MEASUREMENT ERROR IN NONLINEAR MODELS: A MODERN PERSPECTIVE
Second Edition

R. J. Carroll*
D. Ruppert
L. A. Stefanski
C. M. Crainiceanu

* Copyright, 2006.
To our families and friends
| 2.1 | Functional and Structural Models | 25 |
| 2.2 | Models for Measurement Error | 26 |
| 2.2.1 | General Approaches: Berkson and Classical Models | 26 |
| 2.2.2 | Is It Berkson or Classical? | 27 |
| 2.2.3 | Berkson Models from Classical | 28 |
| 2.2.4 | Transportability of Models | 29 |
| 2.2.5 | Potential Dangers of Transporting Models | 30 |
| 2.2.6 | Semicontinuous Variables | 32 |
| 2.2.7 | Misclassification of a Discrete Covariate | 32 |
| 2.3 | Sources of Data | 32 |
| 2.4 | Is There an “Exact” Predictor? What is Truth? | 33 |
| 2.5 | Differential and Nondifferential Error | 36 |
| 2.6 | Prediction | 38 |
| Bibliographic Notes | 39 |

## 3 LINEAR REGRESSION AND ATTENUATION 41

| 3.1 | Introduction | 41 |
| 3.2 | Bias Caused by Measurement Error | 41 |
| 3.2.1 | Simple Linear Regression with Additive Error | 42 |
| 3.2.2 | Regression Calibration: Classical Error as Berkson Error | 44 |
| 3.2.3 | Simple Linear Regression with Berkson Error | 45 |
| 3.2.4 | Simple Linear Regression, More Complex Error Structure | 46 |
| 3.2.5 | Summary of Simple Linear Regression | 49 |
| 3.3 | Multiple and Orthogonal Regression | 52 |
| 3.3.1 | Multiple Regression: Single Covariate Measured with Error | 52 |
| 3.3.2 | Multiple Covariates Measured with Error | 53 |
| 3.4 | Correcting for Bias | 55 |
| 3.4.1 | Method of Moments | 55 |
| 3.4.2 | Orthogonal Regression | 57 |
| 3.5 | Bias Versus Variance | 60 |
| 3.5.1 | Theoretical Bias-Variance Tradeoff Calculations | 61 |
| 3.6 | Attenuation in General Problems | 63 |
| Bibliographic Notes | 64 |

## 4 REGRESSION CALIBRATION 65

<p>| 4.1 | Overview | 65 |
| 4.2 | The Regression Calibration Algorithm | 66 |
| 4.3 | NHANES Example | 66 |
| 4.4 | Estimating the Calibration Function Parameters | 70 |
| 4.4.1 | Overview and First Methods | 70 |</p>
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.6.1</td>
<td>Mixture of Berkson and Classical Error</td>
<td>123</td>
</tr>
<tr>
<td>5.6.2</td>
<td>Misclassification SIMEX</td>
<td>125</td>
</tr>
<tr>
<td>5.6.3</td>
<td>Checking Structural Model Robustness via Re-measurement</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>Bibliographic Notes</td>
<td>128</td>
</tr>
<tr>
<td>6</td>
<td>INSTRUMENTAL VARIABLES</td>
<td>129</td>
</tr>
<tr>
<td>6.1</td>
<td>Overview</td>
<td>129</td>
</tr>
<tr>
<td>6.1.1</td>
<td>A Note on Notation</td>
<td>130</td>
</tr>
<tr>
<td>6.2</td>
<td>Instrumental Variables in Linear Models</td>
<td>131</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Instrumental Variables via Differentiation</td>
<td>131</td>
</tr>
<tr>
<td>6.2.2</td>
<td>Simple Linear Regression with One Instrument</td>
<td>132</td>
</tr>
<tr>
<td>6.2.3</td>
<td>Linear Regression with Multiple Instruments</td>
<td>134</td>
</tr>
<tr>
<td>6.3</td>
<td>Approximate Instrumental Variable Estimation</td>
<td>137</td>
</tr>
<tr>
<td>6.3.1</td>
<td>IV Assumptions</td>
<td>137</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Mean and Variance Function Models</td>
<td>138</td>
</tr>
<tr>
<td>6.3.3</td>
<td>First Regression Calibration IV Algorithm</td>
<td>139</td>
</tr>
<tr>
<td>6.3.4</td>
<td>Second Regression Calibration IV Algorithm</td>
<td>140</td>
</tr>
<tr>
<td>6.4</td>
<td>Adjusted Score Method</td>
<td>140</td>
</tr>
<tr>
<td>6.5</td>
<td>Examples</td>
<td>143</td>
</tr>
<tr>
<td>6.5.1</td>
<td>Framingham Data</td>
<td>143</td>
</tr>
<tr>
<td>6.5.2</td>
<td>Simulated Data</td>
<td>145</td>
</tr>
<tr>
<td>6.6</td>
<td>Other Methodologies</td>
<td>145</td>
</tr>
<tr>
<td>6.6.1</td>
<td>Hybrid Classical and Regression Calibration</td>
<td>145</td>
</tr>
<tr>
<td>6.6.2</td>
<td>Error Model Approaches</td>
<td>147</td>
</tr>
<tr>
<td></td>
<td>Bibliographic Notes</td>
<td>148</td>
</tr>
<tr>
<td>7</td>
<td>SCORE FUNCTION METHODS</td>
<td>151</td>
</tr>
<tr>
<td>7.1</td>
<td>Overview</td>
<td>151</td>
</tr>
<tr>
<td>7.2</td>
<td>Linear and Logistic Regression</td>
<td>152</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Linear Regression Corrected and Conditional Scores</td>
<td>152</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Logistic Regression Corrected and Conditional Scores</td>
<td>157</td>
</tr>
<tr>
<td>7.2.3</td>
<td>Framingham Data Example</td>
<td>159</td>
</tr>
<tr>
<td>7.3</td>
<td>Conditional Score Functions</td>
<td>162</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Conditional Score Basic Theory</td>
<td>162</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Conditional Scores for Basic Models</td>
<td>164</td>
</tr>
<tr>
<td>7.3.3</td>
<td>Conditional Scores for More Complicated Models</td>
<td>166</td>
</tr>
<tr>
<td>7.4</td>
<td>Corrected Score Functions</td>
<td>169</td>
</tr>
<tr>
<td>7.4.1</td>
<td>Corrected Score Basic Theory</td>
<td>170</td>
</tr>
<tr>
<td>7.4.2</td>
<td>Monte Carlo Corrected Scores</td>
<td>170</td>
</tr>
<tr>
<td>7.4.3</td>
<td>Some Exact Corrected Scores</td>
<td>172</td>
</tr>
</tbody>
</table>
CONTENTS

