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

Contents

List of Figures and Tables — *page* xv
Preface — xvii

1 Introduction to Differencing — **1**
- 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

2 Background and Overview — **15**
- 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

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

3 Introduction to Smoothing 27
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

4 Higher-Order Differencing Procedures **57**
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

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

5 Nonparametric Functions of Several Variables 99
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

6 Constrained Estimation and Hypothesis Testing 111
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

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

7 Index Models and Other Semiparametric Specifications 138
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

8 Bootstrap Procedures 154
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

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

Appendixes
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

References 197
Index 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

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 δ 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 β 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 δ in the index model: Percentile method. 169

1 Introduction to Differencing

1.1 A Simple Idea

Consider the nonparametric regression model

$$y = f(x) + \varepsilon \tag{1.1.1}$$

for which little is assumed about the function f except that it is smooth. In its simplest incarnation, the residuals are independently and identically distributed with mean zero and constant variance σ_ε^2, and the x's are generated by a process that ensures they will eventually be dense in the domain. Closeness of the x's combined with smoothness of f provides a basis for estimation of the regression function. By averaging or smoothing observations on y for which the corresponding x's are close to a given point, say x_o, one obtains a reasonable estimate of the regression effect $f(x_o)$.

This premise – that x's that are close will have corresponding values of the regression function that are close – may also be used to remove the regression effect. It is this removal or *differencing* that provides a simple exploratory tool. To illustrate the idea we present four applications:

1. Estimation of the residual variance σ_ε^2,
2. Estimation and inference in the partial linear model $y = z\beta + f(x) + \varepsilon$,
3. A specification test on the regression function f, and
4. A test of equality of nonparametric regression functions.[1]

[1] The first-order differencing estimator of the residual variance in a nonparametric setting appears in Rice (1984). Although unaware of his result at the time, I presented the identical estimator at a conference held at the IC2 Institute at the University of Texas at Austin in May 1984. Differencing subsequently appeared in a series of nonparametric and semiparametric settings, including Powell (1987), Yatchew (1988), Hall, Kay, and Titterington (1990), Yatchew (1997, 1998, 1999, 2000), Lewbel (2000), Fan and Huang (2001), and Horowitz and Spokoiny (2001).

1.2 Estimation of the Residual Variance

Suppose one has data $(y_1, x_1), \ldots, (y_n, x_n)$ on the pure nonparametric regression model (1.1.1), where x is a bounded scalar lying, say, in the unit interval, ε is i.i.d. with $E(\varepsilon \mid x) = 0$, $Var(\varepsilon \mid x) = \sigma_\varepsilon^2$, and all that is known about f is that its first derivative is bounded. Most important, the data have been rearranged so that $x_1 \leq \cdots \leq x_n$. Consider the following estimator of σ_ε^2:

$$s_{diff}^2 = \frac{1}{2n} \sum_{i=2}^{n} (y_i - y_{i-1})^2. \tag{1.2.1}$$

The estimator is consistent because, as the x's become close, differencing tends to remove the nonparametric effect $y_i - y_{i-1} = f(x_i) - f(x_{i-1}) + \varepsilon_i - \varepsilon_{i-1} \cong \varepsilon_i - \varepsilon_{i-1}$, so that[2]

$$s_{diff}^2 \cong \frac{1}{2n} \sum_{i=2}^{n} (\varepsilon_i - \varepsilon_{i-1})^2 \cong \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 - \frac{1}{n} \sum_{i=2}^{n} \varepsilon_i \varepsilon_{i-1}. \tag{1.2.2}$$

An obvious advantage of s_{diff}^2 is that no initial estimate of the regression function f needs to be calculated. Indeed, no consistent estimate of f is implicit in (1.2.1). Nevertheless, the terms in s_{diff}^2 that involve f converge to zero sufficiently quickly so that the asymptotic distribution of the estimator can be derived directly from the approximation in (1.2.2). In particular,

$$n^{1/2}\left(s_{diff}^2 - \sigma_\varepsilon^2\right) \xrightarrow{D} N(0, E(\varepsilon^4)). \tag{1.2.3}$$

Moreover, derivation of this result is facilitated by the assumption that the ε_i are independent so that reordering of the data does not affect the distribution of the right-hand side in (1.2.2).

