

Measurement Error in Nonlinear Models

R.J. Carroll

*Professor of Statistics
Texas A & M University, USA*

D. Ruppert

*Professor of Operations Research and
Industrial Engineering
Cornell University, USA*

and

L.A. Stefanski

*Professor, Department of Statistics
North Carolina State University, USA*

CHAPMAN & HALL/CRC

Boca Raton London New York Washington, D.C.

Contents

Preface	xvii
Guide to Notation	xxiii
1 INTRODUCTION	1
1.1 Measurement Error Examples	1
1.1.1 Nutrition Studies	1
1.1.2 Nurses' Health Study	3
1.1.3 Bioassay in a Herbicide Study	3
1.1.4 Lung Function in Children	4
1.1.5 Coronary Heart Disease and Blood Pressure	4
1.1.6 A-Bomb Survivor Data	5
1.1.7 Blood Pressure and Urinary Sodium Chloride	5
1.2 Functional and Structural Models	6
1.3 Models for Measurement Error	7
1.3.1 General Approaches	7
1.3.2 Transportability of Models	10
1.3.3 Potential Dangers of Transporting Models	11
1.4 Sources of Data	12
1.5 Is There an "Exact" Predictor?	13
1.6 Differential and Nondifferential Error	16
1.7 True and Approximate Replicates	17
1.8 Measurement Error as a Missing Data Problem	18
1.9 Prediction	18
1.10 A Brief Tour	19
2 REGRESSION AND ATTENUATION	21
2.1 Introduction	21

2.2	Bias Caused by Measurement Error	21
2.2.1	Simple Linear Regression with Additive Error	22
2.2.2	Simple Linear Regression, More Complex Error Structure	23
2.2.3	Multiple Regression: Single Covariate Measured with Error	25
2.2.4	Multiple Covariates Measured with Error	26
2.3	Correcting for Bias	27
2.3.1	Method of Moments	27
2.3.2	Orthogonal Regression	28
2.4	Bias Versus Variance	32
2.5	Attenuation in General Problems	34
2.5.1	An Illustration of Nondifferential Measurement Error	36
2.6	Other References	37
2.7	Appendix	38
3	REGRESSION CALIBRATION	40
3.1	Overview	40
3.2	The Regression Calibration Algorithm	41
3.2.1	Correction for Attenuation	42
3.3	NHANES Example	42
3.4	Estimating the Calibration Function Parameters	46
3.4.1	Overview and First Methods	46
3.4.2	Best Linear Approximations Using Replicate Data	47
3.4.3	Nonlinear Calibration Function Models	48
3.4.4	Alternatives When Using Partial Replicates	50
3.4.5	James-Stein Calibration	50
3.5	Standard Errors	50
3.6	Expanded Regression Calibration Models	51
3.6.1	The Expanded Approximation Defined	52
3.6.2	Implementation	54
3.6.3	Models Without Severe Curvature	55
3.7	Bioassay Data	55
3.8	Heuristics and Accuracy of the Approximations	61
3.9	Examples of the Approximations	62
3.9.1	Linear Regression	63
3.9.2	Logistic Regression	63
3.9.3	Loglinear Mean Models	66

3.10	Theoretical Examples	67
3.10.1	Homoscedastic Regression	67
3.10.2	Quadratic Regression with Homoscedastic Regression Calibration	68
3.10.3	Loglinear Mean Model	68
3.10.4	Small Curvature, Heteroscedastic Calibration	69
3.11	Other References	69
3.12	Appendix	69
3.12.1	Error Variance Estimation in the CSFII	69
3.12.2	Standard Errors and Replication	72
3.12.3	Quadratic Regression: Details of The Ex- panded Calibration Model	78
4	SIMULATION EXTRAPOLATION	79
4.1	Overview	79
4.2	Simulation Extrapolation Heuristics	80
4.3	The SIMEX Algorithm	82
4.3.1	The Simulation and Extrapolation Steps	82
4.3.2	Modifications of the Simulation Step	83
4.3.3	Estimating the Measurement Error Variance	83
4.3.4	Extrapolant Function Considerations	84
4.3.5	Inference and Standard Errors	86
4.3.6	Relation to the Jackknife	86
4.4	Nonadditive Measurement Error	87
4.5	Framingham Heart Study	88
4.5.1	Full Replication	89
4.5.2	Partial Replication	92
4.6	SIMEX in Some Important Special Cases	94
4.6.1	Multiple Linear Regression	95
4.6.2	Loglinear Mean Models	95
4.6.3	Quadratic Mean Models	96
4.6.4	Segmented Linear Regression Mean Models	97
4.7	Theory and Variance Estimation	97
4.7.1	Simulation Extrapolation Variance Estimation	99
4.7.2	Estimating Equation Approach to Variance Estimation	101
5	INSTRUMENTAL VARIABLES	107
5.1	Overview	107
5.2	Approximate Instrumental Variable Estimation	108

