Contents at a Glance

Introduction 1

Part 1: Getting Started with Econometrics 5

Chapter 1: Econometrics: The Economist's Approach to Statistical Analysis 7
Chapter 2: Getting the Hang of Probability 21
Chapter 3: Making Inferences and Testing Hypotheses 39

Part 11: Building the Classical Linear Regression Model 59

Chapter 4: Understanding the Objectives of Regression Analysis 61
Chapter 5: Going Beyond Ordinary with the Ordinary Least Squares Technique 75
Chapter 6: Assumptions of OLS Estimation and the Gauss-Markov Theorem 93
Chapter 7: The Normality Assumption and Inference with OLS 111

Part 111: Working with the Classical Regression Model ... 135

Chapter 8: Functional Form, Specification, and Structural Stability 137
Chapter 9: Regression with Dummy Explanatory Variables 153

Part 1V: Violations of Classical Regression Model Assumptions 173

Chapter 10: Multicollinearity 175
Chapter 11: Heteroskedasticity 191
Chapter 12: Autocorrelation 209

Part V: Discrete and Restricted Dependent Variables in Econometrics 229

Chapter 13: Qualitative Dependent Variables 231
Chapter 14: Limited Dependent Variable Models 253

Part V1: Extending the Basic Econometric Model 265

Chapter 15: Static and Dynamic Models 267
Chapter 16: Diving into Pooled Cross-Section Analysis 281
Chapter 17: Panel Econometrics 291

COPYRIGHTED MATERIAL

Part VII: The Part of Tens 305
Chapter 18: Ten Components of a Good Econometrics Research Project 307
Chapter 19: Ten Common Mistakes in Applied Econometrics 315
Appendix: Statistical Tables 321
Index 327

Table of Contents

Introduction 1

About This Book 1
Foolish Assumptions 2
Icons Used in This Book 3
Beyond the Book 3
Where to Go from Here 4

Part 1: Getting Started with Econometrics 5

Chapter 1: Econometrics: The Economist's Approach to Statistical Analysis 7

Evaluating Economic Relationships 8
Using economic theory to describe outcomes and make predictions 8
Relying on sensible assumptions 8
Applying Statistical Methods to Economic Problems 9
Recognizing the importance of data type, frequency, and aggregation 10
Avoiding the data-mining trap 12
Incorporating quantitative and qualitative information 12
Using Econometric Software: An Introduction to STATA 13
Getting acquainted with STATA 14
Creating new variables 18
Estimating, testing, and predicting 19

Chapter 2: Getting the Hang of Probability 21

Reviewing Random Variables and Probability Distributions 21
Looking at all possibilities: Probability density function (PDF) 22
Summing up the probabilities: Cumulative density function (CDF) 25
Putting variable information together: Bivariate or joint probability density 28
Predicting the future using what you know: Conditional probability density 29

Understanding Summary Characteristics of Random Variables 31
Making generalizations with expected value or mean 31
Measuring variance and standard deviation 33
Looking at relationships with covariance and correlation 35

Chapter 3: Making Inferences and Testing Hypotheses 39

Getting to Know Your Data with Descriptive Statistics 40
Calculating parameters and estimators 40
Determining whether an estimator is good 43
Laying the Groundwork of Prediction with the Normal and Standard Normal Distributions 44
Recognizing usual variables: Normal distribution 44
Putting variables on the same scale: Standard normal distribution (Z) 46
Working with Parts of the Population: Sampling Distributions 47
Simulating and using the central limit theorem 48
Defining the chi-squared (χ^2), t, and F distributions 49
Making Inferences and Testing Hypotheses with Probability Distributions 53
Performing a hypothesis test 54
The confidence interval approach 56
The test of significance approach 57

Part II: Building the Classical Linear Regression Model 59

Chapter 4: Understanding the Objectives of Regression Analysis . . .61

Making a Case for Causality 62
Getting Acquainted with the Population Regression Function (PRF) 63
Setting up the PRF model 64
Walking through an example 66
Collecting and Organizing Data for Regression Analysis 68
Taking a snapshot: Cross-sectional data 68
Looking at the past to explain the present: Time-series data 70
Combining the dimensions of space and time: Panel or longitudinal data 71
Joining multiple snapshots: Pooled cross-sectional data 72

Chapter 5: Going Beyond Ordinary with the Ordinary Least Squares Technique. 75

Defining and Justifying the Least Squares Principle 76
Estimating the Regression Function and the Residuals 77

