Identification and Inference for Econometric Models

Essays in Honor of Thomas Rothenberg

EDITED BY

Donald W. K. Andrews

AND

James H. Stock

CAMBRIDGE

Identification and Inference for Econometric Models

Essays in Honor of Thomas Rothenberg

Edited by

Donald W. K. Andrews
Yale University

James H. Stock
Harvard University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo, Mexico City

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521154741

First published 2005
First paperback printing 2010

A catalog record for this publication is available from the British Library

Library of Congress Cataloging in Publication data

Identification and inference for econometric models / edited by Donald W. K.
Andrews, James H. Stock.
p. cm.
"A Festschrift in Honor of Thomas J. Rothenberg."
Includes bibliographical references.
ISBN 0-521-84441-X (hardback)
1. Econometric models. I. Andrews, Donald W. K. II. Stock, James H.
III. Rothenberg, Thomas J. IV. Title.
HB141.I143 2005
330′.01′5195 – dc22 2004025817

ISBN 978-0-521-84441-3 Hardback
ISBN 978-0-521-15474-1 Paperback

Preface

Donald W. K. Andrews and James H. Stock

The chapters in this volume are dedicated to Thomas Rothenberg in honor of his retirement from the Economics Department at the University of California, Berkeley. Tom Rothenberg has made fundamental contributions to econometric theory and has been an inspiring teacher, advisor, and colleague. Rothenberg's early work focused on efficient estimation and identification in simultaneous equations models. In a paper (written with C. T. Leenders) published in *Econometrica* while he was still a graduate student, Rothenberg established the asymptotic efficiency of the linearized maximum likelihood estimator for simultaneous equations models and thus the asymptotic efficiency of three-stage least squares. This line of research was summarized in his monograph *Efficient Estimation with A Priori Information*, where he laid out a unified theory of efficient estimation in simultaneous equations systems.

Because exact optimality results for estimators and tests in simultaneous equations models are generally unavailable, the notion of efficiency in Rothenberg's initial work typically is first-order asymptotic efficiency. Often, however, there are a number of estimators that are asymptotically equivalent to first order; k-class estimators in a single equation with multiple endogenous regressors is a leading example. In finite samples, these estimators have different behavior, but their finite-sample distributions can be either unavailable or so complicated that they fail to provide useful comparisons between the estimators. Thus, Rothenberg undertook to examine the differences between first-order equivalent estimators and tests by studying their higher-order properties using Edgeworth expansions. Much of this work is summarized in his masterful chapter in the 1984 *Handbook of Econometrics*, which remains a key reference for researchers interested in the deviations of the distributions of instrumental variables estimators from their first-order asymptotic distributions. More recently, Rothenberg's interest in efficient inference led him to consider efficient testing in time series with a possible unit root.

Both of the editors of this Festschrift had the privilege of being students of Tom Rothenberg. Like his other students, we benefited from traits that are hallmarks of his research: an insistence on working on problems that are important to econometrics, bringing common sense to both the economics and the econometric theory at hand, an appreciation for the statistical foundations of econometric theory, and a realization that careful analysis of simple models can yield deeper insights about econometric procedures applied in the more complicated settings found in practice.

Most of the papers in this volume fall into one of the three main areas of Rothenberg's research: identification and efficient estimation; analysis of asymptotic approximations, for example, via higher-order asymptotic analysis; and inference involving potentially nonstationary time series. In addition, several papers are in the area of nonparametric and semiparametric inference.

The majority of the papers in this volume were presented at a National Science Foundation conference in honor of Tom Rothenberg held in Berkeley, California, in August 2001. This conference was organized by James Powell and Paul Ruud.

Identification and Efficient Estimation (Part I)

At the request of the editors, this Festschrift starts with a classic unpublished paper in which Rothenberg explores the subtle role of modeling assumptions for causal inferences. By illustrating how seemingly innocuous assumptions can lead to incredible inferential conclusions, the chapter emphasizes the importance of thoughtful consideration of the assumptions underlying a statistical analysis and of focusing on results that are robust to untestable modeling assumptions.

