Monographs on Statistics and Applied Probability 121

Smoothing Splines

Methods and Applications

Yuedong Wang

University of California

Santa Barbara, California, USA

CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the

Taylor & Francis Group an informa business

A CHAPMAN & HALL BOOK

Contents

1 Introduction 1
- 1.1 Parametric and Nonparametric Regression 1
- 1.2 Polynomial Splines 4
- 1.3 Scope of This Book 7
- 1.4 The `assist` Package 9

2 Smoothing Spline Regression 11
- 2.1 Reproducing Kernel Hilbert Space 11
- 2.2 Model Space for Polynomial Splines 14
- 2.3 General Smoothing Spline Regression Models 16
- 2.4 Penalized Least Squares Estimation 17
- 2.5 The `ssr` Function 20
- 2.6 Another Construction for Polynomial Splines 22
- 2.7 Periodic Splines 24
- 2.8 Thin-Plate Splines 26
- 2.9 Spherical Splines 29
- 2.10 Partial Splines 30
- 2.11 *L*-splines 39
  - 2.11.1 Motivation 39
  - 2.11.2 Exponential Spline 41
  - 2.11.3 Logistic Spline 44
  - 2.11.4 Linear-Periodic Spline 46
  - 2.11.5 Trigonometric Spline 48

3 Smoothing Parameter Selection and Inference 53
- 3.1 Impact of the Smoothing Parameter 53
- 3.2 Trade-Offs 57
- 3.3 Unbiased Risk 62
- 3.4 Cross-Validation and Generalized Cross-Validation 64
- 3.5 Bayes and Linear Mixed-Effects Models 67
- 3.6 Generalized Maximum Likelihood 71
- 3.7 Comparison and Implementation 72
- 3.8 Confidence Intervals 75
  - 3.8.1 Bayesian Confidence Intervals 75
  - 3.8.2 Bootstrap Confidence Intervals 81

3.9 Hypothesis Tests 84
3.9.1 The Hypothesis 84
3.9.2 Locally Most Powerful Test 85
3.9.3 Generalized Maximum Likelihood Test 86
3.9.4 Generalized Cross-Validation Test 87
3.9.5 Comparison and Implementation 87

4 Smoothing Spline ANOVA 91
4.1 Multiple Regression 91
4.2 Tensor Product Reproducing Kernel Hilbert Spaces . . 92
4.3 One-Way SS ANOVA Decomposition 93
4.3.1 Decomposition of $\mathbb{R}^a$: One-Way ANOVA 95
4.3.2 Decomposition of $W_2^m[a,b]$ 96
4.3.3 Decomposition of $W_2^m(per)$ 97
4.3.4 Decomposition of $W_2^m(\mathbb{R}^d)$ 97
4.4 Two-Way SS ANOVA Decomposition 98
4.4.1 Decomposition of $\mathbb{R}^a \otimes \mathbb{R}^b$: Two-Way ANOVA . 99
4.4.2 Decomposition of $\mathbb{R}^a \otimes W_2^m[0,1]$ 100
4.4.3 Decomposition of $W_2^{m_1}[0,1] \otimes W_2^{m_2}[0,1]$ 103
4.4.4 Decomposition of $\mathbb{R}^a \otimes W_2^m(per)$ 106
4.4.5 Decomposition of $W_2^{m_1}(per) \otimes W_2^{m_2}[0,1]$ 107
4.4.6 Decomposition of $W_2^2(\mathbb{R}^2) \otimes W_2^m(per)$ 108
4.5 General SS ANOVA Decomposition 110
4.6 SS ANOVA Models and Estimation 111
4.7 Selection of Smoothing Parameters 114
4.8 Confidence Intervals 116
4.9 Examples 117
4.9.1 Tongue Shapes 117
4.9.2 Ozone in Arosa — Revisit 126
4.9.3 Canadian Weather — Revisit 131
4.9.4 Texas Weather 133

5 Spline Smoothing with Heteroscedastic and/or Correlated Errors 139
5.1 Problems with Heteroscedasticity and Correlation . . . 139
5.2 Extended SS ANOVA Models 142
5.2.1 Penalized Weighted Least Squares 142
5.2.2 UBR, GCV and GML Criteria 144
5.2.3 Known Covariance 147
5.2.4 Unknown Covariance 148
5.2.5 Confidence Intervals 150
5.3 Variance and Correlation Structures 150
5.4 Examples 153

5.4.1 Simulated Motorcycle Accident — Revisit 153
5.4.2 Ozone in Arosa — Revisit 154
5.4.3 Beveridge Wheat Price Index 157
5.4.4 Lake Acidity 158

