the volume also includes the related, invited chapters presented at the annual meetings of the American Statistical Association held in San Francisco on August 8–12, 1993. The organizers of the Florence conference, which produced part of the tenth volume, were Mark Salmon and Alan Kirman at the European University Institute in Florence, and David Rand and Robert MacKay from the Mathematics Department at Warwick University in England, while the organizer of the invited American Statistical Association sessions, which produced the other chapters in the volume, was William Barnett, who was Program Chair in Economic and Business Statistics of the American Statistical Association during that year.

The eleventh volume was the proceedings of a conference held at the University of Aarhus, Denmark, on December 14–16, 1995. In addition to being the eleventh in this series, that volume was the proceedings of the Sixth Meeting of the European Conference Series in Quantitative Economics and Econometrics, $(EC)^2$. The organizer of the Aarhus conference was Svend Hylleberg at the University of Aarhus. The editors of the proceedings volume were William A. Barnett, David F. Hendry, Svend Hylleberg, Timo Teräsvirta, Dag Tjøstheim, and Allan Würtz. The topic of the conference and focus of that book was "Nonlinear Econometric Modeling," with an emphasis on nonlinear time series.

The twelfth volume was the proceedings of a conference held at the University of New South Wales in Sydney, Australia, in 1996. The organizers of the Sydney conference were Carl Chiarella, Steve Keen, Bob Marks, and Hermann Schnabl. The editors of the proceedings volume were William Barnett, Carl Chiarella, Steve Keen, Bob Marks, and Hermann Schnabl. The topic of the conference and focus of that book was "Commerce, Complexity, and Evolution: Complexity and Evolutionary Analysis in Economics, Finance, Marketing, and Management," with an emphasis on the new evolutionary approach to analyzing and modeling commercial systems. The research in that volume straddles economics, finance, marketing, and management.

The intention of the volumes in this series is to provide *refereed* journal-quality collections of research chapters of unusual importance in areas of currently highly visible activity within the economics profession. Most of the volumes in the series are proceedings of conferences. Because of the refereeing requirements associated with the editing of the volumes, they do not necessarily contain all of the chapters presented at the corresponding symposia or submitted to the editors of the volume.

William A. Barnett
Washington University in St. Louis

Editors' introduction

The chapters presented in this volume are dedicated to Takeshi Amemiya in honor of his sixty-fifth birthday. Takeshi Amemiya stands out among econometricians as distinctive in the rigor and breadth in his contributions. He has done path-breaking work in areas as diverse as limited dependent variables, discrete choice, nonlinear estimation, duration analysis, panel data, and dynamic models and simultaneous equation models. He has contributed to raising the general analytical and methodological level of econometrics. He has shown the advantages of strict formalization of the analytical techniques, thereby setting the style of generations of econometricians. In spite of the high level of abstraction of much of his work, the advances in the theory achieved have had important bearing on the choice of methods and analytical techniques in applied research. He has been an inspiration to econometricians all over the world.

Besides being a scientist and a scholar, above all, Takeshi Amemiya is a gentleman. He gives generously of himself and is always willing to help the cause of econometrics. We edit this volume on behalf of Takeshi Amemiya's students and colleagues to highlight a small part of his outstanding contribution to the profession. The collection in this volume put together important recent advances in (i) parametric approaches to qualitative response and sample selection models, (ii) nonparametric and semiparametric approaches to qualitative response and sample selection models, and (iii) nonlinear estimation of cross-sectional and time series models.

The chapter by J. J. Heckman and E. J. Vytlacil unites the treatment effect literature and the latent variable literature. It uses the local instrumental variable (LIV) parameters and the Roy model as a unifying concept and a classificatory model, respectively. The LIV can be viewed as a marginal relative of the local average treatment effect (LATE) proposed by Imbens and Angrist (1994), and the chapter shows its close mathematical relation to other causal parameters of interest, the average treatment effect (ATE) and effect of treatment on the treated (TT). The chapter applies these various treatment effect concepts to a particular

model of economic interest, the Roy model of self-selection, and illustrates how the LIV parameter can be used to identify normalized versions of the cost and benefit functions for that model. It also discusses the relative merits of the various treatment effect concepts in traditional cost–benefit analysis of social policy, and give bounds for the treatment parameters when the support of the distribution of propensity scores is not sufficiently rich to identify them exactly.

