

## **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**

Cambridge University Press  
 0521812836 - Semiparametric Regression for the Applied Econometrician  
 Adonis Yatchew  
 Frontmatter  
[More information](#)

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>

© Adonis Yatchew 2003

This book is in copyright. Subject to statutory exception  
 and to the provisions of relevant collective licensing agreements,  
 no reproduction of any part may take place without  
 the written permission of Cambridge University Press.

First published 2003

Printed in the United States of America

*Typeface* Times Roman 10/12 pt.    *System* L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> [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

<i>List of Figures and Tables</i>	<i>page</i> xv
<i>Preface</i>	xvii
<b>1 Introduction to Differencing</b>	<b>1</b>
1.1 A Simple Idea	1
1.2 Estimation of the Residual Variance	2
1.3 The Partial Linear Model	2
1.4 Specification Test	4
1.5 Test of Equality of Regression Functions	4
1.6 Empirical Application: Scale Economies in Electricity Distribution	7
1.7 Why Differencing?	8
1.8 Empirical Applications	11
1.9 Notational Conventions	12
1.10 Exercises	12
<b>2 Background and Overview</b>	<b>15</b>
2.1 Categorization of Models	15
2.2 The Curse of Dimensionality and the Need for Large Data Sets	17
2.2.1 Dimension Matters	17
2.2.2 Restrictions That Mitigate the Curse	17
2.3 Local Averaging Versus Optimization	19
2.3.1 Local Averaging	19
2.3.2 Bias-Variance Trade-Off	19
2.3.3 Naive Optimization	22
2.4 A Bird's-Eye View of Important Theoretical Results	23
2.4.1 Computability of Estimators	23
2.4.2 Consistency	23
2.4.3 Rate of Convergence	23
	ix

x	<b>Contents</b>	
	2.4.4	Bias-Variance Trade-Off 25
	2.4.5	Asymptotic Distributions of Estimators 26
	2.4.6	How Much to Smooth 26
	2.4.7	Testing Procedures 26
<b>3</b>	<b>Introduction to Smoothing</b>	<b>27</b>
3.1	A Simple Smoother	27
3.1.1	The Moving Average Smoother	27
3.1.2	A Basic Approximation	28
3.1.3	Consistency and Rate of Convergence	29
3.1.4	Asymptotic Normality and Confidence Intervals	29
3.1.5	Smoothing Matrix	30
3.1.6	Empirical Application: Engel Curve Estimation	30
3.2	Kernel Smoothers	32
3.2.1	Estimator	32
3.2.2	Asymptotic Normality	34
3.2.3	Comparison to Moving Average Smoother	35
3.2.4	Confidence Intervals	35
3.2.5	Uniform Confidence Bands	36
3.2.6	Empirical Application: Engel Curve Estimation	37
3.3	Nonparametric Least-Squares and Spline Smoothers	37
3.3.1	Estimation	37
3.3.2	Properties	39
3.3.3	Spline Smoothers	40
3.4	Local Polynomial Smoothers	40
3.4.1	Local Linear Regression	40
3.4.2	Properties	41
3.4.3	Empirical Application: Engel Curve Estimation	42
3.5	Selection of Smoothing Parameter	43
3.5.1	Kernel Estimation	43
3.5.2	Nonparametric Least Squares	44
3.5.3	Implementation	46
3.6	Partial Linear Model	47
3.6.1	Kernel Estimation	47
3.6.2	Nonparametric Least Squares	48
3.6.3	The General Case	48
3.6.4	Heteroskedasticity	50
3.6.5	Heteroskedasticity and Autocorrelation	51
3.7	Derivative Estimation	52
3.7.1	Point Estimates	52
3.7.2	Average Derivative Estimation	53
3.8	Exercises	54



Contents	xi
<b>4 Higher-Order Differencing Procedures</b>	<b>57</b>
4.1 Differencing Matrices	57
4.1.1 Definitions	57
4.1.2 Basic Properties of Differencing and Related Matrices	58
4.2 Variance Estimation	58
4.2.1 The $m$ th-Order Differencing Estimator	58
4.2.2 Properties	59
4.2.3 Optimal Differencing Coefficients	60
4.2.4 Moving Average Differencing Coefficients	61
4.2.5 Asymptotic Normality	62
4.3 Specification Test	63
4.3.1 A Simple Statistic	63
4.3.2 Heteroskedasticity	64
4.3.3 Empirical Application: Log-Linearity of Engel Curves	65
4.4 Test of Equality of Regression Functions	66
4.4.1 A Simplified Test Procedure	66
4.4.2 The Differencing Estimator Applied to the Pooled Data	67
4.4.3 Properties	68
4.4.4 Empirical Application: Testing Equality of Engel Curves	69
4.5 Partial Linear Model	71
4.5.1 Estimator	71
4.5.2 Heteroskedasticity	72
4.6 Empirical Applications	73
4.6.1 Household Gasoline Demand in Canada	73
4.6.2 Scale Economies in Electricity Distribution	76
4.6.3 Weather and Electricity Demand	81
4.7 Partial Parametric Model	83
4.7.1 Estimator	83
4.7.2 Empirical Application: CES Cost Function	84
4.8 Endogenous Parametric Variables in the Partial Linear Model	85
4.8.1 Instrumental Variables	85
4.8.2 Hausman Test	86
4.9 Endogenous Nonparametric Variable	87
4.9.1 Estimation	87
4.9.2 Empirical Application: Household Gasoline Demand and Price Endogeneity	88
4.10 Alternative Differencing Coefficients	89
4.11 The Relationship of Differencing to Smoothing	90