1.3 The Partial Linear Model

Consider now the partial linear model $y = z\beta + f(x) + \varepsilon$, where for simplicity all variables are assumed to be scalars. We assume that $E(\varepsilon \mid z, x) = 0$ and that $Var(\varepsilon \mid z, x) = \sigma_\varepsilon^2$.[3] As before, the x's have bounded support, say the unit interval, and have been rearranged so that $x_1 \leq \cdots \leq x_n$. Suppose that the conditional mean of z is a smooth function of x, say $E(z \mid x) = g(x)$ where g' is

[2] To see why this approximation works, suppose that the x_i are equally spaced on the unit interval and that $f' \leq L$. By the mean value theorem, for some $x_i^* \in [x_{i-1}, x_i]$ we have $f(x_i) - f(x_{i-1}) = f'(x_i^*)(x_i - x_{i-1}) \leq L/n$. Thus, $y_i - y_{i-1} = \varepsilon_i - \varepsilon_{i-1} + O(1/n)$. For detailed development of the argument, see Exercise 1. If the x_i have a density function bounded away from zero on the support, then $x_i - x_{i-1} \cong O_P(1/n)$ and $y_i - y_{i-1} \cong \varepsilon_i - \varepsilon_{i-1} + O_P(1/n)$. See Appendix B, Lemma B.2, for a related result.

[3] For extensions to the heteroskedastic and autocorrelated cases, see Sections 3.6 and 4.5.

bounded and $Var(z \mid x) = \sigma_u^2$. Then we may rewrite $z = g(x) + u$. Differencing yields

$$\begin{aligned} y_i - y_{i-1} &= (z_i - z_{i-1})\beta + (f(x_i) - f(x_{i-1})) + \varepsilon_i - \varepsilon_{i-1} \\ &= (g(x_i) - g(x_{i-1}))\beta + (u_i - u_{i-1})\beta \\ &\quad + (f(x_i) - f(x_{i-1})) + \varepsilon_i - \varepsilon_{i-1} \\ &\cong (u_i - u_{i-1})\beta + \varepsilon_i - \varepsilon_{i-1}. \end{aligned} \tag{1.3.1}$$

Thus, the direct effect $f(x)$ of the nonparametric variable x and the indirect effect $g(x)$ that occurs through z are removed. Suppose we apply the OLS estimator of β to the differenced data, that is,

$$\hat{\beta}_{diff} = \frac{\sum (y_i - y_{i-1})(z_i - z_{i-1})}{\sum (z_i - z_{i-1})^2}. \tag{1.3.2}$$

Then, substituting the approximations $z_i - z_{i-1} \cong u_i - u_{i-1}$ and $y_i - y_{i-1} \cong (u_i - u_{i-1})\beta + \varepsilon_i - \varepsilon_{i-1}$ into (1.3.2) and rearranging, we have

$$n^{1/2}(\hat{\beta}_{diff} - \beta) \cong \frac{n^{1/2}\frac{1}{n}\sum (\varepsilon_i - \varepsilon_{i-1})(u_i - u_{i-1})}{\frac{1}{n}\sum (u_i - u_{i-1})^2}. \tag{1.3.3}$$

The denominator converges to $2\,\sigma_u^2$, and the numerator has mean zero and variance $6\,\sigma_\varepsilon^2\sigma_u^2$. Thus, the ratio has mean zero and variance $6\,\sigma_\varepsilon^2\sigma_u^2/(2\sigma_u^2)^2 = 1.5\,\sigma_\varepsilon^2/\sigma_u^2$. Furthermore, the ratio may be shown to be approximately normal (using a finitely dependent central limit theorem). Thus, we have

$$n^{1/2}(\hat{\beta}_{diff} - \beta) \xrightarrow{D} N\left(0, \frac{1.5\,\sigma_\varepsilon^2}{\sigma_u^2}\right). \tag{1.3.4}$$

For the most efficient estimator, the corresponding variance in (1.3.4) would be $\sigma_\varepsilon^2/\sigma_u^2$ so the proposed estimator based on first differences has relative efficiency $2/3 = 1/1.5$. In Chapters 3 and 4 we will produce efficient estimators.

