Semiparametric Regression

Semiparametric regression is concerned with the flexible incorporation of nonlinear functional relationships in regression analyses. Any application area that uses regression analysis can benefit from semiparametric regression. Assuming only a basic familiarity with ordinary parametric regression, this user-friendly book explains the techniques and benefits of semiparametric regression in a concise and modular fashion. The authors make liberal use of graphics and examples plus case studies taken from environmental, financial, and other applications. They include practical advice on implementation and pointers to relevant software.

This book is suitable as a textbook for students with little background in regression as well as a reference book for statistically oriented scientists – such as biostatisticians, econometricians, quantitative social scientists, and epidemiologists – with a good working knowledge of regression and the desire to begin using more flexible semiparametric models. Even experts on semiparametric regression should find something new here.

David Ruppert is the Andrew Schultz, Jr., Professor of Engineering (School of Operations Research and Industrial Engineering) and Professor of Statistical Science at Cornell University. He has served as editor for a number of prestigious series and journals and has published some 80 articles of his own as well as co-authoring two popular books, *Transformation and Weighting in Regression* and *Measurement Error in Nonlinear Models*. He is also winner of the Wilcoxon Prize for best practical applications paper in technometrics and an elected Fellow of the American Statistical Association and the Institute of Mathematical Statistics.

M. P. Wand is Professor of Statistics at the University of New South Wales in Sydney, Australia. He has held faculty appointments at Harvard University, Rice University, and Texas A&M University. Dr. Wand is a Fellow of the American Statistical Association and has served as an associate editor for the *Journal of the American Statistical Association* and *Biometrika*. He is winner of the P. A. P. Moran Medal for statistical research.

R. J. Carroll is Distinguished Professor of Statistics, Nutrition and Toxicology at Texas A&M University. Among his many honors are the COPSS Presidents’ Award, the Fisher Lecture, the Snedecor Award, and the Wilcoxon Prize. He is an elected Fellow of the American Statistical Association and the Institute of Mathematical Statistics as well as an elected member of the International Statistical Institute.
CAMBRIDGE SERIES IN STATISTICAL AND PROBABILISTIC MATHEMATICS

Editorial Board

R. Gill (Department of Mathematics, Utrecht University)
B. D. Ripley (Department of Statistics, University of Oxford)
S. Ross (Department of Industrial Engineering, University of California, Berkeley)
M. Stein (Department of Statistics, University of Chicago)
D. Williams (School of Mathematical Sciences, University of Bath)

This series of high-quality upper-division textbooks and expository monographs covers all aspects of stochastic applicable mathematics. The topics range from pure and applied statistics to probability theory, operations research, optimization, and mathematical programming. The books contain clear presentations of new developments in the field and also of the state of the art in classical methods. While emphasizing rigorous treatment of theoretical methods, the books also contain applications and discussions of new techniques made possible by advances in computational practice.

Already published
1. Bootstrap Methods and Their Application, by A. C. Davison and D. V. Hinkley
2. Markov Chains, by J. Norris
3. Asymptotic Statistics, by A. W. van der Vaart
4. Wavelet Methods for Time Series Analysis, by Donald B. Percival and Andrew T. Walden
5. Bayesian Methods, by Thomas Leonard and John S. J. Hsu
9. The Estimation and Tracking of Frequency, by B. G. Quinn and E. J. Hannan
11. Statistical Models, by A. C. Davison
Semiparametric Regression

DAVID RUPPERT
Cornell University

M. P. WAND
Harvard University

R. J. CARROLL
Texas A&M University
To Anne, with love
—— David

To my wife’s parents, Ayhan and Recep
—— Matt

To Brett and Jeb
—— Raymond
Contents

Preface xiii
Guide to Notation xv

1 Introduction 1
1.1 Assessing the Carcinogenicity of Phenolphthalein 3
1.2 Salinity and Fishing in North Carolina 4
1.3 Management of a Retirement Fund 5
1.4 Biomonitoring of Airborne Mercury 7
1.5 Term Structure of Interest Rates 7
1.6 Air Pollution and Mortality in Milan: The Harvesting Effect 11

2 Parametric Regression 15
2.1 Introduction 15
2.2 Linear Regression Models 15
2.3 Regression Diagnostics 20
2.4 Inference 28
2.5 Parametric Additive Models 36
2.6 Model Selection 44
2.7 Polynomial Regression Models 46
2.8 Nonlinear Regression 48
2.9 Transformations in Regression 51
2.10 Bibliographic Notes 55
2.11 Summary of Formulas 55