Obtaining Estimates of the Regression Parameters 78
Finding the formulas necessary to produce optimal coefficient values 79
Calculating the estimated regression coefficients 80
Interpreting Regression Coefficients 84
Seeing what regression coefficients have to say 84
Standardizing regression coefficients 86
Measuring Goodness of Fit 88
Decomposing variance 88
Measuring proportion of variance with R^2 89
Adjusting the goodness of fit in multiple regression 90
Evaluating fit versus quality 92

Chapter 6: Assumptions of OLS Estimation and the Gauss-Markov Theorem 93

Characterizing the OLS Assumptions 93
Linearity in parameters and additive error 94
Random sampling and variability 95
Imperfect linear relationships among the independent variables 96
Error term has a zero conditional mean; correct specification 97
Error term has a constant variance 98
Correlation of error observations is zero 99
Relying on the CLRM Assumptions: The Gauss-Markov Theorem 101
Proving the Gauss-Markov theorem 101
Summarizing the Gauss-Markov theorem 109

Chapter 7: The Normality Assumption and Inference with OLS 111

Describing the Role of the Normality Assumption 111
The error term and the sampling distribution of OLS coefficients 113
Revisiting the standard normal distribution 114
Deriving a chi-squared distribution from the random error 115
OLS standard errors and the t-distribution 119
Testing the Significance of Individual Regression Coefficients 119
Picking an approach 121
Choosing the level of significance and p-values 123
Analyzing Variance to Determine Overall or Joint Significance 123
Normality, variance, and the F distribution 124
The reported F-statistic from OLS 124
Slope coefficients and the relationship between t and F 127
Joint significance for subsets of variables 128
Applying Forecast Error to OLS Predictions 130
Mean prediction and forecast error 130
Variance of mean prediction 131
All predictions are not the same: The prediction confidence interval 133

Part III: Working with the Classical Regression Model ... 135

Chapter 8: Functional Form, Specification, and Structural Stability 137

Employing Alternative Functions ... 138
Quadratic function: Best for finding minimums and maximums ... 138
Cubic functions: Good for inflexion ... 139
Inverse function: Limiting the value of the dependent variable ... 140
Giving Linearity to Nonlinear Models ... 141
Working both sides to keep elasticity constant: The log-log model ... 141
Making investments and calculating rates of return: The log-linear model ... 143
Decreasing the change of the dependent variable: The linear-log model ... 145
Checking for Misspecification ... 147
Too many or too few: Selecting independent variables ... 147
Sensitivity isn't a virtue: Examining misspecification with results stability ... 150

Chapter 9: Regression with Dummy Explanatory Variables 153

Numbers Please! Quantifying Qualitative Information ... 153
Defining a dummy variable when you have only two possible characteristics ... 154
Juggling multiple characteristics with dummy variables ... 155
Finding Average Differences by Using a Dummy Variable ... 156
Specification ... 156
Interpretation ... 157
Combining Quantitative and Qualitative Data in the Regression Model ... 159
Specification ... 159
Interpretation ... 160
Interacting Quantitative and Qualitative Variables ... 162
Specification ... 162
Interpretation ... 162
Interacting Two (or More) Qualitative Characteristics ... 165
Specification ... 165
Interpretation ... 166
Segregate and Integrate: Testing for Significance ... 168
Revisiting the F-test for joint significance ... 168
Revisiting the Chow test ... 169

Part IV: Violations of Classical Regression Model Assumptions 173

Chapter 10: Multicollinearity 175

Distinguishing between the Types of Multicollinearity 175
Pinpointing perfect multicollinearity 176
Zeroing in on high multicollinearity 178
Rules of Thumb for Identifying Multicollinearity 180
Pairwise correlation coefficients 181
Auxiliary regression and the variance inflation factor (VIF) 183
Knowing When and How to Resolve Multicollinearity Issues 185
Get more data 186
Use a new model 186
Expel the problem variable(s) 188

Chapter 11: Heteroskedasticity 191

Distinguishing between Homoskedastic and Heteroskedastic Disturbances 191
Homoskedastic error versus heteroskedastic error 192
The consequences of heteroskedasticity 193
Detecting Heteroskedasticity with Residual Analysis 194
Examining the residuals in graph form 194
Brushing up on the Breusch-Pagan test 196
Getting acquainted with the White test 197
Trying out the Goldfeld-Quandt test 199
Conducting the Park test 201
Correcting Your Regression Model for the Presence of Heteroskedasticity 203
Weighted least squares (WLS) 203
Robust standard errors (also known as White-corrected standard errors) 205

