

Contents

Preface	1
1. Introduction	3
1.1. Getting Started	3
1.1.1. Software	3
1.1.2. R Scripts	4
1.1.3. Packages	7
1.1.4. File names and the Working Directory	9
1.1.5. Errors and Warnings	10
1.1.6. Other Resources	10
1.2. Objects in R	11
1.2.1. Basic Calculations and Objects	11
1.2.2. Vectors	12
1.2.3. Special Types of Vectors	15
1.2.4. Naming and Indexing Vectors	16
1.2.5. Matrices	17
1.2.6. Lists	20
1.3. Data Frames and Data Files	21
1.3.1. Data Frames	21
1.3.2. Subsets of Data	22
1.3.3. R Data Files	23
1.3.4. Basic Information on a Data Set	23
1.3.5. Import and Export of Text Files	24
1.3.6. Import and Export of Other Data Formats	25
1.3.7. Data Sets in the Examples	26
1.4. Base Graphics	27
1.4.1. Basic Graphs	27
1.4.2. Customizing Graphs with Options	29
1.4.3. Overlaying Several Plots	30
1.4.4. Legends	31
1.4.5. Exporting to a File	33
1.5. Data Manipulation and Visualization: The Tidyverse	34
1.5.1. Data visualization: ggplot Basics	34
1.5.2. Colors and Shapes in ggplot Graphs	36
1.5.3. Fine Tuning of ggplot Graphs	39
1.5.4. Basic Data Manipulation with dplyr	41
1.5.5. Pipes	43
1.5.6. More Advanced Data Manipulation	44
1.6. Descriptive Statistics	48
1.6.1. Discrete Distributions: Frequencies and Contingency Tables	48
1.6.2. Continuous Distributions: Histogram and Density	51
1.6.3. Empirical Cumulative Distribution Function (ECDF)	52
1.6.4. Fundamental Statistics	54
1.7. Probability Distributions	55
1.7.1. Discrete Distributions	55
1.7.2. Continuous Distributions	57
1.7.3. Cumulative Distribution Function (CDF)	58
1.7.4. Random Draws from Probability Distributions	59
1.8. Confidence Intervals and Statistical Inference	61
1.8.1. Confidence Intervals	61
1.8.2. <i>t</i> Tests	64
1.8.3. <i>p</i> Values	65
1.8.4. Automatic calculations	66
1.9. More Advanced R	70
1.9.1. Conditional Execution	70
1.9.2. Loops	70
1.9.3. Functions	71
1.9.4. Outlook	71
1.10. Monte Carlo Simulation	72
1.10.1. Finite Sample Properties of Estimators	72
1.10.2. Asymptotic Properties of Estimators	75
1.10.3. Simulation of Confidence Intervals and <i>t</i> Tests	78

I. Regression Analysis with Cross-Sectional Data	81	6. Multiple Regression Analysis: Further Issues	137
2. The Simple Regression Model	83	6.1. Model Formulae	137
2.1. Simple OLS Regression	83	6.1.1. Data Scaling: Arithmetic Operations Within a Formula	137
2.2. Coefficients, Fitted Values, and Residuals	88	6.1.2. Standardization: Beta Coefficients	139
2.3. Goodness of Fit	91	6.1.3. Logarithms	140
2.4. Nonlinearities	93	6.1.4. Quadratics and Polynomials	140
2.5. Regression through the Origin and Regression on a Constant	94	6.1.5. ANOVA Tables	142
2.6. Expected Values, Variances, and Standard Errors	96	6.1.6. Interaction Terms	144
2.7. Monte Carlo Simulations	98	6.2. Prediction	146
2.7.1. One sample	98	6.2.1. Confidence Intervals for Predictions	146
2.7.2. Many Samples	100	6.2.2. Prediction Intervals	148
2.7.3. Violation of SLR.4	103	6.2.3. Effect Plots for Nonlinear Specifications	148
2.7.4. Violation of SLR.5	103		
3. Multiple Regression Analysis: Estimation	105	7. Multiple Regression Analysis with Qualitative Regressors	151
3.1. Multiple Regression in Practice	105	7.1. Linear Regression with Dummy Variables as Regressors	151
3.2. OLS in Matrix Form	109	7.2. Logical Variables	152
3.3. Ceteris Paribus Interpretation and Omitted Variable Bias	111	7.3. Factor variables	154
3.4. Standard Errors, Multicollinearity, and VIF	113	7.4. Breaking a Numeric Variable Into Categories	156
4. Multiple Regression Analysis: Inference	117	7.5. Interactions and Differences in Regression Functions Across Groups	158
4.1. The <i>t</i> Test	117		
4.1.1. General Setup	117	8. Heteroscedasticity	161
4.1.2. Standard case	118	8.1. Heteroscedasticity-Robust Inference	161
4.1.3. Other hypotheses	120	8.2. Heteroscedasticity Tests	165
4.2. Confidence Intervals	122	8.3. Weighted Least Squares	168
4.3. Linear Restrictions: <i>F</i> -Tests	123		
4.4. Reporting Regression Results	127	9. More on Specification and Data Issues	173
5. Multiple Regression Analysis: OLS Asymptotics	129	9.1. Functional Form Misspecification	173
5.1. Simulation Exercises	129	9.2. Measurement Error	175
5.1.1. Normally Distributed Error Terms	129	9.3. Missing Data and Nonrandom Samples	178
5.1.2. Non-Normal Error Terms	130	9.4. Outlying Observations	181
5.1.3. (Not) Conditioning on the Regressors	133	9.5. Least Absolute Deviations (LAD) Estimation	182
5.2. LM Test	135		

