Contents

1 Introduction **1**
- 1.1 Notational conventions 2
- 1.2 A short review of generalized linear models 3
  - 1.2.1 Historical review 3
  - 1.2.2 Basics 6
  - 1.2.3 Link and variance functions 8
  - 1.2.4 Algorithms 9
- 1.3 Software 11
  - 1.3.1 S-PLUS 12
  - 1.3.2 SAS 13
  - 1.3.3 Stata 13
  - 1.3.4 SUDAAN 14
- 1.4 Exercises 15

2 Model Construction and Estimating Equations **17**
- 2.1 Independent data 17
  - 2.1.1 The FIML estimating equation for linear regression 18
  - 2.1.2 The FIML estimating equation for Poisson regression 21
  - 2.1.3 The FIML estimating equation for Bernoulli regression 22
  - 2.1.4 The LIML estimating equation for GLMs 24
  - 2.1.5 The LIMQL estimating equation for GLMs 27
- 2.2 Estimating the variance of the estimates 28
- 2.3 Panel data 32
  - 2.3.1 Pooled estimators 33
  - 2.3.2 Fixed-effects and random-effects models 34
    - 2.3.2.1 Unconditional fixed-effects models 35
    - 2.3.2.2 Conditional fixed-effects models 36
    - 2.3.2.3 Random-effects models 42
  - 2.3.3 Population-averaged and subject-specific models 49
- 2.4 Estimation 50
- 2.5 Summary 50
- 2.6 Exercises 52

3 Generalized Estimating Equations **55**
- 3.1 Population-averaged (PA) and subject-specific (SS) models 55
- 3.2 The PA-GEE for GLMs 57

3.2.1 Parameterizing the working correlation matrix 58
3.2.1.1 Exchangeable correlation 59
3.2.1.2 Autoregressive correlation 66
3.2.1.3 Stationary correlation 68
3.2.1.4 Nonstationary correlation 71
3.2.1.5 Unstructured correlation 72
3.2.1.6 Fixed correlation 73
3.2.1.7 Free specification 73
3.2.2 Estimating the scale variance (dispersion parameter) 76
3.2.2.1 Independence models 77
3.2.2.2 Exchangeable models 82
3.2.3 Estimating the PA-GEE model 85
3.2.4 Convergence of the estimation routine 89
3.2.5 ALR: Estimating correlations for binomial models 89
3.2.6 Summary 93
3.3 The SS-GEE for GLMs 95
3.3.1 Single random-effects 96
3.3.2 Multiple random-effects 98
3.3.3 Applications of the SS-GEE 99
3.3.4 Estimating the SS-GEE model 103
3.3.5 Summary 104
3.4 The GEE2 for GLMs 104
3.5 GEEs for extensions of GLMs 106
3.5.1 Generalized logistic regression 106
3.5.2 Cumulative logistic regression 108
3.6 Further developments and applications 110
3.6.1 The PA-GEE for GLMs with measurement error 110
3.6.2 The PA-EGEE for GLMs 117
3.6.3 The PA-REGEE for GLMs 119
3.7 Missing data 122
3.8 Choosing an appropriate model 128
3.9 Summary 131
3.10 Exercises 134

4 Residuals, Diagnostics, and Testing 137
4.1 Criterion measures 139
4.1.1 Choosing the best correlation structure 139
4.1.2 Choosing the best subset of covariates 142
4.2 Analysis of residuals 142
4.2.1 A nonparametric test of the randomness of residuals 143
4.2.2 Graphical assessment 143
4.2.3 Quasivariance functions for PA-GEE models 154
4.3 Deletion diagnostics 158
4.3.1 Influence measures 159
4.3.2 Leverage measures 165
4.4 Goodness of fit (population-averaged models) 165

4.4.1 Proportional reduction in variation 165
4.4.2 Concordance correlation 166
4.4.3 A χ^2 goodness of fit test for PA-GEE binomial models 167
4.5 Testing coefficients in the PA-GEE model 169
4.5.1 Likelihood ratio tests 170
4.5.2 Wald tests 172
4.5.3 Score tests 174
4.6 Assessing the MCAR assumption of PA-GEE models 174
4.7 Summary 177
4.8 Exercises 179

5 Programs and Datasets 181
5.1 Programs 181
5.1.1 Fitting PA-GEE models in Stata 182
5.1.2 Fitting PA-GEE models in SAS 183
5.1.3 Fitting PA-GEE models in S-PLUS 184
5.1.4 Fitting ALR models in SAS 185
5.1.5 Fitting PA-GEE models in SUDAAN 186
5.1.6 Calculating QIC in Stata 187
5.1.7 Calculating QICu in Stata 188
5.1.8 Graphing the residual runs test in S-PLUS 189
5.1.9 Using the fixed correlation structure in Stata 190
5.1.10 Fitting quasivariance PA-GEE models in S-PLUS 191
5.2 Datasets 192
5.2.1 Wheeze data 192
5.2.2 Ship accident data 194
5.2.3 Progabide data 196
5.2.4 Simulated logistic data 202
5.2.5 Simulated user-specified correlated data 209
5.2.6 Simulated measurement error data for the PA-GEE 212

References 215

Author index 219

Subject index 221