

Springer Series in Statistics

**Ludwig Fahrmeir
Gerhard Tutz**

**Multivariate
Statistical
Modelling Based
on Generalized
Linear Models**

Second Edition



Springer

Contents

Preface to the Second Edition	v
Preface to the First Edition	vii
List of Examples	xvii
List of Figures	xxi
List of Tables	xxv
1. Introduction	1
1.1 Outline and Examples	2
1.2 Remarks on Notation	13
1.3 Notes and Further Reading	14
2. Modelling and Analysis of Cross-Sectional Data: A Review of Univariate Generalized Linear Models	15
2.1 Univariate Generalized Linear Models	16
2.1.1 Data	16
Coding of Covariates	16
Grouped and Ungrouped Data	17
2.1.2 Definition of Univariate Generalized Linear Models ..	18
2.1.3 Models for Continuous Responses	22
Normal Distribution	22
Gamma Distribution	23
Inverse Gaussian Distribution	24
2.1.4 Models for Binary and Binomial Responses	24
Linear Probability Model	25
Probit Model	26
Logit Model	26
Complementary Log-Log Model	26
Complementary Log-Model	26
Binary Models as Threshold Models of Latent Linear Models	29
Parameter Interpretation	29
Overdispersion	35
2.1.5 Models for Count Data	36
Log-linear Poisson Model	36

	Linear Poisson Model	36
2.2	Likelihood Inference	38
2.2.1	Maximum Likelihood Estimation	38
	Log-likelihood, Score Function and Information Matrix	39
	Numerical Computation of the MLE by Iterative Methods	41
	Uniqueness and Existence of MLEs*	43
	Asymptotic Properties	44
	Discussion of Regularity Assumptions*	46
	Additional Scale or Overdispersion Parameter	47
2.2.2	Hypothesis Testing and Goodness-of-Fit Statistics ...	47
	Goodness-of-Fit Statistics	50
2.3	Some Extensions	55
2.3.1	Quasi-likelihood Models	55
	Basic Models	55
	Variance Functions with Unknown Parameters	58
	Nonconstant Dispersion Parameter	59
2.3.2	Bayesian Models	60
2.3.3	Nonlinear and Nonexponential Family Regression Models*	65
2.4	Notes and Further Reading	67
3.	Models for Multicategorical Responses: Multivariate Extensions of Generalized Linear Models	69
3.1	Multicategorical Response Models	70
3.1.1	Multinomial Distribution	70
3.1.2	Data	71
3.1.3	The Multivariate Model	72
3.1.4	Multivariate Generalized Linear Models	75
3.2	Models for Nominal Responses	77
3.2.1	The Principle of Maximum Random Utility	77
3.2.2	Modelling of Explanatory Variables: Choice of Design Matrix	79
3.3	Models for Ordinal Responses	81
3.3.1	Cumulative Models: The Threshold Approach	83
	Cumulative Logistic Model or Proportional Odds Model	83
	Grouped Cox Model or Proportional Hazards Model ..	86
	Extreme Maximal-value Distribution Model	86
3.3.2	Extended Versions of Cumulative Models	87
3.3.3	Link Functions and Design Matrices for Cumulative Models	88
3.3.4	Sequential Models	92
	Generalized Sequential Models	95
	Link Functions of Sequential Models	98

3.3.5	Strict Stochastic Ordering*	99
3.3.6	Two-Step Models	100
	Link Function and Design Matrix for Two-Step Models	102
3.3.7	Alternative Approaches	103
3.4	Statistical Inference	105
3.4.1	Maximum Likelihood Estimation	105
	Numerical Computation	107
3.4.2	Testing and Goodness-of-Fit	107
	Testing of Linear Hypotheses	107
	Goodness-of-Fit Statistics	107
3.4.3	Power-Divergence Family*	109
	Asymptotic Properties under Classical "Fixed Cells" Assumptions	111
	Sparseness and "Increasing-Cells" Asymptotics	112
3.5	Multivariate Models for Correlated Responses	112
3.5.1	Conditional Models	114
	Asymmetric Models	114
	Symmetric Models	116
3.5.2	Marginal Models	119
	Marginal Models for Correlated Univariate Responses	120
	The Generalized Estimating Approach for Statistical Inference	123
	Marginal Models for Correlated Categorical Responses	129
	Likelihood-based Inference for Marginal Models	135
3.6	Notes and Further Reading	136
	Bayesian Inference	136
4.	Selecting and Checking Models	139
4.1	Variable Selection	139
4.1.1	Selection Criteria	140
4.1.2	Selection Procedures	142
	All-Subsets Selection	142
	Stepwise Backward and Forward Selection	143
4.2	Diagnostics	145
4.2.1	Diagnostic Tools for the Classical Linear Model	146
4.2.2	Generalized Hat Matrix	147
4.2.3	Residuals and Goodness-of-Fit Statistics	151
4.2.4	Case Deletion	156
4.3	General Tests for Misspecification*	161
4.3.1	Estimation under Model Misspecification	162
4.3.2	Hausman-type Tests	165
	Hausman Tests	165
	Information Matrix Test	166



