1. Introducing SAS 1
2. The Simple Linear Regression Model 50
3. Interval Estimation and Hypothesis Testing 82
4. Prediction, Goodness-of-Fit, and Modeling Issues 103
5. The Multiple Regression Model 130
6. Further Inference in the Multiple Regression Model 162
7. Using Indicator Variables 190
8. Heteroskedasticity 207
9. Regression with Time-Series Data: Stationary Variables 264
10. Random Regressors and Moment-Based Estimation 304
11. Simultaneous Equations Models 346
12. Regression with Time-Series Data: Nonstationary Variables 369
13. Vector Error Correction and Vector Autoregressive Models 390
14. Time-Varying Volatility and ARCH Models 406
15. Panel Data Models 428
16. Qualitative and Limited Dependent Variable Models 468

Appendix A. Math Functions 522
Appendix B. Probability 528
Appendix C. Review of Statistical Inference 541
CONTENTS

1. Introducing SAS 1
   1.1 The SAS System 1
   1.2 Starting SAS 1
   1.3 The opening display 1
   1.4 Exiting SAS 3
   1.5 Using Principles of Econometrics, 4E data files 3
       1.5.1 Data definition files 4
   1.6 A working environment 4
   1.7 SAS Program structure 6
       1.7.1 SAS comment statements 6
       1.7.2 Creating a SAS program 7
       1.7.3 Saving a SAS program 8
       1.7.4 Running a SAS program 9
       1.7.5 Printing data with PROC PRINT 10
       1.7.6 Saving SAS output 12
       1.7.7 Opening SAS programs 15
   1.8 Summary Statistics using PROC MEANS 15
   1.9 Making errors in SAS programs 17
       1.9.1 Typing errors 18
       1.9.2 The SAS semi-colon “;” 18
   1.10 SAS Graphics: A scatter diagram 19
       1.10.1 PROC PLOT 19
       1.10.2 PROC GPLOT 20
   1.11 Creating or modifying data sets 22
       1.11.1 The SET statement 22
       1.11.2 Using DROP and KEEP 22
   1.12 Creating new variables 23
       1.12.1 Arithmetic operators 23
       1.12.2 Comparison operators 24
       1.12.3 Logical operators 25
       1.12.4 Using SAS functions 25
       1.12.5 Missing values 26
       1.12.6 Using IF-THEN to recode variables 26
       1.12.7 Creating a data subset 27
       1.12.8 Using SET to combine data sets 27
   1.13 Using SET to open SAS data sets 28
       1.13.1 Using SAS system options 29
       1.13.2 Adding labels 30
   1.14 Using PROC SORT 31
       1.14.1 PROC PRINT with BY 31
       1.14.2 PROC MEANS with BY 32
       1.14.3 PROC SORT on two variables 32
       1.14.4 Sort in descending order 33
   1.15 Merging data sets 33
   Appendix 1A A guide to SAS help and online documentation 34
       1A.1 SAS command line 35
       1A.2 SAS help 35
       1A.3 SAS online documentation 37
       1A.4 SAS online examples 40
       1A.5 Other resources 40
   Appendix 1B Importing data into SAS 41
       1B.1 Reading ASCII data 41
       1B.2 Reading an external ASCII file 42
       1B.3 Importing data in Excel format 47

2. The Simple Linear Regression Model 50
   2.1 Econometric model and estimators 50
   2.2 Example: the food expenditure data 52
   2.3 Scatter diagram using PROC GPLOT 53
   2.4 Using PROC REG for simple regression 54
       2.4.1 Analysis of variance table 54
       2.4.2 ANOVA auxiliary information 56
       2.4.3 PROC MEANS options 56
   2.5 PROC REG options 57
       2.5.1 Covariance matrix 57
       2.5.2 The least squares residuals 58
       2.5.3 Output residuals 58
       2.5.4 PROC UNIVARIATE analysis of residuals 59
   2.6 Prediction with PROC REG 60
       2.6.1 Deleting missing values from data set 62
       2.6.2 Plotting a fitted line using PROC GPLOT 63
   2.7 Creating plots using PROC REG 63
   2.8 SAS ODS graphics 64
2.9 Fitting nonlinear relationships 66
2.10 Using indicator variables 70

Appendix 2A Calculation of least squares estimates: Details 71

Appendix 2B Monte Carlo simulation 75
2B.1 The true estimator variance 76
2B.2 Regression on artificial data 77
2B.3 OUTEST from PROC REG 78
2B.4 Simulating samples using do-loops 79
2B.5 Summarizing parameter estimates 79