Now, in order to use (1.3.4) to perform inference, we will need consistent estimators of σ_ε^2 and σ_u^2. These may be obtained using

$$\begin{aligned} s_\varepsilon^2 &= \frac{1}{2n}\sum_{i=2}^{n}((y_i - y_{i-1}) - (z_i - z_{i-1})\hat{\beta}_{diff})^2 \\ &\cong \frac{1}{2n}\sum_{i=2}^{n}(\varepsilon_i - \varepsilon_{i-1})^2 \xrightarrow{P} \sigma_\varepsilon^2 \end{aligned} \tag{1.3.5}$$

and

$$s_u^2 = \frac{1}{2n}\sum_{i=2}^{n}(z_i - z_{i-1})^2 \cong \frac{1}{2n}\sum_{i=2}^{n}(u_i - u_{i-1})^2 \xrightarrow{P} \sigma_u^2. \tag{1.3.6}$$

The preceding procedure generalizes straightforwardly to models with multiple parametric explanatory variables.

1.4 Specification Test

Suppose, for example, one wants to test the null hypothesis that f is a linear function. Let s^2_{res} be the usual estimate of the residual variance obtained from a linear regression of y on x. If the linear model is correct, then s^2_{res} will be approximately equal to the average of the true squared residuals:

$$s^2_{res} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\gamma}_1 - \hat{\gamma}_2 x_i)^2 \cong \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2. \tag{1.4.1}$$

If the linear specification is incorrect, then s^2_{res} will overestimate the residual variance while s^2_{diff} in (1.2.1) will remain a consistent estimator, thus forming the basis of a test. Consider the test statistic

$$V = \frac{n^{1/2}\left(s^2_{res} - s^2_{diff}\right)}{s^2_{diff}}. \tag{1.4.2}$$

Equations (1.2.2) and (1.4.1) imply that the numerator of V is approximately equal to

$$n^{1/2}\frac{1}{n}\sum \varepsilon_i\,\varepsilon_{i-1} \xrightarrow{D} N\left(0, \sigma_\varepsilon^4\right). \tag{1.4.3}$$

Since s^2_{diff}, the denominator of V, is a consistent estimator of σ_ε^2, V is asymptotically $N(0,1)$ under H_0. (Note that this is a one-sided test, and one rejects for large values of the statistic.)

As we will see later, this test procedure may be used to test a variety of null hypotheses such as general parametric and semiparametric specifications, monotonicity, concavity, additive separability, and other constraints. One simply inserts the restricted estimator of the variance in (1.4.2). We refer to test statistics that compare restricted and unrestricted estimates of the residual variance as "goodness-of-fit" tests.

1.5 Test of Equality of Regression Functions

Suppose we are given data $(y_{A1}, x_{A1}), \ldots, (y_{An}, x_{An})$ and $(y_{B1}, x_{B1}), \ldots,$ (y_{Bn}, x_{Bn}) from two possibly different regression models A and B. Assume x is a scalar and that each data set has been reordered so that the x's are in increasing order. The basic models are

$$\begin{aligned} y_{Ai} &= f_A(x_{Ai}) + \varepsilon_{Ai} \\ y_{Bi} &= f_B(x_{Bi}) + \varepsilon_{Bi} \end{aligned} \tag{1.5.1}$$

where given the x's, the ε's have mean 0, variance σ_ε^2, and are independent within and between populations; f_A and f_B have first derivatives bounded. Using (1.2.1), define consistent "within" differencing estimators of the variance

$$
\begin{aligned}
s_A^2 &= \frac{1}{2n}\sum_i^n (y_{Ai} - y_{Ai-1})^2 \\
s_B^2 &= \frac{1}{2n}\sum_i^n (y_{Bi} - y_{Bi-1})^2.
\end{aligned}
\tag{1.5.2}
$$