xii	<b>Contents</b>	
4.12	Combining Differencing and Smoothing	92
4.12.1	Modular Approach to Analysis of the Partial Linear Model	92
4.12.2	Combining Differencing Procedures in Sequence	92
4.12.3	Combining Differencing and Smoothing	93
4.12.4	Reprise	94
4.13	Exercises	94
<b>5</b>	<b>Nonparametric Functions of Several Variables</b>	<b>99</b>
5.1	Smoothing	99
5.1.1	Introduction	99
5.1.2	Kernel Estimation of Functions of Several Variables	99
5.1.3	Loess	101
5.1.4	Nonparametric Least Squares	101
5.2	Additive Separability	102
5.2.1	Backfitting	102
5.2.2	Additively Separable Nonparametric Least Squares	103
5.3	Differencing	104
5.3.1	Two Dimensions	104
5.3.2	Higher Dimensions and the Curse of Dimensionality	105
5.4	Empirical Applications	107
5.4.1	Hedonic Pricing of Housing Attributes	107
5.4.2	Household Gasoline Demand in Canada	107
5.5	Exercises	110
<b>6</b>	<b>Constrained Estimation and Hypothesis Testing</b>	<b>111</b>
6.1	The Framework	111
6.2	Goodness-of-Fit Tests	113
6.2.1	Parametric Goodness-of-Fit Tests	113
6.2.2	Rapid Convergence under the Null	114
6.3	Residual Regression Tests	115
6.3.1	Overview	115
6.3.2	$U$ -statistic Test – Scalar $x$ 's, Moving Average Smoother	116
6.3.3	$U$ -statistic Test – Vector $x$ 's, Kernel Smoother	117
6.4	Specification Tests	119
6.4.1	Bierens (1990)	119
6.4.2	Härdle and Mammen (1993)	120
6.4.3	Hong and White (1995)	121
6.4.4	Li (1994) and Zheng (1996)	122
6.5	Significance Tests	124

<b>Contents</b>	xiii
6.6 Monotonicity, Concavity, and Other Restrictions	125
6.6.1 Isotonic Regression	125
6.6.2 Why Monotonicity Does Not Enhance the Rate of Convergence	126
6.6.3 Kernel-Based Algorithms for Estimating Monotone Regression Functions	127
6.6.4 Nonparametric Least Squares Subject to Monotonicity Constraints	127
6.6.5 Residual Regression and Goodness-of-Fit Tests of Restrictions	128
6.6.6 Empirical Application: Estimation of Option Prices	129
6.7 Conclusions	134
6.8 Exercises	136
<b>7 Index Models and Other Semiparametric Specifications</b>	<b>138</b>
7.1 Index Models	138
7.1.1 Introduction	138
7.1.2 Estimation	138
7.1.3 Properties	139
7.1.4 Identification	140
7.1.5 Empirical Application: Engel's Method for Estimation of Equivalence Scales	140
7.1.6 Empirical Application: Engel's Method for Multiple Family Types	142
7.2 Partial Linear Index Models	144
7.2.1 Introduction	144
7.2.2 Estimation	146
7.2.3 Covariance Matrix	147
7.2.4 Base-Independent Equivalence Scales	148
7.2.5 Testing Base-Independence and Other Hypotheses	149
7.3 Exercises	151
<b>8 Bootstrap Procedures</b>	<b>154</b>
8.1 Background	154
8.1.1 Introduction	154
8.1.2 Location Scale Models	155
8.1.3 Regression Models	156
8.1.4 Validity of the Bootstrap	157
8.1.5 Benefits of the Bootstrap	157
8.1.6 Limitations of the Bootstrap	159
8.1.7 Summary of Bootstrap Choices	159
8.1.8 Further Reading	160

xiv	<b>Contents</b>	
8.2	Bootstrap Confidence Intervals for Kernel Smoothers	160
8.3	Bootstrap Goodness-of-Fit and Residual Regression Tests	163
	8.3.1 Goodness-of-Fit Tests	163
	8.3.2 Residual Regression Tests	164
8.4	Bootstrap Inference in Partial Linear and Index Models	166
	8.4.1 Partial Linear Models	166
	8.4.2 Index Models	166
8.5	Exercises	171
	<b>Appendixes</b>	
	Appendix A – Mathematical Preliminaries	173
	Appendix B – Proofs	175
	Appendix C – Optimal Differencing Weights	183
	Appendix D – Nonparametric Least Squares	187
	Appendix E – Variable Definitions	194
	<i>References</i>	197
	<i>Index</i>	209