3. Interval Estimation and Hypothesis Testing 82
3.1 Interval estimation 82
3.1.1 Interval estimation details 84
3.2 Hypothesis testing theory 85
3.2.1 Right tail t-tests 86
3.2.2 Left tail t-tests 86
3.2.3 Two-tail t-tests 86
3.2.4 The p-value for t-tests 87
3.3 Hypothesis testing examples 87
3.3.1 Right tail test of significance 87
3.3.2 Right tail test for an economic hypothesis 88
3.3.3 Left tail test of an economic hypothesis 89
3.3.4 Two tail test of an economic hypothesis 90
3.3.5 Two tail test of significance 91
3.4 Testing and estimating linear combinations 91
3.4.1 PROC MODEL 92

Appendix 3A Monte Carlo simulation 94
3A.1 Summarizing interval estimates 94
3A.2 Summarizing t-tests 96
3A.3 Illustrating the central limit theorem 97
3A.4 Monte Carlo experiment with triangular errors 99

4. Prediction, Goodness-of-Fit, and Modeling Issues 103
4.1 Least squares prediction theory 103
4.2 Least squares prediction example 104
4.3 Measuring goodness-of-fit 108
4.4 Residual analysis 109
4.4.1 Using PROC AUTOREG 112
4.5 SAS ODS graphics 113
4.5.1 The SAS Image Editor 114
4.5.2 ODS plots 115
4.6 Nonlinear relationships 118
4.7 Log-linear models 122
4.7.1 A growth model 122
4.7.2 A wage equation 124
4.7.3 Prediction in the log-linear model 127

5. The Multiple Regression Model 130
5.1 Multiple regression theory and methods 130
5.2 Multiple regression example 132
5.2.1 Using PROC REG 133
5.2.2 Using PROC AUTOREG 136
5.2.3 Using PROC MODEL 136
5.3 Polynomial models 137
5.3.1 Using PROC REG 138
5.3.2 Using PROC MODEL 138
5.4 Log-linear models 139
5.4.1 Using PROC REG 140
5.4.2 Using PROC MODEL 141

Appendix 5A The delta method in PROC MODEL 142
5A.1 Monte Carlo study of delta method 143

Appendix 5B Matrix operations 147
5B.1 Vector concepts 148
5B.2 Matrix concepts 149

Appendix 5C Regression calculations in matrix notation 154
5C.1 SAS/IML module for multiple regression 155
5C.2 Estimating a linear combination of parameters 158
5C.3 Testing a single linear hypothesis 158
5C.4 Illustrating computations 159
5C.5 Delta method 160

6. Further Inference in the Multiple Regression Model 162
6.1 Joint hypothesis tests 162
6.1.1 An example 163
6.1.2 PROC REG Test statement 165
6.1.3 F-test of model significance 167
6.1.4 Testing in PROC AUTOREG 167
6.1.5 PROC AUTOREG fit statistics 168
6.1.6 Testing in PROC MODEL 169
6.2 Restricted estimation 170
6.3 Model specification issues 172
6.3.1 The RESET test 174
6.4 Collinearity 175
6.4.1 Consequences of collinearity 176
6.4.2 Diagnosing collinearity 176
6.4.3 Condition indexes 178
6.5 Prediction in multiple regression 179
6A.1 Extending the matrix approach 180
6A.1.1 ANOVA for OLS module 180
6A.2 Prediction and prediction interval 183
6A.3 Tests of a joint hypothesis 184
6A.4 Collinearity diagnostics 187

7. Using Indicator Variables 190
7.1 Indicator variables 190
7.1.1 Slope and intercept effects 190
7.1.2 The Chow test 192
7.2 Using PROC MODEL for log-linear regression 195
7.3 The linear probability model 197
7.4 Treatment effects 198

7.5 Differences-in-differences estimation 202

8. Heteroskedasticity 207
8.1 The nature of heteroskedasticity 207
8.2 Plotting the least squares residuals 207
8.3 Least squares with robust standard errors 209
8.4 Generalized least squares estimation 211
8.4.1 Applying GLS using transformed data 212
8.4.2 Using PROC REG with a WEIGHT statement 213
8.5 Estimating the variance function 213
8.5.1 Model of multiplicative heteroskedasticity 213
8.5.2 A convenient special case 214
8.5.3 Two-step estimator for multiplicative heteroskedasticity 214
8.6 Lagrange multiplier (LM) test for heteroskedasticity 216
8.7 Goldfeld-Quandt test for heteroskedasticity 218
8.8 A heteroskedastic partition 221
8.8.1 The Goldfeld-Quandt test 222
8.8.2 Generalized least squares estimation 224
8.9 Using PROC AUTOREG for heteroskedasticity 225
8.9.1 PROC AUTOREG for a heteroskedastic partition 226
8.9.2 An extended heteroskedasticity model 227
8.10 Using SAS ODS graphics 228
8.11 Using PROC MODEL for heteroskedastic data 229

