Semiparametric Regression for the Applied Econometrician

This book provides an accessible collection of techniques for analyzing nonparametric and semiparametric regression models. Worked examples include estimation of Engel curves and equivalence scales; scale economies; semiparametric Cobb–Douglas, translog, and CES cost functions; household gasoline consumption; hedonic housing prices; and, option prices and state price density estimation. The book should be of interest to a broad range of economists, including those working in industrial organization, labor, development, urban, energy, and financial economics.

A variety of testing procedures are covered such as simple goodness-of-fit tests and residual regression tests. These procedures can be used to test hypotheses such as parametric and semiparametric specifications, significance, monotonicity, and additive separability. Other topics include endogeneity of parametric and nonparametric effects as well as heteroskedasticity and autocorrelation in the residuals. Bootstrap procedures are provided.

Adonis Yatchew teaches economics at the University of Toronto. His principal areas of research are theoretical and applied econometrics. In addition, he has a strong interest in regulatory and energy economics and is Joint Editor of the *Energy Journal*. He has received the social science undergraduate teaching award at the University of Toronto and has taught at the University of Chicago.

Further Praise for *Semiparametric Regression for the Applied Econometrician*

"This fluent book is an excellent source for learning, or updating one's knowledge of semi- and nonparametric methods and their applications. It is a valuable addition to the existent books on these topics."

– Rosa Matzkin, *Northwestern University*

"Yatchew's book is an excellent account of semiparametric regression. The material is nicely integrated by using a simple set of ideas which exploit the impact of differencing and weighting operations on the data. The empirical applications are attractive and will be extremely helpful for those encountering this material for the first time."

– Adrian Pagan, *Australian National University*

"At the University of Toronto Adonis Yatchew is known for excellence in teaching. The key to this excellence is the succinct transparency of his exposition. At its best such exposition transcends the medium of presentation (either lecture or text). This monograph reflects the clarity of the author's thinking on the rapidly expanding fields of semiparametric and nonparametric analysis. Both students and researchers will appreciate the mix of theory and empirical application."

– Dale Poirier, *University of California, Irvine*

Themes in Modern Econometrics

Managing editor
PETER C.B. PHILLIPS, *Yale University*

Series editors
RICHARD J. SMITH, *University of Warwick*
ERIC GHYSELS, *University of North Carolina, Chapel Hill*

Themes in Modern Econometrics is designed to service the large and growing need for explicit teaching tools in econometrics. It will provide an organized sequence of textbooks in econometrics aimed squarely at the student population and will be the first series in the discipline to have this as its express aim. Written at a level accessible to students with an introductory course in econometrics behind them, each book will address topics or themes that students and researchers encounter daily. Although each book will be designed to stand alone as an authoritative survey in its own right, the distinct emphasis throughout will be on pedagogic excellence.

Titles in the series

Statistics and Econometric Models: Volumes 1 and 2
CHRISTIAN GOURIEROUX and ALAIN MONFORT
Translated by QUANG VUONG

Time Series and Dynamic Models
CHRISTIAN GOURIEROUX and ALAIN MONFORT
Translated and edited by GIAMPIERO GALLO

Unit Roots, Cointegration, and Structural Change
G.S. MADDALA and IN-MOO KIM

Generalized Method of Moments Estimation
Edited by LÁSZLÓ MÁTYÁS

Nonparametric Econometrics
ADRIAN PAGAN and AMAN ULLAH

Econometrics of Qualitative Dependent Variables
CHRISTIAN GOURIEROUX
Translated by PAUL B. KLASSEN

The Econometric Analysis of Seasonal Time Series
ERIC GHYSELS and DENISE R. OSBORN

SEMIPARAMETRIC REGRESSION FOR THE APPLIED ECONOMETRICIAN

ADONIS YATCHEW

University of Toronto

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
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2003

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

Yatchew, Adonis.
Semiparametric regression for the applied econometrician / Adonis Yatchew.
p. cm. – (Themes in modern econometrics)
Includes bibliographical references and index.
ISBN 0-521-81283-6 – ISBN 0-521-01226-0 (pbk.)
1. Econometrics. 2. Regression analysis. I. Title. II. Series.
HB139 .Y38 2003
330′.01′519536 – dc21 2002041002

ISBN 0 521 81283 6 hardback
ISBN 0 521 01226 0 paperback

To Marta, Tamara and Mark.
Your smiles are sunlight,
your laughter, the twinkling of stars.

Contents

List of Figures and Tables

Cambridge University Press
0521812836 - Semiparametric Regression for the Applied Econometrician
Adonis Yatchew
Frontmatter
More information

Preface

This book has been largely motivated by pedagogical interests. Nonparametric and semiparametric regression models are widely studied by theoretical econometricians but are much underused by applied economists. In comparison with the linear regression model $y = z\beta + \varepsilon$, semiparametric techniques are theoretically sophisticated and often require substantial programming experience.

