

# Contents

Preface . . . . .	1		
<b>1. Introduction</b>	<b>3</b>		
1.1. Getting Started . . . . .	3		
1.1.1. Software . . . . .	3		
1.1.2. <i>Python</i> Scripts . . . . .	4		
1.1.3. Modules . . . . .	8		
1.1.4. File Names and the Working Directory . . . . .	9		
1.1.5. Errors and Warnings . . . . .	9		
1.1.6. Other Resources . . . . .	10		
1.2. Objects in <i>Python</i> . . . . .	11		
1.2.1. Variables . . . . .	11		
1.2.2. Objects in <i>Python</i> . . . . .	11		
1.2.3. Objects in <i>numpy</i> . . . . .	15		
1.2.4. Objects in <i>pandas</i> . . . . .	19		
1.3. External Data . . . . .	23		
1.3.1. Data Sets in the Examples . . . . .	23		
1.3.2. Import and Export of Data Files . . . . .	24		
1.3.3. Data from other Sources . . . . .	26		
1.4. Base Graphics with <i>matplotlib</i> . . . . .	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. Exporting to a File . . . . .	31		
1.5. Descriptive Statistics . . . . .	33		
1.5.1. Discrete Distributions: Frequencies and Contingency Tables . . . . .	33		
1.5.2. Continuous Distributions: Histogram and Density . . . . .	38		
1.5.3. Empirical Cumulative Distribution Function (ECDF) . . . . .	40		
1.5.4. Fundamental Statistics . . . . .	42		
1.6. Probability Distributions . . . . .	44		
1.6.1. Discrete Distributions . . . . .	44		
1.6.2. Continuous Distributions . . . . .	47		
1.6.3. Cumulative Distribution Function (CDF) . . . . .	47		
1.6.4. Random Draws from Probability Distributions . . . . .	50		
1.7. Confidence Intervals and Statistical Inference . . . . .	52		
1.7.1. Confidence Intervals . . . . .	52		
1.7.2. <i>t</i> Tests . . . . .	55		
1.7.3. <i>p</i> Values . . . . .	57		
1.8. Advanced <i>Python</i> . . . . .	60		
1.8.1. Conditional Execution . . . . .	60		
1.8.2. Loops . . . . .	60		
1.8.3. Functions . . . . .	61		
1.8.4. Object Orientation . . . . .	62		
1.8.5. Outlook . . . . .	66		
1.9. Monte Carlo Simulation . . . . .	66		
1.9.1. Finite Sample Properties of Estimators . . . . .	66		
1.9.2. Asymptotic Properties of Estimators . . . . .	69		
1.9.3. Simulation of Confidence Intervals and <i>t</i> Tests . . . . .	70		
		<b>1. Regression Analysis with Cross-Sectional Data</b>	<b>75</b>
		<b>2. The Simple Regression Model</b>	<b>77</b>
		2.1. Simple OLS Regression . . . . .	77
		2.2. Coefficients, Fitted Values, and Residuals . . . . .	82
		2.3. Goodness of Fit . . . . .	85
		2.4. Nonlinearities . . . . .	88
		2.5. Regression through the Origin and Regression on a Constant . . . . .	89
		2.6. Expected Values, Variances, and Standard Errors . . . . .	91
		2.7. Monte Carlo Simulations . . . . .	94
		2.7.1. One Sample . . . . .	94
		2.7.2. Many Samples . . . . .	96
		2.7.3. Violation of SLR.4 . . . . .	98
		2.7.4. Violation of SLR.5 . . . . .	99
		<b>3. Multiple Regression Analysis: Estimation</b>	<b>101</b>
		3.1. Multiple Regression in Practice . . . . .	101
		3.2. OLS in Matrix Form . . . . .	107
		3.3. Ceteris Paribus Interpretation and Omitted Variable Bias . . . . .	110
		3.4. Standard Errors, Multicollinearity, and VIF . . . . .	112

<b>4. Multiple Regression Analysis: Inference</b>	<b>115</b>	8.2. Heteroscedasticity Tests . . . . .	168