7.4.4 SIMEX Connection 173
7.4.5 Corrected Scores with Replicate Measurements 173
7.5 Computation and Asymptotic Approximations 174
  7.5.1 Known Measurement Error Variance 175
  7.5.2 Estimated Measurement Error Variance 176
7.6 Comparison of Conditional and Corrected Scores 177
7.7 Bibliographic Notes 178

Bibliographic Notes 178

8 LIKELIHOOD AND QUASILIKELIHOOD 181
  8.1 Introduction 181
    8.1.1 Step 1: The Likelihood if \( X \) Were Observable 183
    8.1.2 A General Concern: Identifiable Models 184
  8.2 Steps 2 and 3: Constructing Likelihoods 184
    8.2.1 The Discrete Case 185
    8.2.2 Likelihood Construction for General Error Models 186
    8.2.3 The Berkson Model 188
    8.2.4 Error Model Choice 189
  8.3 Step 4: Numerical Computation of Likelihoods 190
  8.4 Cervical Cancer and Herpes 190
  8.5 Framingham Data 192
  8.6 Nevada Test Site Reanalysis 193
    8.6.1 Regression Calibration Implementation 195
    8.6.2 Maximum Likelihood Implementation 196
  8.7 Bronchitis Example 197
    8.7.1 Calculating the Likelihood 198
    8.7.2 Effects of Measurement Error on Threshold Models 199
    8.7.3 Simulation Study and Maximum Likelihood 199
    8.7.4 Berkson Analysis of the Data 201
  8.8 Quasilikelihood and Variance Function Models 201
    8.8.1 Details of Step 3 for QVF Models 202
    8.8.2 Details of Step 4 for QVF Models 203
Bibliographic Notes 203

9 BAYESIAN METHODS 205
  9.1 Overview 205
    9.1.1 Problem Formulation 205
    9.1.2 Posterior Inference 207
    9.1.3 Bayesian Functional and Structural Models 208
    9.1.4 Modularity of Bayesian MCMC 209
  9.2 The Gibbs Sampler 209
  9.3 Metropolis-Hastings Algorithm 211
  9.4 Linear Regression 213
### 9 Nonlinear Models

- **9.4.1 Example** 216
- **9.5 Nonlinear Models** 219
  - **9.5.1 A General Model** 219
  - **9.5.2 Polynomial Regression** 220
  - **9.5.3 Multiplicative Error** 221
  - **9.5.4 Segmented Regression** 222
- **9.6 Logistic Regression** 223
- **9.7 Berkson Errors** 225
  - **9.7.1 Nonlinear Regression with Berkson Errors** 225
  - **9.7.2 Logistic Regression with Berkson Errors** 227
  - **9.7.3 Bronchitis Data** 227
- **9.8 Automatic implementation** 230
  - **9.8.1 Implementation and simulations in WinBUGS** 231
  - **9.8.2 More complex models** 234
- **9.9 Cervical Cancer and Herpes** 235
- **9.10 Framingham Data** 237
- **9.11 OPEN Data: A Variance Components Model** 238