Appendix 8A Monte Carlo simulations 231
8A.1 Simulating heteroskedastic data 231
8A.2 Heteroskedastic data Monte Carlo experiment 233
8A.3 Using PROC IML to compute true variances 238
### Chapter 9 Regression with Time-Series Data:

**Stationary Variables**  
9.1 Time-series data 264  
9.2 Finite distributed lags 264  
9.2.1 Lag and difference operators 265  
9.2.2 Time-series plots 267  
9.2.3 Model estimation 268  
9.3 Serial correlation 269  
9.3.1 Residual correlogram 272  
9.4 Testing for serially correlated errors 274  
9.4.1 A Lagrange multiplier (LM) test 274  
9.4.2 Durbin-Watson test 276  
9.5 Estimation with serially correlated errors 277

**10. Random Regressors and Moment-Based Estimation**  
10.1 The consequences of random regressors 304  
10.2 Instrumental variables estimation 305  
10.2.1 Two-stage least squares estimation 306  
10.3 An illustration using simulated data 307  
10.3.1 Using two-stage least squares 309  
10.4 A wage equation 315  
10.4.1 Robust specification tests 319  
10.5 Using PROC MODEL 321  
10.5.1 Robust 2SLS estimation 321

### Appendix 8B Two-step estimation 245

8B.1 Simulating heteroskedastic data 245  
8B.2 Feasible GLS in multiplicative model 246  
8B.3 Feasible GLS in PROC IML 247

### Appendix 8C Multiplicative model Monte Carlo 249

8C.1 Simulating heteroskedastic data 249  
8C.2 The least squares estimator 250  
8C.3 Maximum likelihood estimation 251

### Appendix 8D Multiplicative model MLE 253

8D.1 Using PROC AUTOREG 253  
8D.2 Numerical optimization in the multiplicative model 254  
8D.3 MLE based tests for heteroskedasticity 257  
8D.4 MLE using analytic derivatives 258  
8D.5 MLE using method of scoring 260

### Appendix 9A Estimation and forecasting with PROC ARIMA 291

9A.1 Finite distributed lag models in PROC ARIMA 291  
9A.2 Serially correlated error models in PROC ARIMA 293  
9A.3 Autoregressive distributed lag models in PROC ARIMA 295  
9A.4 Autoregressive models and forecasting in PROC ARIMA 297

### Appendix 9B GLS estimation of AR(1) error model 299
10.5.2 Using the Hausman test command 321

Appendix 10A Simulating endogenous regressors 323
10A.1 Simulating the data 323
10A.2 The Cholesky decomposition 326

Appendix 10B Using PROC IML for 2SLS 328
10B.1 The model, estimators and tests 328
10B.2 PROC IML commands 330

Appendix 10C The repeated sampling properties of IV/2SLS 336
10D.1 The model, estimators and tests 342
10D.2 Using PROC MODEL and IML 342

Appendix 10D Robust 2SLS and GMM 342
10D.1 The model, estimators and tests 342
10D.2 Using PROC MODEL and IML 342

Appendix 10E Using PROC IML for 2SLS 328
10E.1 The model, estimators and tests 328
10E.2 PROC IML commands 330

Appendix 10F Using PROC IML for 2SLS 328
10F.1 The model, estimators and tests 328
10F.2 PROC IML commands 330

Appendix 11A Alternatives to two-stage least squares 356
11A.1 The LIML estimator 357
11A.2 Fuller’s modified LIML 357
11A.3 Advantages of LIML 358
11A.4 Stock-Yogo weak IV tests for LIML 358
11A.5 LIML and k-class algebra 358
11A.6 PROC IML for LIML and k-class 359

Appendix 11A Alternatives to two-stage least squares 356
11A.1 The LIML estimator 357
11A.2 Fuller’s modified LIML 357
11A.3 Advantages of LIML 358
11A.4 Stock-Yogo weak IV tests for LIML 358
11A.5 LIML and k-class algebra 358
11A.6 PROC IML for LIML and k-class 359

Appendix 11A Alternatives to two-stage least squares 356
11A.1 The LIML estimator 357
11A.2 Fuller’s modified LIML 357
11A.3 Advantages of LIML 358
11A.4 Stock-Yogo weak IV tests for LIML 358
11A.5 LIML and k-class algebra 358
11A.6 PROC IML for LIML and k-class 359