The chapter by N. E. Savin and A. H. Würtz compares the properties of hypothesis tests for limited dependent variable models that use critical values based on first order asymptotic approximations with tests that use critical values calculated by the bootstrap. The authors argue persuasively that critical values based on first order asymptotic approximations sometimes get the size wrong. They find that for several permutations of a probit and tobit model in a sample of fifty observations, the tests that use the bootstrap critical values have more accurate size.

The chapter by L.-F. Lee considers simulation estimation of sample selection models. Simulation estimation techniques are useful when sample selection criteria are complicated with polychotomous choice alternatives and correlated disturbances. Simulated likelihood and two-stage estimation methods are considered. For two-stage estimation, possible simulation estimation of outcome regression equations based on Gibbs sampler and variance-reduction techniques are considered. Monte Carlo results are provided for finite sample comparison on performance of various simulation methods. Issues of sensitivity of parameter estimates to distributional assumptions are investigated. Semiparametric estimation methods with simulation are also studied in Monte Carlo experiments. A useful decomposition of the finite sample variances of the estimators into their simulation-based and sample-variation-based components is also provided.

The chapter by K. Ryu proposes a new approach to tackling attrition problems in panel studies. Panel data sets are often unbalanced due to attrition. Over time, cross-sectional units are dissipated, possibly in a nonrandom fashion, thus negating the initial randomization. If the variables of interest are stochastically related to the factors governing the attrition process, then the attrition is called *nonrandom*, and is said to cause *attrition bias*. This chapter proposes three different ways of accounting for attrition problem in unbalanced panel data. They key element of the Ryu's approach is to identify the nature of selection arising from panel attrition. Using grouped duration techniques with heterogeneity, the author proposes to estimate the attrition process. Once the nature of attrition is identified, Hausman–Wise–Ridder joint maximum likelihood estimation, sequential maximum likelihood estimation, and Heckman (1976, 1979) sequential estimation are proposed.

The chapter by F. Goto discusses estimation of parametric duration models which are subject to left-censoring and in which the distribution of starting times

is not parametrically specified. The authors shows that if one has a sample of the completed spells, sampled from the spells that are in progress at a point in time, then the semiparametric maximum likelihood estimator (MLE) that takes this sampling scheme into account is efficient. This is true for both the parametric and nonparametric components.

The chapter by J. L. Powell proposes semiparametric estimators of a class of bivariate latent dependent variable models, including the censored sample selection model and the disequilibrium regression model. Estimation proceeds in two steps. In the first step, consistent semiparametric estimates of the coefficients of the "selection" or "regime" equation are obtained for the *single index* variables characterizing the selectivity bias in the equation of interest. In the second step a weighted instrumental variables method is applied to the pairwise differences of the equation of interest. The chapter proves the root-n consistency and asymptotic normality of the proposed estimator and proposes consistent estimators of its asymptotic covariance matrix.

The chapter by Y. Nishiyama and P. M. Robinson considers semiparametric estimates, up to scale, of single index models. A valid Edgeworth expansion is established for the limiting distribution of density-weighted semiparametric averaged derivative estimates, which depends on the choice of bandwidth and kernel order.

The epidemiological literature has devoted much attention to relative risks and attributable risks, that is, to ratios and differences of conditional response probabilities at distinct covariate values. The chapter by C. F. Manski provides interesting bounds on these measures using auxiliary distributional information. Under the less stringent assumptions that only the marginal distribution of the response, $P(y)$, is known, but not the marginal distribution of the covariate, $P(z)$, or that $P(z)$ is known, but not $P(y)$, Manski proves that the available information implies informative sharp bounds on response probabilities, relative risks, and attributable risks.

A common problem in applied regression analysis is to select the variables that enter a regression model. It is often computationally convenient to estimate such models by least squares, with variables selected from possible candidates by enumeration, grid search, or Gauss–Newton iteration to maximize their conventional least squares significance level; this method is termed *prescreened least squares* (PLS). D. McFadden shows that PLS in equivalent to direct estimation by nonlinear least squares, and thus provides a practical computational shortcut that shares the statistical properties of the nonlinear least squares solution. However, standard errors and test statistics produced by least squares software are biased by variable selection. The chapter also gives consistent estimators for covariance and test statistics and shows in examples that kernel-smoothing or bootstrap methods appear to give adequate approximation in samples of moderate size.