### Bibliographic Notes

240

### 10 HYPOTHESIS TESTING

- **10.1 Overview** 243
  - **10.1.1 Simple Linear Regression, Normally Distributed** 243
  - **X**
  - **10.1.2 Analysis of Covariance** 246
  - **10.1.3 General Considerations: What is a Valid Test?** 248
  - **10.1.4 Summary of Major Results** 248
- **10.2 The Regression Calibration Approximation** 249
  - **10.2.1 Testing \( H_0 : \beta_x = 0 \)** 250
  - **10.2.2 Testing \( H_0 : \beta_z = 0 \)** 250
  - **10.2.3 Testing \( H_0 : (\beta^t_x, \beta^t_z)^t = 0 \)** 250
- **10.3 Illustration: OPEN Data** 251
- **10.4 Hypotheses about Sub-vectors of \( \beta_x \) and \( \beta_z \)** 251
  - **10.4.1 Illustration: Framingham Data** 252
- **10.5 Efficient Score Tests of \( H_0 : \beta_x = 0 \)** 253
  - **10.5.1 Generalized Score Tests** 254

### Bibliographic Notes

257

### 11 LONGITUDINAL DATA AND MIXED MODELS

- **11.1 Mixed Models for Longitudinal Data** 259
  - **11.1.1 Simple Linear Mixed Models** 259
  - **11.1.2 The General Linear Mixed Model** 260
  - **11.1.3 The Linear Logistic Mixed Model** 261
  - **11.1.4 The Generalized Linear Mixed Model** 261

### Bibliographic Notes

261
CONTENTS

11.2 Mixed Measurement Error Models 262
  11.2.1 The Variance Components Model Revisited 262
  11.2.2 General Considerations 263
  11.2.3 Some Simple Examples 263
  11.2.4 Models for Within-Subject X-correlation 265
11.3 A Bias Corrected Estimator 265
11.4 SIMEX for GLMMEMs 267
11.5 Regression Calibration for GLMMs 267
11.6 Maximum Likelihood Estimation 267
11.7 Joint Modeling 268
11.8 Other Models and Applications 269
  11.8.1 Models With Random Effects Multiplied by X 269
  11.8.2 Models With Random Effects Depending Nonlinearly on X 270
  11.8.3 Inducing a True-Data Model From a Standard Observed Data Model 270
  11.8.4 Autoregressive Models in Longitudinal Data 271
11.9 Example: The CHOICE Study 272
Bibliographic Notes 275

12 NONPARAMETRIC ESTIMATION 277
12.1 Deconvolution 277
  12.1.1 The Problem 277
  12.1.2 Fourier Inversion 278
  12.1.3 Methodology 278
  12.1.4 Properties of Deconvolution Methods 279
  12.1.5 Is it Possible to Estimate the Bandwidth? 280
  12.1.6 Parametric Deconvolution 282
  12.1.7 Estimating Distribution Functions 285
  12.1.8 Optimal Score Tests 286
  12.1.9 Framingham Data 287
  12.1.10 NHANES Data 288
  12.1.11 Bayesian Density Estimation by Normal Mixtures 289
12.2 Nonparametric Regression 291
  12.2.1 Local Polynomial, Kernel-Weighted Regression 291
  12.2.2 Splines 292
  12.2.3 QVF and Likelihood Models 293
  12.2.4 SIMEX For Nonparametric Regression 294
  12.2.5 Regression Calibration 295
  12.2.6 Structural Splines 295
  12.2.7 Taylex and Other Methods 296
12.3 Baseline Change Example 297
  12.3.1 Discussion of the Baseline Change Controls Data 299
13 SEMIPARAMETRIC REGRESSION 301
13.1 Overview 301
13.2 Additive Models 301
13.3 MCMC for Additive Spline Models 302
13.4 Monte-Carlo EM-Algorithm 303
  13.4.1 Starting Values 304
  13.4.2 Metropolis Hastings Fact 304
  13.4.3 The Algorithm 304
13.5 Simulation with Classical Errors 307
13.6 Simulation with Berkson Errors 309
13.7 Semiparametrics: $X$ Modeled Parametrically 310
13.8 Parametric Models: No Assumptions on $X$ 312
  13.8.1 Deconvolution Methods 312
  13.8.2 Models Linear in Functions of $X$ 313
  13.8.3 Linear Logistic Regression With Replicates 314
  13.8.4 Doubly-Robust Parametric Modeling 315
13.9 Bibliographic Notes 316
Bibliographic Notes 316

14 SURVIVAL DATA 317
14.1 Notation and Assumptions 317
14.2 Induced Hazard Function 318
14.3 Regression Calibration for Survival Analysis 319
  14.3.1 Methodology and Asymptotic Properties 319
  14.3.2 Risk Set Calibration 320
14.4 SIMEX for Survival Analysis 321
14.5 Chronic Kidney Disease Progression 322
  14.5.1 Regression Calibration for CKD Progression 323
  14.5.2 SIMEX for CKD Progression 324
14.6 Semi and Nonparametric Methods 327
  14.6.1 Nonparametric Estimation With Validation Data 328
  14.6.2 Nonparametric Estimation With Replicated Data 330
  14.6.3 Likelihood Estimation 331
14.7 Likelihood Inference for Frailty Models 334
Bibliographic Notes 335

15 RESPONSE VARIABLE ERROR 337
15.1 Response Error and Linear Regression 337
15.2 Other Forms of Additive Response Error 341
  15.2.1 Biased Responses 341
  15.2.2 Response Error in Heteroscedastic Regression 342
CONTENTS  xi
Subject Index  443