THE FUNDAMENTALS OF POLITICAL SCIENCE RESEARCH

Paul M. Kellstedt’s and Guy D. Whitten’s *The Fundamentals of Political Science Research* provides an introduction to the scientific study of politics, supplying students with the basic tools needed to be both critical consumers and producers of scholarly research in political science. The book begins with a discussion of what it means to take a scientific approach to the study of politics. At the core of such an approach is the development of causal theories. Because there is no magic formula by which theories are developed, the authors present a series of strategies and develop an integrated approach to research design and empirical analyses that allows students to determine the plausibility of their causal theories. The text’s accessible presentation of mathematical concepts and regression models with two or more independent variables is a key component to this process, along with the integration of examples from political science and the real world to help students grasp the key concepts.

Paul M. Kellstedt is Associate Professor of Political Science and Director of the American Politics Program at Texas A&M University. He is the author of *The Mass Media and the Dynamics of American Racial Attitudes* (2003), which won the Goldsmith Book Prize. Professor Kellstedt is also the author or co-author of articles appearing in scholarly journals such as *American Journal of Political Science*, *British Journal of Political Science*, and *Political Analysis*, as well as several book chapters. He has been an Academic Visitor at Nuffield College, Oxford, and a Harvard University Fellow in the Joan Shorenstein Center on the Press, Politics, and Public Policy in the Kennedy School of Government.

Guy D. Whitten is Associate Professor of Political Science and Director of the European Union Center at Texas A&M University. He has published a variety of papers in scholarly journals, including *American Journal of Political Science*, *British Journal of Political Science*, and *Electoral Studies*. Professor Whitten serves on the editorial board of *Electoral Studies* and has previously served on the editorial boards of *Journal of Politics* and *Political Research Quarterly*. He has been a visiting researcher at the University of Amsterdam and is a frequent instructor at the Summer School for Social Science Data Analysis and Collection at the University of Essex in the United Kingdom.
THE FUNDAMENTALS OF
Political Science Research

Paul M. Kellstedt
Texas A&M University

Guy D. Whitten
Texas A&M University
Dedicated to

Lyman A. Kellstedt, Charmaine C. Kellstedt,
David G. Whitten, and Jo Wright-Whitten,
the best teachers we ever had

– PMK and GDW
Contents

Figures xv
Tables xvii
Acknowledgments xix

1 The Scientific Study of Politics .................................. 1

Overview 1
1.1 Political Science? 1
1.2 Approaching Politics Scientifically: The Search for Causal Explanations 3
1.3 Thinking about the World in Terms of Variables and Causal Explanations 7
1.4 Models of Politics 14
1.5 Rules of the Road to Scientific Knowledge about Politics 15
  1.5.1 Make Your Theories Causal 15
  1.5.2 Don’t Let Data Alone Drive Your Theories 16
  1.5.3 Consider Only Empirical Evidence 17
  1.5.4 Avoid Normative Statements 17
  1.5.5 Pursue Both Generality and Parsimony 18
1.6 A Quick Look Ahead 18
Concepts Introduced in This Chapter 19
Exercises 20

2 The Art of Theory Building .................................... 22

Overview 22
2.1 Good Theories Come from Good Theory-Building Strategies 22
2.2 Identifying Interesting Variation 23
  2.2.1 Time-Series Example 24
  2.2.2 Cross-Sectional Example 25
2.3 Learning to Use Your Knowledge 26
  2.3.1 Moving from a Specific Event to More General Theories 26
Contents

2.3.2 Know Local, Think Global: Can You Drop the Proper Nouns? 27

2.4 Examine Previous Research 28
2.4.1 What Did the Previous Researchers Miss? 29
2.4.2 Can Their Theory Be Applied Elsewhere? 29
2.4.3 If We Believe Their Findings, Are There Further Implications? 30
2.4.4 How Might This Theory Work at Different Levels of Aggregation (Micro⇐⇒Macro)? 30

2.5 Think Formally about the Causes That Lead to Variation in Your Dependent Variable 31
2.5.1 Utility and Expected Utility 32
2.5.2 The Puzzle of Turnout 34

2.6 Think about the Institutions: The Rules Usually Matter 36
2.6.1 Legislative Rules 36
2.6.2 The Rules Matter! 38

2.7 Extensions 39

2.8 How Do I Know If I Have a “Good” Theory? 40
2.8.1 Is Your Theory Causal? 40
2.8.2 Can You Test Your Theory on Data That You Have Not Yet Observed? 41
2.8.3 How General Is Your Theory? 41
2.8.4 How Parsimonious Is Your Theory? 41
2.8.5 How New Is Your Theory? 41
2.8.6 How Nonobvious Is Your Theory? 42