5.2.1	First Regression Calibration Instrumental Variable Algorithm	109
5.2.2	Second Regression Calibration Instrumental Variable Algorithm	110
5.3	An Example	111
5.4	Derivation of the Estimators	112
5.4.1	First Regression Calibration Instrumental Variable Algorithm	113
5.4.2	Second Regression Calibration Instrumental Variable Algorithm	114
5.5	Asymptotic Distribution Approximations	116
5.5.1	Two-Stage Estimation	117
5.5.2	Computing Estimates and Standard Errors	120
6	FUNCTIONAL METHODS	122
6.1	Overview	122
6.2	Linear, Logistic and Gamma-Loglinear Models	123
6.3	Framingham Data	125
6.4	Unbiased Score Functions via Conditioning	125
6.4.1	Linear and Logistic Regression	128
6.4.2	Other Canonical Models	129
6.4.3	Computation	129
6.4.4	Inference	130
6.5	Exact Corrected Estimating Equations	131
6.5.1	Likelihoods With Exponentials and Powers	132
6.5.2	Asymptotic Distribution Approximation	133
6.6	Estimated Σ_{uu}	133
6.7	Infinite Series Corrected Estimating Equations	134
6.7.1	Rare-Event Logistic Regression	135
6.7.2	Extensions to Mean and Variance Function Models	136
6.8	Comparison of Methods	137
6.9	Appendix	139
6.9.1	Technical Complements to Conditional Score Theory	139
6.9.2	Technical Complements to Distribution Theory for Estimated Σ_{uu}	139
7	LIKELIHOOD AND QUASILIKELIHOOD	141
7.1	Introduction	141

7.1.1	Identifiable Models	143
7.2	Measurement Error Models and Missing Data	144
7.3	Likelihood Methods when \mathbf{X} is Unobserved	146
7.3.1	Error Models	147
7.3.2	Likelihood and External Second Measures	149
7.3.3	The Berkson Model	150
7.3.4	Error Model Choice	151
7.4	Likelihood When \mathbf{X} is Partly Observed	152
7.5	Numerical Computation of Likelihoods	153
7.6	Framingham Data	154
7.7	Bronchitis Example	156
7.8	Quasilikelihood and Variance Function Models	160
7.9	Appendix	161
7.9.1	Monte-Carlo Computation of Integrals	161
7.9.2	Linear, Probit and Logistic Regression	162
8	BAYESIAN METHODS	165
8.1	Overview	165
8.2	The Gibbs Sampler	168
8.2.1	Direct Sampling without Measurement Error	168
8.2.2	The Weighted Bootstrap	169
8.2.3	Forming Complete Data	170
8.3	Importance Sampling	171
8.4	Cervical Cancer	173
8.5	Framingham Data	175
8.5.1	Details of the Gibbs Sampler and Weighted Bootstrap	178
9	SEMIPARAMETRIC METHODS	182
9.1	Using Only Complete Data	183
9.2	Special Two-Stage Designs for Binary Responses	184
9.3	Pseudolikelihood	185
9.4	Mean Score Method	187
9.5	General Unbiased Estimating Functions	188
9.5.1	Using Polynomials	190
9.5.2	Optimal Moment-Based Estimators	191
9.5.3	Mean Based Moment-Based Estimators	191
9.6	Semiparametric Regression Calibration	192
9.7	Comparison of the Methods	193
9.8	Appendix	194