As we will do frequently, we have dropped the subscript "*diff*". Now pool *all* the data and reorder so that the pooled x's are in increasing order: $(y_1^*, x_1^*), \ldots\ldots, (y_{2n}^*, x_{2n}^*)$. (Note that the pooled data have only one subscript.) Applying the differencing estimator once again, we have

$$
s_p^2 = \frac{1}{4n}\sum_j^{2n}\left(y_j^* - y_{j-1}^*\right)^2. \tag{1.5.3}
$$

The basic idea behind the test procedure is to compare the pooled estimator with the average of the within estimators. If $f_A = f_B$, then the within and pooled estimators are consistent and should yield similar estimates. If $f_A \neq f_B$, then the within estimators remain consistent, whereas the pooled estimator overestimates the residual variance, as may be seen in Figure 1.1.

To formalize this idea, define the test statistic

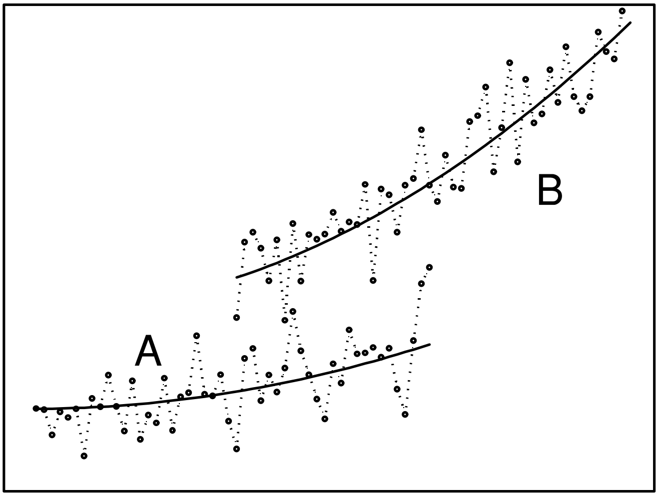
$$
\Upsilon \equiv (2n)^{1/2}\left(s_p^2 - 1/2\left(s_A^2 + s_B^2\right)\right). \tag{1.5.4}
$$

If $f_A = f_B$, then differencing removes the regression effect sufficiently quickly in both the within and the pooled estimators so that

$$
\begin{aligned}
\Upsilon &\equiv (2n)^{1/2}\left(s_p^2 - 1/2\left(s_A^2 + s_B^2\right)\right) \\
&\cong \frac{(2n)^{1/2}}{4n}\left(\sum_j^{2n}\left(\varepsilon_j^* - \varepsilon_{j-1}^*\right)^2 - \sum_i^n(\varepsilon_{Ai} - \varepsilon_{Ai-1})^2 - \sum_i^n(\varepsilon_{Bi} - \varepsilon_{Bi-1})^2\right) \\
&\cong \frac{(2n)^{1/2}}{2n}\left(\sum_j^{2n}\varepsilon_j^{*2} - \varepsilon_j^*\varepsilon_{j-1}^* - \sum_i^n\varepsilon_{Ai}^2 - \varepsilon_{Ai}\varepsilon_{Ai-1} - \sum_i^n\varepsilon_{Bi}^2 - \varepsilon_{Bi}\varepsilon_{Bi-1}\right) \\
&\cong \frac{1}{(2n)^{1/2}}\left(\sum_i^n\varepsilon_{Ai}\varepsilon_{Ai-1} + \sum_i^n\varepsilon_{Bi}\varepsilon_{Bi-1}\right) - \frac{1}{(2n)^{1/2}}\left(\sum_j^{2n}\varepsilon_j^*\varepsilon_{j-1}^*\right).
\end{aligned}
\tag{1.5.5}
$$

Consider the two terms in the last line. In large samples, each is approximately $N(0, \sigma_\varepsilon^4)$. If observations that are consecutive in the individual data