Two natural extensions to the linear model that allow greater flexibility are the partial linear model $y = z\beta + f(x) + \varepsilon$, which adds a nonparametric function, and the index model $y = f(z\beta) + \varepsilon$, which applies a nonparametric function to the linear index $z\beta$. Together, these models and their variants comprise the most commonly used semiparametric specifications in the applied econometrics literature. A particularly appealing feature for economists is that these models permit the inclusion of multiple explanatory variables without succumbing to the "curse of dimensionality."

We begin by describing the idea of differencing, which provides a simple way to analyze the partial linear model because it allows one to remove the nonparametric effect $f(x)$ and to analyze the parametric portion of the model $z\beta$ as if the nonparametric portion were not there to begin with. Thus, one can draw not only on the reservoir of parametric human capital but one can also make use of existing software. By the end of the first chapter, the reader will be able to estimate the partial linear model and apply it to a real data set (the empirical example analyzes scale economies in electricity distribution using a semiparametric Cobb–Douglas specification).

Chapter 2 describes the broad contours of nonparametric and semiparametric regression modeling, the categorization of models, the "curse of dimensionality," and basic theoretical results.

Chapters 3 and 4 are devoted to smoothing and differencing, respectively. The techniques are reinforced by empirical examples on Engel curves, gasoline demand, the effect of weather on electricity demand, and semiparametric translog and CES cost function models. Methods that incorporate heteroskedasticity, autocorrelation, and endogeneity of right-hand-side variables are included.

Chapter 5 focuses on nonparametric functions of several variables. The example on hedonic pricing of housing attributes illustrates the benefits of nonparametric modeling of location effects.

Economic theory rarely prescribes a specific functional form. Typically, the implications of theory involve constraints such as monotonicity, concavity, homotheticity, separability, and so on. Chapter 6 begins by outlining two broad classes of tests of these and other properties: goodness-of-fit tests that compare restricted and unrestricted estimates of the residual variance, and residual regression tests that regress residuals from a restricted regression on all the explanatory variables to see whether there is anything left to be explained. Both of these tests have close relatives in the parametric world. The chapter then proceeds to constrained estimation, which is illustrated by an option pricing example.

Chapter 7 addresses the index model with an application to equivalence scale estimation using South African household survey data. Chapter 8 describes bootstrap techniques for various procedures described in earlier chapters.

A cornerstone of the pedagogical philosophy underlying this book is that the second best way to learn econometric techniques is to actually apply them. (The best way is to teach them.[1]) To this purpose, data and sample programs are available for the various examples and exercises at www.chass.utoronto.ca/~yatchew/. With the exception of constrained estimation of option prices, all code is in *S-Plus*.[2] The reader should be able to translate the code into other programs such as *Stata* easily enough.

By working through the examples and exercises,[3] the reader should be able to

- estimate nonparametric regression, partial linear, and index models;
- test various properties using large sample results and bootstrap techniques;
- estimate nonparametric models subject to constraints such as monotonicity and concavity.

Well-known references in the nonparametrics and semiparametrics literature include Härdle (1990), Stoker (1991), Bickel et al. (1993), Horowitz (1998),

[1] Each year I tell my students the apocryphal story of a junior faculty member complaining to a senior colleague of his inability to get through to his students. After repeating the same lecture to his class on three different occasions, he exclaims in exasperation "I am so disappointed. Today I thought I had finally gotten through to them. This time even *I* understood the material, and they still did not understand."

[2] Krause and Olson (1997) have provided a particularly pleasant introduction to *S-Plus*. See also Venables and Ripley (1994).

[3] Many of the examples and portions of the text draw upon previously published work, in particular, Yatchew (1997, 1998, 1999, 2000), Yatchew and Bos (1997), Yatchew and No (2001), and Yatchew, Sun, and Deri (2001). The permission for use of these materials is gratefully acknowledged.

and Pagan and Ullah (1999).[4] It is hoped that this book is worthy of being squeezed onto a nearby bookshelf by providing an applied approach with numerical examples and adaptable code. It is intended for the applied economist and econometrician working with cross-sectional or possibly panel data.[5] It is expected that the reader has had a good basic course in econometrics and is thoroughly familiar with estimation and testing of the linear model and associated ideas such as heteroskedasticity and endogeneity. Some knowledge of nonlinear regression modeling and inference is desirable but not essential. Given the presence of empirical examples, the book could be used as a text in an advanced undergraduate course and certainly at the graduate level.

I owe a great intellectual debt to too many to name them individually, and regrettably not all of them appear in the references. Several anonymous reviewers provided extensive and valuable comments for which I am grateful. Thanks are also due to Scott Parris at Cambridge University Press for his unflagging efforts in this endeavor. My sister Oenone kindly contributed countless hours of proofreading time. Finally, it is indeed a special privilege to thank Peter Phillips, whose intellectual guidance shaped several aspects of this book. It was Peter who from the start insisted on reproducible empirical exercises. Those who are acquainted with both of us surely know to whom the errors belong.

[4] There are also several surveys: Delgado and Robinson (1992), Härdle and Linton (1994), Powell (1994), Linton (1995a), and Yatchew (1998). See also DiNardo and Tobias (2001).

[5] With the exception of correlation in the residuals, time-dependent data issues have not been covered here.