4.3.3	Tests for Nonnested Hypotheses	167
	Tests Based on Artificial Nesting	168
	Generalized Wald and Score Tests	168
4.4	Notes and Further Reading	170
	Bayesian Model Determination	170
	Robust Estimates	172
	Model Tests Against Smooth Alternatives	172
5.	Semi- and Nonparametric Approaches to	
	Regression Analysis	173
5.1	Smoothing Techniques for Continuous Responses	174
5.1.1	Regression Splines and Other Basis Functions	174
	Regression Splines	176
	Other Basis Functions	178
	Regularization	179
5.1.2	Smoothing Splines	181
5.1.3	Local Estimators	183
	Simple Neighborhood Smoothers	183
	Local Regression	184
	Bias-Variance Trade-off	187
	Relation to Other Smoothers	189
5.1.4	Selection of Smoothing Parameters	190
5.2	Smoothing for Non-Gaussian Data	193
5.2.1	Basis Function Approach	193
	Fisher Scoring for Penalized Likelihood*	194
5.2.2	Penalization and Spline Smoothing	195
	Fisher Scoring for Generalized Spline Smoothing* ...	196
	Choice of Smoothing Parameter	197
5.2.3	Localizing Generalized Linear Models	198
	Local Fitting by Weighted Scoring	201
5.3	Modelling with Multiple Covariates	202
5.3.1	Modelling Approaches	207
	Generalized Additive Models	207
	Partially Linear Models	208
	Varying-Coefficient Models	208
	Projection Pursuit Regression	209
	Basis Function Approach	210
5.3.2	Estimation Concepts	213
	Backfitting Algorithm for Generalized	
	Additive Models	213
	Backfitting with Spline Functions	217
	Choice of Smoothing Parameter	220
	Partial Linear Models	220
5.4	Semiparametric Bayesian Inference for	
	Generalized Regression	221

5.4.1	Gaussian Responses	221
	Smoothness Priors Approaches	221
	Basis Function Approaches	227
	Models with Multiple Covariates	228
5.4.2	Non-Gaussian Responses	231
	Latent Variable Models for Categorical Responses ...	234
5.5	Notes and Further Reading	239
6.	Fixed Parameter Models for Time Series and Longitudinal Data	241
6.1	Time Series	242
6.1.1	Conditional Models	242
	Generalized Autoregressive Models	242
	Quasi-Likelihood Models and Generalized Autore- gression Moving Average Models	246
6.1.2	Statistical Inference for Conditional Models	249
6.1.3	Marginal Models	255
	Estimation of Marginal Models	258
6.2	Longitudinal Data	260
6.2.1	Conditional Models	261
	Generalized Autoregressive Models, Quasi-Likelihood Models	261
	Statistical Inference	262
	Transition Models	264
	Subject-specific Approaches and Conditional Likelihood	264
6.2.2	Marginal Models	267
	Statistical Inference	268
6.2.3	Generalized Additive Models for Longitudinal Data ..	274
6.3	Notes and Further Reading	278
7.	Random Effects Models	283
7.1	Linear Random Effects Models for Normal Data	285
7.1.1	Two-stage Random Effects Models	285
	Random Intercepts	286
	Random Slopes	287
	Multilevel Models	288
7.1.2	Statistical Inference	289
	Known Variance-Covariance Components	289
	Unknown Variance-Covariance Components	289
	Derivation of the EM algorithm*	291
7.2	Random Effects in Generalized Linear Models	292
	Generalized Linear Models with Random Effects	293
	Examples	294
7.3	Estimation Based on Posterior Modes	298