Within estimators of residual variance



Pooled estimator of residual variance

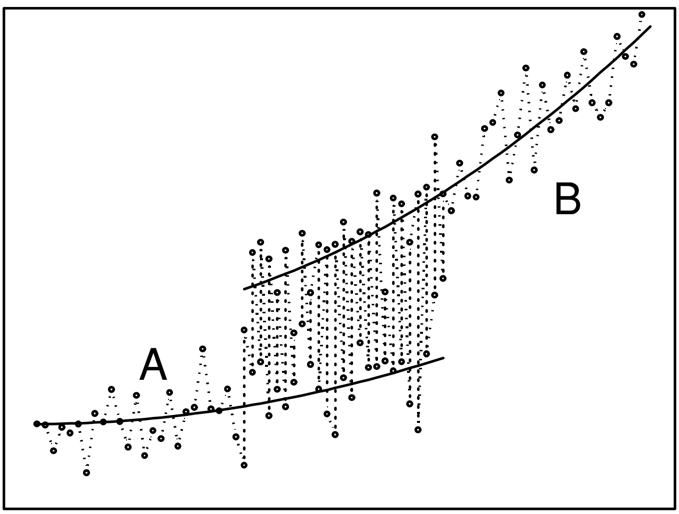


Figure 1.1. Testing equality of regression functions.

sets tend to be consecutive after pooling and reordering, then the *covariance* between the two terms will be large. In particular, the covariance is approximately $\sigma_\varepsilon^4(1-\pi)$, where π equals the probability that consecutive observations in the pooled reordered data set come from *different* populations.

It follows that under $H_o : f_A = f_B$,

$$\Upsilon \xrightarrow{D} N\left(0, 2\pi\sigma_\varepsilon^4\right). \tag{1.5.6}$$

For example, if reordering the pooled data is equivalent to stacking data sets A and B – because the two sets of x's, x_A and x_B, do not intersect – then $\pi \cong 0$ and indeed the statistic Υ becomes degenerate. This is not surprising, since observing nonparametric functions over different domains cannot provide a basis for testing whether they are the same. If the pooled data involve a simple interleaving of data sets A and B, then $\pi \cong 1$ and $\Upsilon \to N(0, 2\sigma_\varepsilon^4)$. If x_A and x_B are independent of each other but have the same distribution, then for the pooled reordered data the probability that consecutive observations come from different populations is ½ and $\Upsilon \to N(0, \sigma_\varepsilon^4)$.[4] To implement the test, one may obtain a consistent estimate $\hat{\pi}$ by taking the proportion of observations in the pooled reordered data that are preceded by an observation from a different population.

1.6 Empirical Application: Scale Economies in Electricity Distribution[5]

To illustrate these ideas, consider a simple variant of the Cobb–Douglas model for the costs of distributing electricity

$$\begin{aligned} tc = {} & f(cust) + \beta_1 wage + \beta_2 pcap \\ & + \beta_3 PUC + \beta_4 kwh + \beta_5 life + \beta_6 lf + \beta_7\, kmwire + \varepsilon \end{aligned} \tag{1.6.1}$$

where *tc* is the log of total cost per customer, *cust* is the log of the number of customers, *wage* is the log wage rate, *pcap* is the log price of capital, *PUC* is a dummy variable for public utility commissions that deliver additional services and therefore may benefit from economies of scope, *life* is the log of the remaining life of distribution assets, *lf* is the log of the load factor (this measures capacity utilization relative to peak usage), and *kmwire* is the log of kilometers of distribution wire per customer. The data consist of 81 municipal distributors in Ontario, Canada, during 1993. (For more details, see Yatchew, 2000.)

[4] For example, distribute n men and n women randomly along a stretch of beach facing the sunset. Then, for any individual, the probability that the person to the left is of the opposite sex is ½. More generally, if x_A and x_B are independent of each other and have different distributions, then π depends on the relative density of observations from each of the two populations.

[5] Variable definitions for empirical examples are contained in Appendix E.