## List of Figures and Tables

Figure 1.1.	Testing equality of regression functions.	page 6
Figure 1.2.	Partial linear model – log-linear cost function: Scale economies in electricity distribution.	9
Figure 2.1.	Categorization of regression functions.	16
Figure 2.2.	Naive local averaging.	20
Figure 2.3.	Bias-variance trade-off.	21
Figure 2.4.	Naive nonparametric least squares.	24
Figure 3.1.	Engel curve estimation using moving average smoother.	31
Figure 3.2.	Alternative kernel functions.	33
Figure 3.3.	Engel curve estimation using kernel estimator.	38
Figure 3.4.	Engel curve estimation using kernel, spline, and lowess estimators.	42
Figure 3.5.	Selection of smoothing parameters.	45
Figure 3.6.	Cross-validation of bandwidth for Engel curve estimation.	46
Figure 4.1.	Testing linearity of Engel curves.	65
Figure 4.2.	Testing equality of Engel curves.	70
Figure 4.3.	Household demand for gasoline.	74
Figure 4.4.	Household demand for gasoline: Monthly effects.	75
Figure 4.5.	Scale economies in electricity distribution.	77
Figure 4.6.	Scale economies in electricity distribution: PUC and non-PUC analysis.	79
Figure 4.7.	Weather and electricity demand.	82
Figure 5.1.	Hedonic prices of housing attributes.	108
Figure 5.2.	Household gasoline demand in Canada.	109
Figure 6.1.	Constrained and unconstrained estimation and testing.	113
Figure 6.2A.	Data and estimated call function.	131
Figure 6.2B.	Estimated first derivative.	132
Figure 6.2C.	Estimated SPDs.	133
Figure 6.3.	Constrained estimation – simulated expected mean- squared error.	135

xvi	<b>List of Figures and Tables</b>	
Figure 7.1.	Engel's method for estimating equivalence scales.	141
Figure 7.2.	Parsimonious version of Engel's method.	144
Figure 8.1.	Percentile bootstrap confidence intervals for Engel curves.	162
Figure 8.2.	Equivalence scale estimation for singles versus couples: Asymptotic versus bootstrap methods.	170
Table 3.1.	Asymptotic confidence intervals for kernel estimators – implementation.	36
Table 4.1.	Optimal differencing weights.	61
Table 4.2.	Values of $\delta$ for alternate differencing coefficients.	62
Table 4.3.	Mixed estimation of PUC/non-PUC effects: Scale economies in electricity distribution.	80
Table 4.4.	Scale economies in electricity distribution: CES cost function.	85
Table 4.5.	Symmetric optimal differencing weights.	90
Table 4.6.	Relative efficiency of alternative differencing sequences.	90
Table 5.1.	The backfitting algorithm.	103
Table 6.1.	Bierens (1990) specification test – implementation.	120
Table 6.2.	Härdle and Mammen (1993) specification test – implementation.	122
Table 6.3.	Hong and White (1995) specification test – implementation.	123
Table 6.4.	Li (1994), Zheng (1996) residual regression test of specification – implementation.	123
Table 6.5.	Residual regression test of significance – implementation.	125
Table 7.1.	Distribution of family composition.	143
Table 7.2.	Parsimonious model estimates.	145
Table 8.1.	Wild bootstrap.	157
Table 8.2.	Bootstrap confidence intervals at $f(x_o)$ .	161
Table 8.3.	Bootstrap goodness-of-fit tests.	164
Table 8.4.	Bootstrap residual regression tests.	165
Table 8.5.	Percentile- $t$ bootstrap confidence intervals for $\beta$ in the partial linear model.	167
Table 8.6.	Asymptotic versus bootstrap confidence intervals: Scale economies in electricity distribution.	168
Table 8.7.	Confidence intervals for $\delta$ in the index model: Percentile method.	169

## 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.

xviii **Preface**

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.<sup>1</sup>) To this purpose, data and sample programs are available for the various examples and exercises at [www.chass.utoronto.ca/~yatchew/](http://www.chass.utoronto.ca/~yatchew/). With the exception of constrained estimation of option prices, all code is in *S-Plus*.<sup>2</sup> The reader should be able to translate the code into other programs such as *Stata* easily enough.

By working through the examples and exercises,<sup>3</sup> 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),

<sup>1</sup> 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.”

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

<sup>3</sup> 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).<sup>4</sup> 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.<sup>5</sup> 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.

<sup>4</sup> 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).

<sup>5</sup> With the exception of correlation in the residuals, time-dependent data issues have not been covered here.