The chapter by Arthur Goldberger continues this theme. Goldberger considers studies of twins in behavioral genetics. He illustrates how modeling assumptions that seem plausible on their face can lead to implausibly strong conclusions that are not robust to questionable assumptions on unobservables – specifically, assumptions about correlations between genetic characteristics and the environment.

Jeffrey Wooldridge's chapter addresses the identification and estimation of causal effects in nonlinear models and examines how certain estimands are more robust than others to violation of assumptions on unmodeled heterogeneity. In particular, he shows that, under certain conditional independence assumptions, it is possible to estimate average partial effects in nonlinear models consistently, even with unobserved heterogeneity and even though this heterogeneity can lead to inconsistency of estimated parameters (such as probit slope coefficients) of standard nonlinear models.

David Freedman's chapter also considers what assumptions are needed to provide a causal interpretation to regression coefficients estimated using nonexperimental data and emphasizes the importance of having prior information about causal mechanisms – that is, a model in which one believes – if one is to draw causal inferences. Freedman makes these arguments using graphical causal models, a framework more commonly encountered outside rather than inside the field of econometrics. His conclusions reinforce those in the chapters by Rothenberg and Goldberger about the key role played in identification by subsidiary modeling assumptions.

James Stock and Motohiro Yogo consider a different aspect of identification in econometrics: instrumental variables regression when the coefficient of interest is identified but, for the sample size at hand, the marginal explanatory power of the instruments is small, that is, the instruments are weak. As Rothenberg and others have shown, in this case the distributions of IV estimators are poorly approximated by their first-order asymptotic distributions, and Stock and Yogo propose tests of the hypothesis that the instruments are weak against the alternative that they are strong. In a companion chapter, they also derive alternative asymptotic distributions for k-class IV estimators when there are many weak instruments.

The chapter by Douglas Steigerwald and Richard Vagnoni examines the role of modeling assumptions in achieving identification in the context of a dynamic financial model of stock and stock option prices. The model captures salient stylized empirical facts, including serial correlation in stock trades, serial correlation in stock price changes, and more persistent serial correlation in stock trades than in squared stock price changes. Steigerwald and Vagnoni use this model to illustrate how subsidiary modeling assumptions (in this case, assumptions about the process of trader arrival) play an important role in the identification of the model parameters.

Asymptotic Approximations (Part II)

Rothenberg's teaching and research have emphasized the virtues of using alternative asymptotic frameworks, beyond conventional $\sqrt{n}$-normal asymptotics, to understand and compare the performance of estimators and test statistics. For example, Rothenberg's work on higher-order expansions is well known. The chapters by Hidehiko Ichimura and Oliver Linton, by Donald Andrews, by Guido Imbens and Richard Spady, and by Whitney Newey, Joaquim Ramalho, and Richard Smith all follow this approach and employ higher-order expansions to analyze and improve methods based on first-order asymptotics.

Ichimura and Linton calculate higher-order expansions for semiparametric estimators of treatment effects. They use these expansions to define a method for bandwidth selection and to specify a degrees of freedom–like bias correction.

Andrews uses Edgeworth expansions to compare competing bootstrap methods for parametric time series models. In particular, he shows that a parametric bootstrap based on the maximum likelihood estimator achieves greater improvements in coverage probabilities than the nonparametric block bootstrap. Moreover, he shows that these improvements can be achieved using a linearized k-step version of the estimator, resulting in substantial computational savings.

Imbens and Spady calculate higher-order biases and mean-squared errors of generalized method of moments (GMM) and generalized empirical likelihood (GEL) estimators in a simple model with a sequence of moment conditions. Their analysis suggests that GEL estimators outperform feasible GMM estimators. In addition, they find that the relative performances of different GEL estimators depend on the magnitudes of third moments of the moment conditions.

Newey, Ramalho, and Smith establish stochastic expansions for GMM and GEL estimators that may depend on preliminary nuisance parameters. Examples considered include estimators of models with sample selection corrections and estimators of covariance structures. Their results also cover two-step GMM estimators with sample splitting employed to estimate the weight matrix. The stochastic expansions are used to analytically bias-correct the GMM and GEL estimators. Simulation experiments are used to show that this method works well in the case of covariance structure models.