Because the data have been reordered so that the nonparametric variable *cust* is in increasing order, first differencing (1.6.1) tends to remove the nonparametric effect f. We also divide by $\sqrt{2}$ so that the residuals in the differenced Equation (1.6.2) have the same variance as those in (1.6.1). Thus, we have

$$
\begin{aligned}
&[tc_i - tc_{i-1}]/\sqrt{2} \\
&\quad \cong \beta_1[wage_i - wage_{i-1}]/\sqrt{2} + \beta_2[pcap_i - pcap_{i-1}]/\sqrt{2} \\
&\qquad + \beta_3[PUC_i - PUC_{i-1}]/\sqrt{2} + \beta_4[kwh_i - kwh_{i-1}]/\sqrt{2} \\
&\qquad + \beta_5[life_i - life_{i-1}]/\sqrt{2} + \beta_6[lf_i - lf_{i-1}]/\sqrt{2} \\
&\qquad + \beta_7[kmwire_i - kmwire_{i-1}]/\sqrt{2} + [\varepsilon_i - \varepsilon_{i-1}]/\sqrt{2}. \qquad (1.6.2)
\end{aligned}
$$

Figure 1.2 summarizes our estimates of the parametric effects β using the differenced equation. It also contains estimates of a pure parametric specification in which the scale effect f is modeled with a quadratic. Applying the specification test (1.4.2), where s^2_{diff} is replaced with (1.3.5), yields a value of 1.50, indicating that the quadratic model may be adequate.

Thus far our results suggest that by differencing we can perform inference on β as if there were no nonparametric component f in the model to begin with. But, having estimated β, we can then proceed to apply a variety of nonparametric techniques to analyze f as if β were known. Such a modular approach simplifies implementation because it permits the use of existing software designed for pure nonparametric models.

More precisely, suppose we assemble the ordered pairs $(y_i - z_i\hat{\beta}_{diff}, x_i)$; then, we have

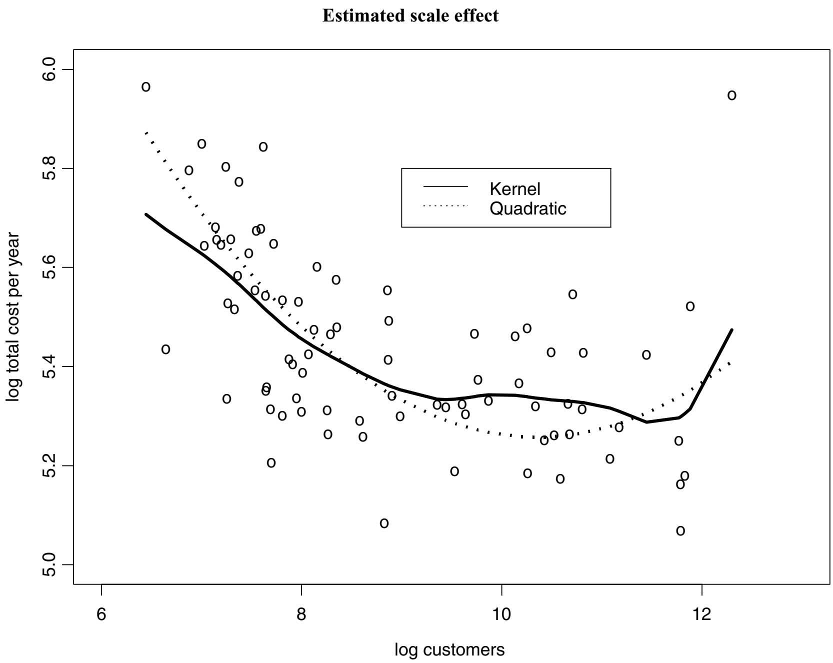
$$y_i - z_i\hat{\beta}_{diff} = z_i(\beta - \hat{\beta}_{diff}) + f(x_i) + \varepsilon_i \cong f(x_i) + \varepsilon_i. \qquad (1.6.3)$$

If we apply conventional smoothing methods to these ordered pairs such as kernel estimation (see Section 3.2), then consistency, optimal rate of convergence results, and the construction of confidence intervals for f remain valid because $\hat{\beta}_{diff}$ converges sufficiently quickly to β that the approximation in the last part of (1.6.3) leaves asymptotic arguments unaffected. (This is indeed why we could apply the specification test after removing the *estimated* parametric effect.) Thus, in Figure 1.2 we have also plotted a nonparametric (kernel) estimate of f that can be compared with the quadratic estimate. In subsequent sections, we will elaborate this example further and provide additional ones.