4.1. The <i>t</i> Test . . . . .	115	8.3. Weighted Least Squares . . . . .	171
4.1.1. General Setup . . . . .	115	<b>9. More on Specification and Data Issues</b>	<b>177</b>
4.1.2. Standard Case . . . . .	116	9.1. Functional Form Misspecification .	177
4.1.3. Other Hypotheses . . . . .	118	9.2. Measurement Error . . . . .	180
4.2. Confidence Intervals . . . . .	121	9.3. Missing Data and Nonrandom Samples . . . . .	184
4.3. Linear Restrictions: <i>F</i> -Tests . . . . .	123	9.4. Outlying Observations . . . . .	188
<b>5. Multiple Regression Analysis: OLS Asymptotics</b>	<b>127</b>	9.5. Least Absolute Deviations (LAD) Estimation . . . . .	190
5.1. Simulation Exercises . . . . .	127	<b>II. Regression Analysis with Time Series Data</b>	<b>191</b>
5.1.1. Normally Distributed Error Terms . . . . .	127	<b>10. Basic Regression Analysis with Time Series Data</b>	<b>193</b>
5.1.2. Non-Normal Error Terms .	128	10.1. Static Time Series Models . . . . .	193
5.1.3. (Not) Conditioning on the Regressors . . . . .	132	10.2. Time Series Data Types in <i>Python</i> .	194
5.2. LM Test . . . . .	135	10.2.1. Equispaced Time Series in <i>Python</i> . . . . .	194
<b>6. Multiple Regression Analysis: Further Issues</b>	<b>137</b>	10.2.2. Irregular Time Series in <i>Python</i> . . . . .	197
6.1. Model Formulae . . . . .	137	10.3. Other Time Series Models . . . . .	199
6.1.1. Data Scaling: Arithmetic Operations Within a Formula	137	10.3.1. Finite Distributed Lag Models . . . . .	199
6.1.2. Standardization: Beta Coef- ficients . . . . .	138	10.3.2. Trends . . . . .	201
6.1.3. Logarithms . . . . .	140	10.3.3. Seasonality . . . . .	202
6.1.4. Quadratics and Polynomials	140	<b>11. Further Issues in Using OLS with Time Series Data</b>	<b>205</b>
6.1.5. Hypothesis Testing . . . . .	142	11.1. Asymptotics with Time Series . . .	205
6.1.6. Interaction Terms . . . . .	143	11.2. The Nature of Highly Persistent Time Series . . . . .	210
6.2. Prediction . . . . .	144	11.3. Differences of Highly Persistent Time Series . . . . .	213
6.2.1. Confidence and Prediction Intervals for Predictions . .	144	11.4. Regression with First Differences .	213
6.2.2. Effect Plots for Nonlinear Specifications . . . . .	147	<b>12. Serial Correlation and Heteroscedas- ticity in Time Series Regressions</b>	<b>217</b>
<b>7. Multiple Regression Analysis with Qualitative Regressors</b>	<b>151</b>	12.1. Testing for Serial Correlation of the Error Term . . . . .	217
7.1. Linear Regression with Dummy Variables as Regressors . . . . .	151	12.2. FGLS Estimation . . . . .	222
7.2. Boolean Variables . . . . .	154	12.3. Serial Correlation-Robust Infer- ence with OLS . . . . .	223
7.3. Categorical Variables . . . . .	155	12.4. Autoregressive Conditional Het- eroscedasticity . . . . .	224
7.3.1. ANOVA Tables . . . . .	157		
7.4. Breaking a Numeric Variable Into Categories . . . . .	159		
7.5. Interactions and Differences in Re- gression Functions Across Groups	161		
<b>8. Heteroscedasticity</b>	<b>165</b>		
8.1. Heteroscedasticity-Robust Inference	165		

<b>III. Advanced Topics</b>	<b>227</b>	<b>18. Advanced Time Series Topics</b>	<b>289</b>