The chapter by Ron Mittelhammer, George Judge, and Ron Schoenberg uses Monte Carlo simulation methods to analyze the finite-sample properties of GEL, GMM, and two-stage least-squares estimators in a linear structural model. They also provide an algorithm for computation of GEL estimators.

The chapter by Ole Barndorff-Nielsen and Neil Shephard considers asymptotic approximations in time series models. The authors numerically compare different first-order equivalent approximations to the distribution of the local sum of squared financial returns (the so-called realized variance).

The chapter by Gene Savin and Allan Würtz considers tests concerning the transformation parameter in Box–Cox regression models with unknown error distributions. Using Monte Carlo simulations, they find that Wald tests based on first-order asymptotics have poor size properties. In contrast, they find that GMM residual-based bootstrap tests have only small discrepancies between nominal and true null rejection probabilities.

Inference Involving Potentially Nonstationary Time Series (Part III)

The chapters by Michael Jansson, by Samuel Thompson, and by Andrew Harvey consider inference about the degree of persistence in time series. Jansson considers tests of the null hypothesis that a vector time series is cointegrated. Specifically, he applies the theory of point optimal tests for a unit moving average root to the residual from a cointegrating regression to develop a new family of tests of the null hypothesis of cointegration. Thompson focuses on the problem of constructing confidence intervals for autoregressive coefficients when the true value is nearly one. Thompson shows that intervals based on inverting robust tests can result in substantial improvements over procedures using only second moments when the errors are heavy-tailed. In his chapter, Harvey proposes a unified framework for testing for stationarity and unit roots in both univariate and multivariate time series. The unifying concept is that the tests have generalized Cramér–von Mises distributions, and Harvey shows how to derive such tests via the Lagrange multiplier principle.

The chapters by Jushan Bai and Serena Ng and by Brownwyn Hall and Jacques Mairesse examine inference in potentially persistent panel data. Bai and Ng consider a common components model and study tests for the stationarity of the common components against the alternative that one or more common components have a unit root. In their chapter, Hall and Mairesse use Monte Carlo simulations to compare the performance of various unit root tests that have been proposed for panel data, focusing on the common case in which there are few time series observations on a large number of individuals or firms. They find that many existing tests have substantial size distortions, especially when there is firm-level heteroskedasticity.

David Hendry and Grayham Mizon consider forecasting in the presence of a different sort of nonstationarity: structural breaks and policy regime shifts. They develop a framework in which structural shifts in causal structural models lead those causal models to produce poor forecasts, whereas nonstructural models can produce reliable forecasts; one of their conclusions is that forecast failure of an econometric model need not rule out its usefulness for forecasting.

Nonparametric and Semiparametric Inference (Part IV)

The chapter by Peter Bickel, Ya'acov Ritov, and Tom Stoker examines the fundamental question of the choice of regressors in a regression model. In contrast to much of the literature on this problem, they analyze a nonparametric regression model rather than a linear model. They develop tests for exclusion restrictions in the nonparametric regression context.

Bo Honoré and James Powell exploit the pairwise differencing approach commonly used to eliminate a fixed effect in a linear panel data model to estimate various semiparametric nonlinear models, including the partially linear logit model. They establish $\sqrt{n}$-consistency and asymptotic normality of estimators that are minimizers of kernel-weighted U-statistics.

The chapter by Whitney Newey and Paul Ruud considers semiparametric estimation of single-index models. The authors establish $\sqrt{n}$-consistency and asymptotic normality of the inverse-density-weighted quasi–maximum likelihood estimator introduced by Ruud in 1986. This estimator has an advantage over alternative estimators in that it allows for discontinuities in the unknown transformation function.