1.7 Why Differencing?

An important advantage of differencing procedures is their simplicity. Consider once again the partial linear model $y = z\beta + f(x) + \varepsilon$. Conventional

Variable	Quadratic model		Partial linear model[a]	
	Coef	SE	Coef	SE
cust	−0.833	0.175	–	–
cust^2	0.040	0.009	–	–
wage	0.833	0.325	0.448	0.367
pcap	0.562	0.075	0.459	0.076
PUC	−0.071	0.039	−0.086	0.043
kwh	−0.017	0.089	−0.011	0.087
life	−0.603	0.119	−0.506	0.131
lf	1.244	0.434	1.252	0.457
kmwire	0.445	0.086	0.352	0.094
s_ε^2	.021		.018	
R^2	.618		.675	



[a] Test of quadratic versus nonparametric specification of scale effect: $V = n^{1/2}(s_{res}^2 - s_{diff}^2)/s_{diff}^2 = 81^{1/2}(.021 - .018)/.018 = 1.5$, where V is $N(0,1)$, Section 1.4.

Figure 1.2. Partial linear model – Log-linear cost function: Scale economies in electricity distribution.

estimators, such as the one proposed by Robinson (1988) (see Section 3.6), require one to estimate $E(y \mid x)$ and $E(z \mid x)$ using nonparametric regressions. The estimated residuals from each of these regressions (hence the term "double residual method") are then used to estimate the *parametric* regression

$$y - E(y \mid x) = (z - E(z \mid x))\beta + \varepsilon. \tag{1.7.1}$$

If z is a vector, then a separate nonparametric regression is run for each component of z, where the independent variable is the nonparametric variable x. In contrast, differencing eliminates these first-stage regressions so that estimation of β can be performed – regardless of its dimension – even if nonparametric regression procedures are not available within the software being used. Similarly, tests of parametric specifications against nonparametric alternatives and tests of equality of regression functions across two or more (sub-) samples can be carried out without performing a nonparametric regression.

As should be evident from the empirical example of the last section, differencing may easily be combined with other procedures. In that example, we used differencing to estimate the parametric component of a partial linear model. We then removed the estimated parametric effect and applied conventional nonparametric procedures to analyze the nonparametric component. Such modular analysis does require theoretical justification, which we will provide in Section 4.12.

As we have seen, the partial linear model permits a simple semiparametric generalization of the Cobb–Douglas model. Translog and other linear-in-parameters models may be generalized similarly. If we allow the parametric portion of the model to be nonlinear – so that we have a partial parametric model – then we may also obtain simple semiparametric generalizations of models such as the constant elasticity of substitution (CES) cost function. These, too, may be estimated straightforwardly using differencing (see Section 4.7). The key requirement is that the parametric and nonparametric portions of the model be additively separable.

Other procedures commonly used by the econometrician may be imported into the differencing setting with relative ease. If some of the parametric variables are potentially correlated with the residuals, instrumental variable techniques can be applied, with suitable modification, as can the Hausman endogeneity test (see Section 4.8). If the residuals are potentially not homoskedastic, then well-known techniques such as White's heteroskedasticity-consistent standard errors can be adapted (see Section 4.5). The reader will no doubt find other procedures that can be readily transplanted.

Earlier we have pointed out that the first-order differencing estimator of β in the partial linear model is inefficient when compared with the most efficient estimator (see Section 1.3). The same is true for the first-order differencing estimator of the residual variance (see Section 1.2). This problem can be corrected using higher-order differencing, as demonstrated in Chapter 4.