The chapter by T. E. MaCurdy develops new estimators for regression and simultaneous equations that make use of the second and higher moments of the disturbances of the equation under consideration. The estimator are in general asymptotically strictly efficient relative to the conventional least squares and two-stage least squares procedures, except when these latter procedures correspond directly to the application of maximum likelihood methods. The small scale Monte Carlo findings suggest that substantial efficiency gains may be attainable by implementing these estimation procedures.

In proving the weak convergence of the standardized empirical spectral distribution function of a stationary process to a Gaussian process, Grenander and Rosenblatt (1952) assumed that the eighth moment of the process is finite. T. W. Anderson and L. You's chapter gives a neat and direct proof that the existence of the second moment is sufficient for weak convergence of the standardized spectral distribution when the process consists of independently identically distributed variables.

The chapters by H. Lütkepohl, C. Müller, and P. Saikkonen and by K. Morimune and M. Nakagawa consider unit root tests. These tests are important because the trending properties of a set of time series determine to some extent which model and statistical procedures are suitable for analyzing their relationship. When the breakpoint is known, this is useful information which should be taken into account in testing for unit roots. Lütkepohl, Müller, and Saikkonen propose a framework in which the shift is modeled as part of the intercept term of the stationary part of the data-generating process, which is separated from the unit root part. In this framework simple shift functions result in a smooth transition from one state to another, both under the null hypothesis of a unit root and under the alternative hypothesis of stationarity. A major advantage of their approach is that for conducting a standard unit root test, e.g. of the Dickey-Fuller type, the estimation of the nuisance parameters reduces to a fairly simple nonlinear least squares problem. Moreover, their asymptotic null distribution is known from the unit root literature, and tables of critical values exist.

Morimune and Nakagawa's chapter proposes a unified framework for the unit root tests that allows for multiple breaks in both the intercepts and trend coefficients. They derive the asymptotic expressions for the t-test statistics. The small-sample power of the Dickey–Fuller test (1981) and Perron (1989) tests are compared by simulations, and the Dickey–Fuller test is found to be more powerful.

The chapter by N. Kunitomo and S. Sato introduces a simultaneous switching autoregressive model (SSAR) as a generalization of the Tobit model. The distinction between the structural form and the reduced form SSAR is stressed. The reduced form SSAR is shown as a nonlinear Markovian time series model. This type of time series modelling can be useful for the analysis of financial time

series data. The chapter has given some conditions for the geometrical ergodicity and proposes maximum likelihood and instrumental variable estimators that are consistent and asymptotically normally distributed.

One of the issues in panel data analysis is to control the effect of unobserved heterogeneity on inference about characteristics of the population. When cross-sectional dimension is large and the time series dimension is finite, the presence of individual specific effects introduces the classical incidental parameter problem (Neyman and Scott 1948). If the model is linear in individual specific effects (incidental parameters), one can difference out the individual specific effects and obtain, from the differenced equation, consistent estimators of the parameters that are common across individuals and over time (structural parameters). If the model is nonlinear in the effects, this differencing method does not work. The structural parameters usually cannot be consistently estimated, because of the presence of incidental parameters. The chapter by T. Lancaster proposes a novel approach of orthogonal reparameterization of the fixed effect for each individual to a new fixed effect which is independent of the structural parameters in the information matrix sense. The chapter shows how we can derive consistent estimators of structural parameters for a duration model using two consecutive uncensored spelled lengths for many people.

The chapter by S. J. Yhee, J. B. Nugent, and C. Hsiao uses a censored switching regression approach to evaluate the effect of sunk costs and firm level disequilibrium on export performance using panel data on South Korean small and medium sized enterprises. The effect of sunk costs is captured as the difference between the technology capability and the marketing capability, each depending on the characteristics of the individual firm, and the results are used to measure the firm-specific effectiveness of export support systems.

Takeshi Amemiya has had a lasting influence on nearly every field of econometrics. Clearly no single volume can do justice to the wide variety of topics to which he has made pioneering contributions. Nevertheless, the included chapters provide a rich selection of techniques and applications that are inspired by Amemiya's work.