3 Scatterplot Smoothing 57
3.1 Introduction 57
3.2 Preliminary Ideas 58
3.3 Practical Implementation 62
3.4 Automatic Knot Selection 64
3.5 Penalized Spline Regression 65
3.6 Quadratic Spline Bases 67
3.7 Other Spline Models and Bases 69
3.8 Other Penalties 74
3.9 General Definition of a Penalized Spline 75
3.10 Linear Smoothers 76
3.11 Error of a Smoother 76
## Contents

7.3 Semiparametric Binary Offset Model 162  
7.4 Additivity and Interactions 164  
7.5 General Parametric Component 164  
7.6 Inference 167  
7.7 Bibliographical Notes 168

8 Additive Models 170  
8.1 Introduction 170  
8.2 Fitting an Additive Model 171  
8.3 Degrees of Freedom 174  
8.4 Smoothing Parameter Selection 176  
8.5 Hypothesis Testing 181  
8.6 Model Selection 183  
8.7 Bibliographical Notes 185

9 Semiparametric Mixed Models 186  
9.1 Introduction 186  
9.2 Additive Mixed Models 186  
9.3 Subject-Specific Curves 191  
9.4 Bibliographical Notes 192

10 Generalized Parametric Regression 194  
10.1 Introduction 194  
10.2 Binary Response Data 194  
10.3 Logistic Regression 195  
10.4 Other Generalized Linear Models 197  
10.5 Iteratively Reweighted Least Squares 200  
10.6 Hat Matrix, Degrees of Freedom, and Standard Errors 201  
10.7 Overdispersion and Variance Functions: Pseudolikelihood 201  
10.8 Generalized Linear Mixed Models 203  
10.9 Deviance 209  
10.10 Technical Details 210  
10.11 Bibliographical Notes 213

11 Generalized Additive Models 214  
11.1 Introduction 214  
11.2 Generalized Scatterplot Smoothing 215  
11.3 Generalized Additive Mixed Models 217  
11.4 Degrees-of-Freedom Approximations 219  
11.5 Automatic Smoothing Parameter Selection 220  
11.6 Hypothesis Testing 220  
11.7 Model Selection 221  
11.8 Density Estimation 221  
11.9 Bibliographical Notes 222

12 Interaction Models 223  
12.1 Introduction 223  
12.2 Binary-by-Continuous Interaction Models 224
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.3</td>
<td>Factor-by-Curve Interactions in Additive Models</td>
<td>226</td>
</tr>
<tr>
<td>12.4</td>
<td>Varying Coefficient Models</td>
<td>234</td>
</tr>
<tr>
<td>12.5</td>
<td>Continuous-by-Continuous Interactions</td>
<td>235</td>
</tr>
<tr>
<td>12.6</td>
<td>Bibliographical Notes</td>
<td>237</td>
</tr>
<tr>
<td>13</td>
<td>Bivariate Smoothing</td>
<td>238</td>
</tr>
<tr>
<td>13.1</td>
<td>Introduction</td>
<td>238</td>
</tr>
<tr>
<td>13.2</td>
<td>Choice of Bivariate Basis Functions</td>
<td>240</td>
</tr>
<tr>
<td>13.3</td>
<td>Kriging</td>
<td>242</td>
</tr>
<tr>
<td>13.4</td>
<td>General Radial Smoothing</td>
<td>248</td>
</tr>
<tr>
<td>13.5</td>
<td>Default Automatic Bivariate Smoother</td>
<td>256</td>
</tr>
<tr>
<td>13.6</td>
<td>Geoadditive Models</td>
<td>258</td>
</tr>
<tr>
<td>13.7</td>
<td>Additive Plus Interaction Models</td>
<td>259</td>
</tr>
<tr>
<td>13.8</td>
<td>Generalized Bivariate Smoothing</td>
<td>259</td>
</tr>
<tr>
<td>13.9</td>
<td>Appendix: Equivalence of BLUP using $Z_R$ and $Z_P$</td>
<td>259</td>
</tr>
<tr>
<td>13.10</td>
<td>Bibliographical Notes</td>
<td>260</td>
</tr>
<tr>
<td>14</td>
<td>Variance Function Estimation</td>
<td>261</td>
</tr>
<tr>
<td>14.1</td>
<td>Introduction</td>
<td>261</td>
</tr>
<tr>
<td>14.2</td>
<td>Formulation</td>
<td>263</td>
</tr>
<tr>
<td>14.3</td>
<td>Application to the LIDAR Data</td>
<td>264</td>
</tr>
<tr>
<td>14.4</td>
<td>Quasilikelihood and Variance Functions</td>
<td>266</td>
</tr>
<tr>
<td>14.5</td>
<td>Bibliographical Notes</td>
<td>267</td>
</tr>
<tr>
<td>15</td>
<td>Measurement Error</td>
<td>268</td>
</tr>
<tr>
<td>15.1</td>
<td>Introduction</td>
<td>268</td>
</tr>
<tr>
<td>15.2</td>
<td>Formulation</td>
<td>269</td>
</tr>
<tr>
<td>15.3</td>
<td>The Expectation Maximization (EM) Algorithm</td>
<td>270</td>
</tr>
<tr>
<td>15.4</td>
<td>Simulated Example Revisited</td>
<td>273</td>
</tr>
<tr>
<td>15.5</td>
<td>Sensitivity Analysis Example</td>
<td>273</td>
</tr>
<tr>
<td>15.6</td>
<td>Bibliographical Notes</td>
<td>275</td>
</tr>
<tr>
<td>16</td>
<td>Bayesian Semiparametric Regression</td>
<td>276</td>
</tr>
<tr>
<td>16.1</td>
<td>Introduction</td>
<td>276</td>
</tr>
<tr>
<td>16.2</td>
<td>General Framework</td>
<td>277</td>
</tr>
<tr>
<td>16.3</td>
<td>Scatterplot Smoothing</td>
<td>279</td>
</tr>
<tr>
<td>16.4</td>
<td>Linear Mixed Models</td>
<td>285</td>
</tr>
<tr>
<td>16.5</td>
<td>Generalized Linear Mixed Models</td>
<td>288</td>
</tr>
<tr>
<td>16.6</td>
<td>Rao–Blackwellization</td>
<td>291</td>
</tr>
<tr>
<td>16.7</td>
<td>Bibliographical Notes</td>
<td>292</td>
</tr>
<tr>
<td>17</td>
<td>Spatially Adaptive Smoothing</td>
<td>293</td>
</tr>
<tr>
<td>17.1</td>
<td>Introduction</td>
<td>293</td>
</tr>
<tr>
<td>17.2</td>
<td>A Local Penalty Method</td>
<td>294</td>
</tr>
<tr>
<td>17.3</td>
<td>Completely Automatic Algorithm</td>
<td>295</td>
</tr>
<tr>
<td>17.4</td>
<td>Bayesian Inference</td>
<td>296</td>
</tr>
</tbody>
</table>
Preface