7.3.1	Known Variance-Covariance Components	298
7.3.2	Unknown Variance-Covariance Components	299
7.3.3	Algorithmic Details*	300
	Fisher Scoring for Given Variance-Covariance Components	300
	EM Type Algorithm	302
7.4	Estimation by Integration Techniques	303
7.4.1	Maximum Likelihood Estimation of Fixed Parameters Direct Maximization Using Fitting Techniques for GLMs	303 305
	Nonparametric Maximum Likelihood for Finite Mix- tures	308
7.4.2	Posterior Mean Estimation of Random Effects	310
7.4.3	Indirect Maximization Based on the EM Algorithm*	311
7.4.4	Algorithmic Details for Posterior Mean Estimation*	315
7.5	Examples	318
7.6	Bayesian Mixed Models	321
	Bayesian Generalized Mixed Models	321
	Generalized Additive Mixed Models	322
7.7	Marginal Estimation Approach to Random Effects Models	325
7.8	Notes and Further Reading	328
8.	State Space and Hidden Markov Models	331
8.1	Linear State Space Models and the Kalman Filter	332
8.1.1	Linear State Space Models	332
8.1.2	Statistical Inference	337
	Linear Kalman Filtering and Smoothing	338
	Kalman Filtering and Smoothing as Posterior Mode Estimation*	340
	Unknown Hyperparameters	342
	EM Algorithm for Estimating Hyperparameters*	343
8.2	Non-Normal and Nonlinear State Space Models	345
8.2.1	Dynamic Generalized Linear Models	345
	Categorical Time Series	347
8.2.2	Nonlinear and Nonexponential Family Models*	349
8.3	Non-Normal Filtering and Smoothing	350
8.3.1	Posterior Mode Estimation	351
	Generalized Extended Kalman Filter and Smoother*	352
	Gauss-Newton and Fisher-Scoring Filtering and Smoothing*	354
	Estimation of Hyperparameters*	356
	Some Applications	356
8.3.2	Markov Chain Monte Carlo and Integration-based Approaches	361
	MCMC Inference	362

	Integration-based Approaches	365
8.4	Longitudinal Data	369
8.4.1	State Space Modelling of Longitudinal Data	369
8.4.2	Inference For Dynamic Generalized Linear Mixed Models	372
8.5	Spatial and Spatio-temporal Data	376
8.6	Notes and Further Reading	383
9.	Survival Models	385
9.1	Models for Continuous Time	385
9.1.1	Basic Models	385
	Exponential Distribution	386
	Weibull Distribution	387
	Piecewise Exponential Model	388
9.1.2	Parametric Regression Models	388
	Location-Scale Models for $\log T$	388
	Proportional Hazards Models	389
	Linear Transformation Models and Binary Regression Models	390
9.1.3	Censoring	391
	Random Censoring	391
	Type I Censoring	392
9.1.4	Estimation	393
	Exponential Model	394
	Weibull Model	394
	Piecewise Exponential Model	395
9.2	Models for Discrete Time	396
9.2.1	Life Table Estimates	397
9.2.2	Parametric Regression Models	400
	The Grouped Proportional Hazards Model	400
	A Generalized Version: The Model of Aranda-Ordaz	402
	The Logistic Model	403
	Sequential Model and Parameterization of the Baseline Hazard	403
9.2.3	Maximum Likelihood Estimation	404
9.2.4	Time-varying Covariates	408
	Internal Covariates*	411
	Maximum Likelihood Estimation*	412
9.3	Discrete Models for Multiple Modes of Failure	414
9.3.1	Basic Models	414
9.3.2	Maximum Likelihood Estimation	417
9.4	Smoothing in Discrete Survival Analysis	420
9.4.1	Smoothing Life Table Estimates	420
9.4.2	Smoothing with Covariates	422
9.4.3	Dynamic Discrete-Time Survival Models	423



Posterior Mode Smoothing	423
Fully Bayesian Inference via MCMC	425
9.5 Remarks and Further Reading	429
A.	433
A.1 Exponential Families and Generalized Linear Models	433
A.2 Basic Ideas for Asymptotics	437
A.3 EM Algorithm	442
A.4 Numerical Integration	443
A.5 Monte Carlo Methods	449
B. Software for Fitting Generalized Linear Models and Extensions	455
Bibliography	467
Author Index	505
Subject Index	512