Contents

List of Contributors

Donald W. K. Andrews
Yale University

Jushan Bai
New York University and Tsinghua University, Beijing

Ole E. Barndorff-Nielsen
University of Aarhus

Peter J. Bickel
University of California, Berkeley

David A. Freedman
University of California, Berkeley

Arthur S. Goldberger
University of Wisconsin, Madison

Brownwyn H. Hall
University of California, Berkeley

Andrew C. Harvey
University of Cambridge

David F. Hendry
University of Oxford

Bo E. Honoré
Princeton University

Hidehiko Ichimura
University College, London

Guido W. Imbens
University of California, Berkeley

Michael Jansson
University of California, Berkeley

George G. Judge
University of California, Berkeley

Oliver Linton
London School of Economics

Jacques Mairesse
CREST-INSEE and EHESS

Ron C. Mittelhammer
Washington State University

Grayham E. Mizon
University of Southampton

Whitney K. Newey
Massachusetts Institute of Technology

Serena Ng
University of Michigan

James L. Powell
University of California, Berkeley

Joaquim J. S. Ramalho
University of Evora

Ya'acov Ritov
The Hebrew University of Jerusalem

Thomas J. Rothenberg
University of California, Berkeley

Paul A. Ruud
University of California, Berkeley

N. E. Savin
University of Iowa

Ron Schoenberg
Aptech Systems

Neil Shephard
Nuffield College, Oxford

Richard J. Smith
University of Warwick

Richard H. Spady
Northwestern University and Nuffield College, Oxford

Douglas G. Steigerwald
University of California, Santa Barbara

James H. Stock
Harvard University

Thomas M. Stoker
Massachusetts Institute of Technology

Samuel B. Thompson
Harvard University

Richard J. Vagnoni
University of California, Santa Barbara

Jeffrey M. Wooldridge
Michigan State University

Allan H. Würtz
University of Aarhus

Motohiro Yogo
University of Pennsylvania

PART I

IDENTIFICATION AND EFFICIENT ESTIMATION

CHAPTER 1

Incredible Structural Inference

Thomas J. Rothenberg

1. INTRODUCTION

In the course of their everyday work, economists routinely employ statistical techniques to analyze data. Typically, these techniques are based on probability models for the observations and justified by an appeal to the theory of statistical inference. An important example is the estimation of structural equations relating economic variables. Such equations are interpreted as representing causal mechanisms and are widely used for forecasting and policy analysis. This econometric approach is arguably the dominant research methodology today among applied economists both in and out of academia.

The econometric approach is not without its critics. Scholars from other disciplines often seem puzzled by the emphasis that economists place on regression analysis. Statisticians express surprise that their techniques should be applicable to so many situations. Recently, a number of leading econometricians have added to the critique. In his paper "Let's Take the Con Out of Econometrics," Ed Leamer (1983) chides economists for ignoring the fragility of their estimates. The title of this paper comes from Christopher Sims's (1980) paper "Macroeconomics and Reality," which argues that the economic and statistical assumptions underlying most macromodels are not believable. They are, he asserts, literally "incredible."

Although my purpose is similar to that of Leamer and Sims, my approach will be rather different. In any area of application there will always be differences of opinion on what constitutes a reasonable set of assumptions on which to base the statistical analysis. Particularly in macroeconomics, where one is trying to summarize in a manageable aggregate model the behavior of millions of decision makers with regard to thousands of products, the disagreements are bound to be enormous. Therefore, instead of discussing typical economic examples

Presented at the International Symposium on Foundations of Statistical Inference, December 1985, Tel Aviv, Israel. This paper evolved from a series of lectures given in June 1985 at the University of Canterbury, Christchurch, New Zealand. I am grateful to Yoel Haitovsky and Richard Manning for providing me with the opportunity to discuss these ideas in such marvelous settings.

where assumptions are always controversial, I shall go to the other extreme and discuss two very simple, almost trivial, examples of statistical inference where the assumptions are quite conventional yet the inferences could naturally be called incredible. Although the examples have nothing to do with economics, I hope to persuade the reader that the key problems with econometric inference are illuminated by their analysis.

2. EXAMPLE ONE: A MEASUREMENT PROBLEM

In order to learn the dimensions of a rectangular table, I ask my research assistant to measure its length and width a number of times. The measuring device is imperfect, so the measurements do not yield the exact length and width. I believe, however, that the measurement errors behave like unpredictable random noise, with any particular error having equal probability of being positive or negative. Therefore, I decide to treat the measurement errors as independent, identically distributed random variables, each with median zero. In addition, I assume that the common error distribution is symmetric and possesses finite fourth moment. For example, the normal probability curve (truncated to insure the measurements are positive) might serve as an approximate model for the error distribution.