13. Pooling Cross-Sections Across Time: Simple Panel Data Methods	229	18.1. Infinite Distributed Lag Models . . .	289
13.1. Pooled Cross-Sections . . . . .	229	18.2. Testing for Unit Roots . . . . .	291
13.2. Difference-in-Differences . . . . .	230	18.3. Spurious Regression . . . . .	292
13.3. Organizing Panel Data . . . . .	233	18.4. Cointegration and Error Correction Models . . . . .	295
13.4. First Differenced Estimator . . . . .	234	18.5. Forecasting . . . . .	295
14. Advanced Panel Data Methods	239	<b>19. Carrying Out an Empirical Project</b>	<b>299</b>
14.1. Fixed Effects Estimation . . . . .	239	19.1. Working with <i>Python</i> Scripts . . .	299
14.2. Random Effects Models . . . . .	240	19.2. Logging Output in Text Files . . .	301
14.3. Dummy Variable Regression and Correlated Random Effects . . . . .	244	19.3. Formatted Documents with Jupyter Notebook . . . . .	302
14.4. Robust (Clustered) Standard Errors	247	19.3.1. Getting Started . . . . .	302
15. Instrumental Variables Estimation and Two Stage Least Squares	249	19.3.2. Cells . . . . .	302
15.1. Instrumental Variables in Simple Regression Models . . . . .	249	19.3.3. Markdown Basics . . . . .	303
15.2. More Exogenous Regressors . . . . .	251	<b>IV. Appendices</b>	<b>309</b>
15.3. Two Stage Least Squares . . . . .	254	<b>Python Scripts</b>	<b>311</b>
15.4. Testing for Exogeneity of the Regressors . . . . .	256	1. Scripts Used in Chapter 01 . . . . .	311
15.5. Testing Overidentifying Restrictions	257	2. Scripts Used in Chapter 02 . . . . .	334
15.6. Instrumental Variables with Panel Data . . . . .	259	3. Scripts Used in Chapter 03 . . . . .	343
16. Simultaneous Equations Models	261	4. Scripts Used in Chapter 04 . . . . .	347
16.1. Setup and Notation . . . . .	261	5. Scripts Used in Chapter 05 . . . . .	349
16.2. Estimation by 2SLS . . . . .	262	6. Scripts Used in Chapter 06 . . . . .	352
16.3. Outlook: Estimation by 3SLS . . . . .	263	7. Scripts Used in Chapter 07 . . . . .	357
17. Limited Dependent Variable Models and Sample Selection Corrections	265	8. Scripts Used in Chapter 08 . . . . .	361
17.1. Binary Responses . . . . .	265	9. Scripts Used in Chapter 09 . . . . .	365
17.1.1. Linear Probability Models . . . . .	265	10. Scripts Used in Chapter 10 . . . . .	371
17.1.2. Logit and Probit Models: Estimation . . . . .	267	11. Scripts Used in Chapter 11 . . . . .	374
17.1.3. Inference . . . . .	270	12. Scripts Used in Chapter 12 . . . . .	378
17.1.4. Predictions . . . . .	271	13. Scripts Used in Chapter 13 . . . . .	383
17.1.5. Partial Effects . . . . .	273	14. Scripts Used in Chapter 14 . . . . .	386
17.2. Count Data: The Poisson Regression Model . . . . .	276	15. Scripts Used in Chapter 15 . . . . .	390
17.3. Corner Solution Responses: The Tobit Model . . . . .	279	16. Scripts Used in Chapter 16 . . . . .	395
17.4. Censored and Truncated Regression Models . . . . .	281	17. Scripts Used in Chapter 17 . . . . .	396
17.5. Sample Selection Corrections . . . . .	286	18. Scripts Used in Chapter 18 . . . . .	405
		19. Scripts Used in Chapter 19 . . . . .	409
		<b>Bibliography</b>	<b>411</b>
		<b>List of Wooldridge (2019) Examples</b>	<b>413</b>
		<b>Index</b>	<b>415</b>