Chapter 12: Autocorrelation 209

Examining Patterns of Autocorrelation 209
Positive versus negative autocorrelation 210
Misspecification and autocorrelation 212
Illustrating the Effect of Autoregressive Errors 212
Analyzing Residuals to Test for Autocorrelation 214
Taking the visual route: Graphical inspection of residuals 215
Using the normal distribution to identify residual sequences: The run test 215

Detecting autocorrelation of an AR(1) process: The Durbin-Watson test 219
Detecting autocorrelation of an AR(q) process: The Breusch-Godfrey test 221
Remedying Harmful Autocorrelation 223
Feasible generalized least squares (FGLS) 223
Serial correlation robust standard errors 225

Part V: Discrete and Restricted Dependent Variables in Econometrics 229

Chapter 13: Qualitative Dependent Variables 231

Modeling Discrete Outcomes with the Linear Probability Model (LPM) 231
Estimating LPM with OLS 232
Interpreting your results 234
Presenting the Three Main LPM Problems 237
Non-normality of the error term 237
Heteroskedasticity 238
Unbounded predicted probabilities 239
Specifying Appropriate Nonlinear Functions: The Probit and Logit Models 240
Working from the standard normal CDF: The probit model 241
Basing off of the logistic CDF: The logit model 242
Using Maximum Likelihood (ML) Estimation 244
Constructing the likelihood function 244
The log transformation and ML estimates 245
Interpreting Probit and Logit Estimates 246
Probit coefficients 249
Logit coefficients 251

Chapter 14: Limited Dependent Variable Models 253

The Nitty-Gritty of Limited Dependent Variables 253
Censored dependent variables 254
Truncated dependent variables 255
Modifying Regression Analysis for Limited Dependent Variables 256
Tobin's Tobit 256
Truncated regression 258
Oh, what the heck if I self select? Heckman's selection bias correction 260

Part VI: Extending the Basic Econometric Model 265

Chapter 15: Static and Dynamic Models 267

Using Contemporaneous and Lagged Variables in Regression Analysis 267
Examining problems with dynamic models 268
Testing and correcting for autocorrelation in dynamic models 270
Projecting Time Trends with OLS 271
Spurious correlation and time series 273
Detrending time-series data 274
Using OLS for Seasonal Adjustments 276
Estimating seasonality effects 276
Deseasonalizing time-series data 278

Chapter 16: Diving into Pooled Cross-Section Analysis 281

Adding a Dynamic Time Element to the Mix 281
Examining intercepts and/or slopes that change over time 282
Incorporating time dummy variables 283
Using Experiments to Estimate Policy Effects with Pooled Cross Sections 285
Benefitting from random assignment: A true experiment 286
Working with predetermined subject groups: A natural (or quasi) experiment 287

Chapter 17: Panel Econometrics 291

Estimating the Uniqueness of Each Individual Unit 291
First difference (FD) transformation 294
Dummy variable (DV) regression 294
Fixed effects (FE) estimator 295
Increasing the Efficiency of Estimation with Random Effects 298
The composite error term and assumptions of random effects model 298
The random effects (RE) estimator 299
Testing Efficiency against Consistency with the Hausman Test 302

Part VII: The Part of Tens 305

Chapter 18: Ten Components of a Good Econometrics Research Project 307

Introducing Your Topic and Posing the Primary Question of Interest 307
Discussing the Relevance and Importance of Your Topic 308

Reviewing the Existing Literature....308
Describing the Conceptual or Theoretical Framework....309
Explaining Your Econometric Model....310
Discussing the Estimation Method(s)....310
Providing a Detailed Description of Your Data....311
Constructing Tables and Graphs to Display Your Results....312
Interpreting the Reported Results....312
Summarizing What You Learned....313

Chapter 19: Ten Common Mistakes in Applied Econometrics.....315

Failing to Use Your Common Sense and Knowledge of Economic Theory....315
Asking the Wrong Questions First....316
Ignoring the Work and Contributions of Others....316
Failing to Familiarize Yourself with the Data....316
Making It Too Complicated....317
Being Inflexible to Real-World Complications....318
Looking the Other Way When You See Bizarre Results....319
Obsessing over Measures of Fit and Statistical Significance....319
Forgetting about Economic Significance....320
Assuming Your Results Are Robust....320

Appendix: Statistical Tables..... 321

The Standard Normal Distribution....321
t-Distribution....322
Chi-Squared Distribution....323
F-Distribution....325
Durbin-Watson d-Statistic....325

Index..... 327