Appendix 11B Monte Carlo simulation 364

12. Regression with Time-Series Data: Nonstationary Variables 369
12.1 Stationary and nonstationary variables 369
12.1.1 The first-order autoregressive model 375
12.1.2 Random walk models 375
12.2 Spurious regressions 375
12.3 Unit root tests for stationarity 378
12.3.1 The Dickey-Fuller tests: an example 380
12.3.2 Order of integration 382
12.4 Cointegration 385
12.4.1 An example of a cointegration test 386
12.4.2 The error correction model 388

13. Vector Error Correction and Vector Autoregressive Models 390
13.1 VEC and VAR models 390
13.2 Estimating a vector error correction model 391
13.3 Estimating a VAR model 397
13.4 Impulse responses and variance decompositions 403

14. Time-Varying Volatility and ARCH Models 406
14.1 Time-varying volatility 406
14.2 Testing, estimating and forecasting 411
14.2.1 Testing for ARCH effects 412
14.2.2 Estimating an ARCH model 415
14.2.3 Forecasting volatility 417
14.3 Extensions 418
14.3.1 The GARCH model—generalized ARCH 418
14.3.2 Allowing for an asymmetric effect—threshold GARCH 421
14.3.3 GARCH-in-mean and time-varying risk premium 424

15. Panel Data Models 428
15.1 A microeconometric panel 428
15.2 A pooled model 429
15.2.1 Cluster-robust standard errors 430
15.3 The fixed effects model 432
15.3.1 The fixed effects estimator 435
15.3.2 The fixed effects estimator using PROC PANEL 438
15.3.3 Fixed effects using the complete panel 439
15.4 Random effects estimation 440
15.4.1 The Breusch-Pagan test 442
15.4.2 The Hausman test 442
15.5 Sets of regression equations 444
15.5.1 Seemingly unrelated regressions 447
15.5.2 Using PROC MODEL for SUR 449
15A.1 NLS examples 452
15A.2 Using PROC IML 453

Appendix 15B Panel data estimation details 454
15B.1 Estimating variance components 454
15B.2 Using PROC PANEL 456
15B.3 Using PROC IML 458
15C Robust fixed effects estimation 461
15D The Hausman-Taylor estimator 464

16. Qualitative and Limited Dependent Variable Models 468
16.1 Models with binary dependent variables 468
16.1.1 The linear probability model 469
16.1.2 The probit model 472
16.1.3 The logit model 476
16.1.4 A labor force participation model 476
16.2 Probit for consumer choice 479
16.2.1 Wald tests 482
16.2.2 Likelihood ratio tests 483
16.3 Multinomial logit 484
16.3.1 Example: post-secondary education choice 485
16.4 Conditional logit 491
16.4.1 Marginal effects 496
16.4.2 Testing the IIA assumption 497
16.5 Ordered choice models 501
16.6 Models for count data 504
16.7 Limited dependent variable models 507
16.7.1 Censored variable models 508
16.7.2 Sample selection model 512

Appendix 16A Probit maximum likelihood estimation 515
16A.1 Probit estimation 515
16A.2 Predicted probabilities 516
16A.3 Marginal effects 517
16A.4 SAS/IML code for probit 517

Appendix A. Math Functions 522
A.1 SAS math and logical operators 522
A.2 Math functions 523
A.3 Matrix manipulation 525

Appendix B. Probability 528
B.1 Probability calculations 528
B.2 Quantiles 531
B.3 Plotting probability density functions 532
B.3.1 Normal distribution 533
B.3.2 t-distribution 535
B.3.3 Chi-square distribution 536
B.3.4 F-distribution 537
B.4 Random numbers 538
Appendix C. Review of Statistical Inference
541
C.1 Histogram 541
C.2 Summary statistics 542
  C.2.1 Estimating higher moments 544
  C.2.2 Jarque-Bera normality test 545
C.3 Confidence interval for the mean 546
C.4 Testing the population mean 546
  C.4.1 A right-tail test 547
  C.4.2 A two-tail test 547
  C.4.3 Automatic tests using PROC TTEST 548
C.5 Maximum likelihood estimation: one parameter 549
  C.5.1 A coin flip example 549
  C.5.2 Statistical inference 551
  C.5.3 Inference in the coin flip example 553
C.6 Maximum likelihood estimation 554
  C.6.1 Exponential distribution example 556
  C.6.2 Gamma distribution example 558
  C.6.3 Testing the gamma distribution 559
C.7 Exponential model using SAS/IML 560
  C.7.1 Direct maximization 561
  C.7.2 Using SAS optimizers 562
  C.7.3 Maximum likelihood estimation of gamma model 565
  C.7.4 Testing the gamma model 567

Index 571