II. Regression Analysis with Time Series Data	183	14.3. Dummy Variable Regression and Correlated Random Effects	230
10. Basic Regression Analysis with Time Series Data	185	14.4. Robust (Clustered) Standard Errors	234
10.1. Static Time Series Models	185	15. Instrumental Variables Estimation and Two Stage Least Squares	237
10.2. Time Series Data Types in <i>R</i>	186	15.1. Instrumental Variables in Simple Regression Models	237
10.2.1. Equispaced Time Series in <i>R</i>	186	15.2. More Exogenous Regressors	239
10.2.2. Irregular Time Series in <i>R</i> .	187	15.3. Two Stage Least Squares	240
10.3. Other Time Series Models	191	15.4. Testing for Exogeneity of the Regressors	242
10.3.1. The <i>dynlm</i> Package	191	15.5. Testing Overidentifying Restrictions	243
10.3.2. Finite Distributed Lag Models	191	15.6. Instrumental Variables with Panel Data	243
10.3.3. Trends	194	16. Simultaneous Equations Models	247
10.3.4. Seasonality	195	16.1. Setup and Notation	247
11. Further Issues In Using OLS with Time Series Data	197	16.2. Estimation by 2SLS	248
11.1. Asymptotics with Time Series . . .	197	16.3. Joint Estimation of System	249
11.2. The Nature of Highly Persistent Time Series	200	16.4. Outlook: Estimation by 3SLS . . .	251
11.3. Differences of Highly Persistent Time Series	203	17. Limited Dependent Variable Models and Sample Selection Corrections	253
11.4. Regression with First Differences .	204	17.1. Binary Responses	253
12. Serial Correlation and Heteroscedasticity in Time Series Regressions	205	17.1.1. Linear Probability Models .	253
12.1. Testing for Serial Correlation of the Error Term	205	17.1.2. Logit and Probit Models: Estimation	255
12.2. FGLS Estimation	209	17.1.3. Inference	258
12.3. Serial Correlation-Robust Inference with OLS	210	17.1.4. Predictions	259
12.4. Autoregressive Conditional Heteroscedasticity	211	17.1.5. Partial Effects	260
III. Advanced Topics	213	17.2. Count Data: The Poisson Regression Model	263
13. Pooling Cross-Sections Across Time: Simple Panel Data Methods	215	17.3. Corner Solution Responses: The Tobit Model	266
13.1. Pooled Cross-Sections	215	17.4. Censored and Truncated Regression Models	269
13.2. Difference-in-Differences	216	17.5. Sample Selection Corrections . . .	271
13.3. Organizing Panel Data	219	18. Advanced Time Series Topics	273
13.4. Panel-specific computations	220	18.1. Infinite Distributed Lag Models . .	273
13.5. First Differenced Estimator	222	18.2. Testing for Unit Roots	275
14. Advanced Panel Data Methods	225	18.3. Spurious Regression	278
14.1. Fixed Effects Estimation	225	18.4. Cointegration and Error Correction Models	280
14.2. Random Effects Models	227	18.5. Forecasting	281
14.3. Dummy Variable Regression and Correlated Random Effects	230	19. Carrying Out an Empirical Project	285
14.4. Robust (Clustered) Standard Errors	234	19.1. Working with <i>R</i> Scripts	285

19.2. Logging Output in Text Files	287
19.3. Formatted Documents and Reports with R Markdown	287
19.3.1. Basics	287
19.3.2. Advanced Features	288
19.3.3. Bottom Line	291
19.4. Combining R with LaTeX	293
19.4.1. Automatic Document Generation using <i>Sweave</i> and <i>knitr</i>	293
19.4.2. Separating R and L ^A T _E X code	298

IV. Appendices 303

R Scripts 305

1. Scripts Used in Chapter 01	305
2. Scripts Used in Chapter 02	317
3. Scripts Used in Chapter 03	324
4. Scripts Used in Chapter 04	326
5. Scripts Used in Chapter 05	327
6. Scripts Used in Chapter 06	329
7. Scripts Used in Chapter 07	332
8. Scripts Used in Chapter 08	333
9. Scripts Used in Chapter 09	335
10. Scripts Used in Chapter 10	338
11. Scripts Used in Chapter 11	340
12. Scripts Used in Chapter 12	342
13. Scripts Used in Chapter 13	344
14. Scripts Used in Chapter 14	346
15. Scripts Used in Chapter 15	348
16. Scripts Used in Chapter 16	350
17. Scripts Used in Chapter 17	350
18. Scripts Used in Chapter 18	355
19. Scripts Used in Chapter 19	357

Bibliography 359

List of Wooldridge (2019) Examples 361

Index 363