9.8.1	Use of Complete Data Only	194
9.8.2	Theory for Complete Data Only	196
9.8.3	Theory of Moment-Estimating Functions	197
10	UNKNOWN LINK FUNCTIONS	199
10.1	Overview	199
10.1.1	Constants of Proportionality	200
10.2	Estimation Methods	201
10.2.1	Some Basic Facts	201
10.2.2	Least Squares and Sliced Inverse Regression	201
10.2.3	Details of Implementation	202
10.3	Framingham Heart Study	203
10.4	Appendix	203
10.4.1	Basic Theory	203
11	HYPOTHESIS TESTING	206
11.1	Overview	206
11.2	The Regression Calibration Approximation	207
11.2.1	Testing $H_0 : \beta_x = 0$	208
11.2.2	Testing $H_0 : \beta_z = 0$	208
11.2.3	Testing $H_0 : (\beta_x^t, \beta_z^t)^t = 0$	208
11.3	Hypotheses about Subvectors of β_x and β_z	209
11.4	Efficient Score Tests of $H_0 : \beta_x = 0$	210
11.4.1	Generalized Score Tests	211
12	DENSITY ESTIMATION AND NONPARAMET- RIC REGRESSION	215
12.1	Deconvolution	215
12.1.1	Parametric Deconvolution via Moments	219
12.1.2	Estimating Distribution Functions	219
12.1.3	Optimal Score Tests	220
12.1.4	Framingham Data	220
12.1.5	NHANES Data	222
12.2	Nonparametric Regression	223
12.2.1	SIMEX	224
12.2.2	Regression Calibration	225
12.2.3	QVF and Likelihood Models	226
12.2.4	Framingham Data	226
12.2.5	Other Methods	228

13 RESPONSE VARIABLE ERROR	229
13.1 Additive/Multiplicative Error and QVF Models	230
13.1.1 Unbiased Measures of True Response	230
13.1.2 Recommendations	233
13.1.3 Biased Responses	233
13.1.4 Calibration	233
13.2 Likelihood Methods	235
13.2.1 General Likelihood Theory and Surrogates	235
13.2.2 Use of Complete Data Only	237
13.2.3 Other Methods	238
13.3 Semiparametric Methods	238
13.3.1 Pseudolikelihood—Simple Random Subsampling	238
13.3.2 Modified Pseudolikelihood—Other Types of Subsampling	239
13.4 Example	240
14 OTHER TOPICS	243
14.1 Logistic Case-Control Studies	243
14.1.1 The Case that \mathbf{X} is Observed	243
14.1.2 Measurement Error	244
14.1.3 Normal Discriminant Model	245
14.2 Differential Measurement Error	245
14.2.1 Likelihood Formulation	245
14.2.2 Functional Methods in Two-Stage Studies	246
14.2.3 Comparison of Functional and Likelihood Approaches	247
14.3 Mixture Methods as Functional Modeling	247
14.3.1 Overview	247
14.3.2 Nonparametric Mixture Likelihoods	248
14.3.3 A Cholesterol Example	250
14.3.4 Covariates Measured Without Error	251
14.4 Design of Two-Stage Validation and Replication Studies	251
14.5 Misclassification	253
14.6 Survival Analysis	254
14.6.1 General Considerations	254
14.6.2 Rare Events	254
14.6.3 Risk Set Calibration	255

A	FITTING METHODS AND MODELS	257
A.1	Overview	257
A.2	Likelihood Methods	257
A.2.1	Notation	257
A.2.2	Maximum likelihood Estimation	258
A.2.3	Likelihood Ratio Tests	259
A.2.4	Profile Likelihood and Likelihood Ratio Confidence Intervals	259
A.2.5	Efficient Score Tests	260
A.3	Unbiased Estimating Equations	261
A.3.1	Introduction and Basic Large Sample Theory	261
A.3.2	Sandwich Formula Example: Linear Regres- sion Without Measurement Error	264
A.3.3	Sandwich Method and Likelihood-type In- ference	265
A.3.4	Unbiased, But Conditionally Biased, Esti- mating Equations	266
A.3.5	Biased Estimating Equations	267
A.3.6	Stacking Estimating Equations: Using Prior Estimates of Some Parameters	267
A.4	Quasilikelihood and Variance Function (QVF) Models	269
A.4.1	General Ideas	269
A.4.2	Estimation and Inference for QVF Models	270
A.5	Generalized Linear Models	273
A.6	Bootstrap Methods	273
A.6.1	Introduction	273
A.6.2	Nonlinear Regression Without Measurement Error	274
A.6.3	Bootstrapping Heteroscedastic Regression Models	276
A.6.4	Bootstrapping Logistic Regression Models	277
A.6.5	Bootstrapping Measurement Error Models	277
A.6.6	Bootstrap Confidence Intervals	278
	References	280
	Author index	298
	Subject index	301