6 Generalized Smoothing Spline ANOVA 163
6.1 Generalized SS ANOVA Models 163
6.2 Estimation and Inference 164
6.2.1 Penalized Likelihood Estimation 164
6.2.2 Selection of Smoothing Parameters 167
6.2.3 Algorithm and Implementation 168
6.2.4 Bayes Model, Direct GML and Approximate Bayesian Confidence Intervals 170
6.3 Wisconsin Epidemiological Study of Diabetic Retinopathy 172
6.4 Smoothing Spline Estimation of Variance Functions . . 176
6.5 Smoothing Spline Spectral Analysis 182
6.5.1 Spectrum Estimation of a Stationary Process . . 182
6.5.2 Time-Varying Spectrum Estimation of a Locally Stationary Process 183
6.5.3 Epileptic EEG 185

7 Smoothing Spline Nonlinear Regression 195
7.1 Motivation 195
7.2 Nonparametric Nonlinear Regression Models 196
7.3 Estimation with a Single Function 197
7.3.1 Gauss–Newton and Newton–Raphson Methods . 197
7.3.2 Extended Gauss–Newton Method 199
7.3.3 Smoothing Parameter Selection and Inference . . 201
7.4 Estimation with Multiple Functions 204
7.5 The nnr Function 205
7.6 Examples 206
7.6.1 Nonparametric Regression Subject to Positive Constraint 206
7.6.2 Nonparametric Regression Subject to Monotone Constraint 207
7.6.3 Term Structure of Interest Rates 212
7.6.4 A Multiplicative Model for Chickenpox Epidemic 218
7.6.5 A Multiplicative Model for Texas Weather 223

8 Semiparametric Regression **227**
8.1 Motivation 227
8.2 Semiparametric Linear Regression Models 228
8.2.1 The Model 228
8.2.2 Estimation and Inference 229
8.2.3 Vector Spline 233
8.3 Semiparametric Nonlinear Regression Models 240
8.3.1 The Model 240
8.3.2 SNR Models for Clustered Data 241
8.3.3 Estimation and Inference 242
8.3.4 The `snr` Function 245
8.4 Examples 247
8.4.1 Canadian Weather — Revisit 247
8.4.2 Superconductivity Magnetization Modeling . . . 254
8.4.3 Oil-Bearing Rocks 257
8.4.4 Air Quality 259
8.4.5 The Evolution of the Mira Variable R Hydrae . . 262
8.4.6 Circadian Rhythm 267

9 Semiparametric Mixed-Effects Models **273**
9.1 Linear Mixed-Effects Models 273
9.2 Semiparametric Linear Mixed-Effects Models 274
9.2.1 The Model 274
9.2.2 Estimation and Inference 275
9.2.3 The `slm` Function 279
9.2.4 SS ANOVA Decomposition 280
9.3 Semiparametric Nonlinear Mixed-Effects Models 283
9.3.1 The Model 283
9.3.2 Estimation and Inference 284
9.3.3 Implementation and the `snm` Function 286
9.4 Examples 288
9.4.1 Ozone in Arosa — Revisit 288
9.4.2 Lake Acidity — Revisit 291
9.4.3 Coronary Sinus Potassium in Dogs 294
9.4.4 Carbon Dioxide Uptake 305
9.4.5 Circadian Rhythm — Revisit 310

A Data Sets **323**
A.1 Air Quality Data 324
A.2 Arosa Ozone Data 324
A.3 Beveridge Wheat Price Index Data 324
A.4 Bond Data 324
A.5 Canadian Weather Data 325

A.6 Carbon Dioxide Data 325
A.7 Chickenpox Data 325
A.8 Child Growth Data 326
A.9 Dog Data 326
A.10 Geyser Data 326
A.11 Hormone Data 327
A.12 Lake Acidity Data 327
A.13 Melanoma Data 327
A.14 Motorcycle Data 328
A.15 *Paramecium caudatum* Data 328
A.16 Rock Data 328
A.17 Seizure Data 328
A.18 Star Data 329
A.19 Stratford Weather Data 329
A.20 Superconductivity Data 329
A.21 Texas Weather Data 330
A.22 Ultrasound Data 330
A.23 USA Climate Data 331
A.24 Weight Loss Data 331
A.25 WESDR Data 331
A.26 World Climate Data 332

B Codes for Fitting Strictly Increasing Functions **333**
B.1 C and R Codes for Computing Integrals 333
B.2 R Function `inc` 336

C Codes for Term Structure of Interest Rates **339**
C.1 C and R Codes for Computing Integrals 339
C.2 R Function for One Bond 341
C.3 R Function for Two Bonds 342

References **347**

Author Index **355**

Subject Index **359**