The primary aim of this book is to guide researchers needing to flexibly incorporate nonlinear relationships into their regression analyses. Flexible nonlinear regression is traditionally known as *nonparametric regression*; it differs from parametric regression in that the shape of the functional relationships are not predetermined but can adjust to capture unusual or unexpected features of the data.

Almost all existing regression texts treat either parametric or nonparametric regression exclusively. The level of exposition between books of either type differs quite alarmingly. In this book we argue that nonparametric regression can be viewed as a relatively simple extension of parametric regression and treat the two together. We refer to this combination as *semiparametric regression*. Our approach to semiparametric regression is based on penalized regression splines and mixed models. Indeed, every model in this book is a special case of the linear mixed model or its generalized counterpart. This makes the methodology modular and is in keeping with our general philosophy of *minimalist statistics* (see Section 19.2), where the amount of methodology, terminology, and so on is kept to a minimum. This is the first smoothing book that makes use of the mixed model representation of smoothers.

Unlike many other texts on nonparametric regression, this book is very much problem-driven. Examples from our collaborative research (and elsewhere) have driven the selection of material and emphases and are used throughout the book.

The book is suitable for several audiences. One audience consists of students or working scientists with only a moderate background in regression, though familiarity with matrix and linear algebra is assumed. Marginal notes and the appendices are intended for beginners, especially those from interface disciplines. We make liberal use of graphics because visualization is a particularly effective tool for acquiring intuition in a new subject.

Another audience that we are aiming at consists of statistically oriented scientists (e.g., biostatisticians, econometricians, quantitative social scientists, and epidemiologists) who have a good working knowledge of linear models and the desire to begin using more flexible semiparametric models. There are many connections between linear and nonparametric regression. Our goal is to exploit them and the reader’s knowledge of linear models to provide a foundation for understanding nonparametric modeling.

There is enough new material to be of interest even to experts on smoothing, and they are a third possible audience.
There are several competing approaches to nonparametric modeling: smoothing splines (e.g., Eubank 1988, 1999; Wahba 1990; Green and Silverman 1994); series-based smoothers, including wavelets (Tarter and Lock 1993; Ogden 1996); kernel methods, including local regression (Wand and Jones 1995; Fan and Gijbels 1996); and regression splines (Friedman 1991; Stone et al. 1997; Hansen and Kooperberg 2002). All four approaches can be used effectively and have their devotees. We believe that the nature of the data should play a role in the choice among them. For example, wavelets are more suited to highly oscillatory functions. Apart from this, the choice of a nonparametric regression method is a matter somewhat of individual taste and background. Based on our motivating applications and personal tastes, the approach to nonparametric regression used throughout this book is what we call penalized splines, although they are also labeled as P-splines, pseudosplines, and low-rank spline smoothers in the literature. Penalized splines are quite similar to smoothing splines; in fact, they are a generalization of smoothing splines that allow more flexible choices of the spline model, the basis functions for that model, and the penalty.