These assumptions would not usually be called incredible. They might not be valid for every measurement situation, but they could be reasonable for many such situations. (One might worry about my ruling out thick-tailed distributions that could capture the effects of gross measurement errors. I do that to simplify my story; the analysis could be conducted using medians rather than means, but only with harder distribution theory.) Now I shall make one further assumption. My research assistant mistakenly thinks I care only about the *area* of the table and hence multiplies the length and width measurements. Instead of receiving n length measurements $L_1, L_2, \ldots, L_n$ and n width measurements $W_1, W_2, \ldots, W_n$, I get only n area measurements $A_1 = L_1 W_1, A_2 = L_2 W_2, \ldots, A_n = L_n W_n$. Worse yet, my research assistant throws away the original data so they are lost forever.

Can I get reasonable estimates of the true length and width of the table using only these area measurements? Can I salvage anything from this badly reported experiment? If there were no measurement error, the answer is clearly no; I will learn the true area of the table, but there are an infinity of length and width pairs that are consistent with any given area. Length and width are simply not identifiable in this experiment. In the presence of measurement error, the answer is quite different. Both length and width are identifiable and can be well estimated from a moderately large sample. In this case credible assumptions seem to lead us to incredible inference!

To demonstrate that inference about length and width is possible, some notation will prove useful. Suppose α is the true length of the table and β is the true width. Let u_i be the error in the ith length measurement, let v_i be the error in the ith width measurement, and let σ^2 be the common error variance. Then we can write

$$A_i = \alpha\beta + \alpha v_i + \beta u_i + u_i v_i. \tag{1.1}$$

Given the assumption that u_i and v_i are independent random variables distributed symmetrically about zero and possessing third moments, we find:

$$E[A_i] = \alpha\beta, \quad \text{Var}\,[A_i] = \sigma^2(\alpha^2 + \beta^2 + \sigma^2)$$
$$E(A_i - \alpha\beta)^3 = 6\alpha\beta\sigma^4.$$

By convention, $\alpha \geq \beta > 0$. Simple algebra demonstrates that the three population moments uniquely determine the three parameters α, β, and σ^2. Furthermore, under our assumptions, the sample moments converge in probability to the population moments as n tends to infinity. Denoting the sample mean of the area measurements by M_1, the sample variance by M_2, and the sample third central moment by M_3, a natural method of moments estimator of σ^4 is $M_3/6M_1$. Assuming this is positive and denoting its square root by S, we can estimate $(\alpha+\beta)^2$ by the equation

$$(\alpha + \beta)^2 = \frac{M_2}{S} - S + 2M_1. \tag{1.2}$$

If $\sigma^2 > 0$, the probability that both estimates are positive goes to 1 as n tends to infinity. Define A to be the square root of expression (1.2) if real, and zero otherwise. Then A is a consistent estimate of $\alpha + \beta$. A natural estimate of $(\alpha - \beta)^2$ is

$$(\alpha - \beta)^2 = \frac{M_2}{S} - S - 2M_1. \tag{1.3}$$

If this expression is positive, its square root is a consistent estimate of $\alpha - \beta$. However, if the table is almost square, a negative value for (1.3) is quite likely. Define B to be the square root of expression (1.3) if real, and zero otherwise. Then $(A + B)/2$ and $(A - B)/2$ should be reasonable estimates of α and β.

These method of moments estimates will converge in probability to the true values as long as there is some measurement error. Central limit theory can be employed to develop large sample approximations of their sampling distributions. These approximate distributions are typically normal, although things get slightly more complicated when the table is square (because then the length and width estimates are confounded). To avoid this technical problem in the asymptotic distribution theory, I shall continue the discussion using $\alpha + \beta$ as the parameter of interest and A as my estimate. The essential feature of my example – that the parameter is estimable in the presence of measurement error but not otherwise – is unchanged.

If $\sigma > 0$ and the errors possess finite sixth moments, then the standardized estimator $\sqrt{n}(A - \alpha - \beta)$ converges in distribution to a zero-mean normal