Finally, we would like to thank the economics editor of the Cambridge University Press, Scott Parris, for his support of this project. We are also grateful to a number of our colleagues who have acted as anonymous referees. Their comments have led to the improvement of the chapters published in this volume.

REFERENCES

Dickey, D. A. and Fuller, W. A. (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, **49**, 1057–1072.

Grenander, U. and Rosenblatt, M. (1952), "On Spectral Analysis of Stationary Time Series," *Proceedings of the National Academy of Sciences*, **38**, 519–521.

Heckman, J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for such Models," *Annals of Economics and Social Measurement*, **5**, 474–492.

(1979), "Sample Selection Bias as a Specification Error," *Econometrica*, **47**, 153–161.

Imbens, G. and Angrist, J. (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, **62**, 467–476.

Neyman, J. and Scott, E. (1948), "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, **16**, 1–32.

Perron, P. (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, **57**, 1361–1401.

Contributors

T. W. Anderson
Stanford University

Fumihiro Goto
Aoyama Gakuin University

James J. Heckman
University of Chicago

C. Hsiao
University of Southern California
and Hong Kong University
of Science and Technology

Naoto Kunitomo
University of Tokyo

Tony Lancaster
Brown University

Lung-fei Lee
Hong Kong University of Science
and Technology

Helmut Lütkepohl
Humboldt University

Thomas E. MaCurdy
Stanford University

Charles F. Manski
Northwestern University

Daniel McFadden
University of California, Berkeley

Kimio Morimune
Kyoto University

Christian Müller
Humboldt University

Mitsuru Nakagawa
Osaka City University

Y. Nishiyama
Nagoya University

J. B. Nugent
University of Southern
California

James L. Powell
University of California, Berkeley

P. M. Robinson
London School of Economics

Keunkwan Ryu
Seoul National University
and Hong Kong University
of Science and Technology

Pentti Saikkonen
University of Helsinki

Seisho Sato
Institute of Statistical Mathematics, Tokyo

N. E. Savin
University of Iowa

Edward J. Vytlacil
Stanford University

Allan H. Würtz
University of Aarhus

Seung-Jae Yhee
American Express

L. Y. You
Salomon Smith Barney

Contents

CHAPTER 1

Local instrumental variables

James J. Heckman and Edward J. Vytlacil

1 Introduction

Takeshi Amemiya made basic contributions to the econometrics of discrete choice and limited dependent variable models. His fundamental 1973 paper on the censored normal regression model was the first systematic application in econometrics of the uniform law of large numbers and the central limit theorems required to establish consistency and asymptotic normality of nonlinear econometric models. That paper and his later work on nonlinear least squares are summarized in his magisterial text *Advanced Econometrics* (1985). His research and his text set the standard for a generation of econometricians, and provided the framework for modern structural econometric analysis of index models or latent variable models.

Latent variable models of the type analyzed by Amemiya arise in well-posed economic problems. The latent variables can be utilities, potential wages (or home wages as in Gronau 1974 and Heckman 1974) or potential profitability. This class of models, which originates in psychology in the work of Thurstone (1930, 1959), has been widely developed in the econometrics of discrete choice (see McFadden 1981, and the survey of index models in labor economics presented by Heckman and MaCurdy 1985). Amemiya provided

James J. Heckman is Henry Schultz Distinguished Service Professor of Economics at the University of Chicago and a Senior Fellow at the American Bar Foundation. Edward Vytlacil is a Sloan Fellow at the University of Chicago. We would like to thank Arild Aakvik, Jaap Abbring, Victor Aguirregabiria, Lars Hansen, Justin Tobias, two anonymous referees, and especially Cheng Hsiao for close and helpful readings of this manuscript. We thank participants in the Canadian Econometric Studies Group (September, 1998), The Midwest Econometrics Group (September, 1998), the University of Upsalla (November, 1998), the University of Chicago (December, 1998), University College London (December, 1998), the University of California, Berkeley (April, 1999), and the North American Summer Meetings of the Econometric Society (June, 1999). This research was supported by NIH:R01-HD3498-01, NIH:R01-HD3208-03, NSF97-09-873, and the Donner Foundation.