Penalized splines have close ties with ridge regression, mixed models, and Bayesian statistics, ties that were discovered by researchers working on smoothing splines. These ties allow techniques from mixed models – for example, (restricted) maximum likelihood estimation and likelihood ratio tests – to be added to penalized spline methodology. Similarly, Bayesian techniques based on Markov chain Monte Carlo provide what we believe to be the most satisfactory approach to fitting complex semiparametric models as well as the direction that semiparametric regression is most likely to take in the future. This book includes introductions to mixed models and to Bayesian modeling.

Acknowledgments

We are especially grateful to Ciprian Crainiceanu and Bhaswati Ganguli for their assistance in the preparation of this book. Ciprian wrote the WinBugs program in Appendix B and wrote the programs used for simulations-based p-values for likelihood ratio tests. Several other of our colleagues and collaborators have contributed to the book in various ways. We would like to thank Marc Aerts, Babette Brumback, Tianxi Cai, Gerda Claeskens, Brent Coull, Maria Durban, Garrett Fitzmaurice, Jonathan French, Robert Gentleman, Bob Gray, Nick Horton, Joe Ibrahim, Erin Kammann, Göran Kauermann, Robert Kohn, Nan Laird, Nick Lange, Mary Lindstrom, Long Ngo, Doug Nychka, Michael O’Connell, Helen Parise, José Pinheiro, Louise Ryan, Misha Salganik, Joel Schwartz, John Staudenmayer, Sally Thurston, Carrie Wager, Naisyin Wang, Jim Ware, Antonella Zanobetti, and Yihua Zhao for their collaboration, interest, and comments.

We thank Lauren Cowles for being a very supportive and patient editor. The second author lovingly acknowledges the support of his wife, Handan, and children, Declan and Jaida, throughout this project. Support of the Department of Biostatistics, Harvard University, is also gratefully acknowledged.
This chapter gives a brief overview of notational conventions used in the book. Please see the Notation Index for more specialized notation.

The symbol “≡” means “equal by definition”.

We use both lower- and uppercase letters (e.g., $x$, $X$, and $\lambda$) to denote scalar quantities, either fixed or random. Lowercase bold letters (e.g., $\mathbf{x}$ and $\lambda$) will be used for vectors. Uppercase bold fonts (e.g., $X$ and $\Lambda$) will denote matrices. The entries of a vector or matrix use the same letter and case as the vector or matrix itself but are not bold. Thus,

$$ \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} $$

and

$$ \mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}. $$

If a matrix is partitioned then the submatrices are in bold; for example,

$$ \mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}. $$

We will indicate the row index of a matrix to the right and the column index below, as in:

$$ \mathbf{C} = \begin{bmatrix} c_{ik} \\ 1 \leq k \leq K \\ 1 \leq i \leq n \end{bmatrix}. $$

The transpose of $\mathbf{A}$ is denoted by $\mathbf{A}^T$. If $\mathbf{A}$ is an invertible square matrix, then $\mathbf{A}^{-1}$ denotes its inverse. Any vector is assumed to be a column, so its transpose is a row.

The norm of a vector $\mathbf{x}$ is denoted by $\|\mathbf{x}\|$: that is,

$$ \|\mathbf{x}\| \equiv \sqrt{\mathbf{x}^T \mathbf{x}}. $$

The real line will be denoted by $\mathbb{R}$, and $d$-dimensional space will be denoted by $\mathbb{R}^d$.

For a function $f(x)$ of a scalar $x$,

$$ f^{(r)}(x) \equiv \left( \frac{d^r}{dx^r} \right) f(x), $$

the $r$th derivative of $f(x)$.
Guide to Notation

If $f(x)$ is a function from $\mathbb{R}^d$ to $\mathbb{R}$ then the derivative vector is a $1 \times d$ row vector with $j$th entry equal to $(\partial/\partial x_j)f(x)$, the partial derivative of $f(x)$ with respect to $x_j$, and is denoted by $Df(x)$.

The Hessian matrix is a $d \times d$ matrix whose $(i, j)$ entry is equal to

$$\frac{\partial^2}{\partial x_i \partial x_j} f(x);$$

it is denoted by $Hf(x)$.

If $x$ and $y$ are random variables, then $E(x)$, $\text{var}(x)$, and $\text{st.dev.}(x)$ are the mean, variance, and standard deviation of $x$, and $\text{cov}(x, y)$ is the covariance between $x$ and $y$. $\text{Cov}(x)$ is the covariance matrix of a random vector $x$; see Appendix A for its definition.