2.9 Conclusion 42

Concepts Introduced in This Chapter 43
Exercises 43

3 Evaluating Causal Relationships .......................... 45

Overview 45

3.1 Causality and Everyday Language 45

3.2 Four Hurdles along the Route to Establishing Causal Relationships 48
3.2.1 Putting It All Together – Adding Up the Answers to Our Four Questions 50
3.2.2 Identifying Causal Claims Is an Essential Thinking Skill 50
3.2.3 What Are the Consequences of Failing to Control for Other Possible Causes? 53

3.3 Why Is Studying Causality So Important? Three Examples from Political Science 54
3.3.1 Life Satisfaction and Democratic Stability 54
3.3.2 School Choice and Student Achievement 55
3.3.3 Electoral Systems and the Number of Political Parties 57
3.4 Why Is Studying Causality So Important? Three Examples from Everyday Life

3.4.1 Alcohol Consumption and Income 61
3.4.2 Treatment Choice and Breast Cancer Survival 62
3.4.3 Explicit Lyrics and Teen Sexual Behavior 63

3.5 Wrapping Up 65
Concepts Introduced in This Chapter 65
Exercises 65

4 Research Design ................................ 67
Overview 67
4.1 Comparison as the Key to Establishing Causal Relationships 67
4.2 Experimental Research Designs 68
4.2.1 “Random Assignment” versus “Random Sampling” 74
4.2.2 Are There Drawbacks to Experimental Research Designs? 74
4.3 Observational Studies (in Two Flavors) 77
4.3.1 Datum, Data, Data Set 79
4.3.2 Cross-Sectional Observational Studies 81
4.3.3 Time-Series Observational Studies 82
4.3.4 The Major Difficulty with Observational Studies 83
4.4 Summary 83
Concepts Introduced in This Chapter 84
Exercises 84

5 Measurement .................................... 86
Overview 86
5.1 Why Measurement Matters 86
5.2 Social Science Measurement: The Varying Challenges of Quantifying Humanity 88
5.3 Problems in Measuring Concepts of Interest 91
5.3.1 Conceptual Clarity 91
5.3.2 Reliability 92
5.3.3 Measurement Bias and Reliability 93
5.3.4 Validity 94
5.3.5 The Relationship between Validity and Reliability 95
5.4 Controversy 1: Measuring Democracy 96
5.5 Controversy 2: Measuring Political Tolerance 99
5.6 Are There Consequences to Poor Measurement? 101
5.7 Conclusions 101
Concepts Introduced in This Chapter 102
Exercises 102

6 Descriptive Statistics and Graphs ............................. 104
Overview 104
6.1 Know Your Data 104
Contents

9.3 Which Line Fits Best? Estimating the Regression Line 162
9.4 Measuring Our Uncertainty about the OLS Regression Line 165
  9.4.1 Goodness-of-Fit: Root Mean-Squared Error 167
  9.4.2 Goodness-of-Fit: R-Squared Statistic 167
  9.4.3 Is That a “Good” Goodness-of-Fit? 169
  9.4.4 Uncertainty about Individual Components of the Sample Regression Model 169
  9.4.5 Confidence Intervals about Parameter Estimates 171
  9.4.6 Hypothesis Testing: Overview 172
  9.4.7 Two-Tailed Hypothesis Tests 173
  9.4.8 The Relationship between Confidence Intervals and Two-Tailed Hypothesis Tests 175
  9.4.9 One-Tailed Hypothesis Tests 175
9.5 Assumptions, More Assumptions, and Minimal Mathematical Requirements 177
  9.5.1 Assumptions about the Population Stochastic Component 177
  9.5.2 Assumptions about Our Model Specification 180
  9.5.3 Minimal Mathematical Requirements 181
  9.5.4 How Can We Make All of These Assumptions? 181
Concepts Introduced in This Chapter 182
Exercises 182

10 Multiple Regression Models I: The Basics .................................. 183
  Overview 183
  10.1 Modeling Multivariate Reality 183
  10.2 The Population Regression Function 184
  10.3 From Two-Variable to Multiple Regression 184
  10.4 What Happens When We Fail to Control for Z? 188
    10.4.1 An Additional Minimal Mathematical Requirement in Multiple Regression 192
  10.5 Interpreting Multiple Regression 193
  10.6 Which Effect Is “Biggest”? 196
  10.7 Statistical and Substantive Significance 198
  10.8 Implications 199
  Concepts Introduced in This Chapter 200
  Exercises 200

11 Multiple Regression Models II: Crucial Extensions .................... 202
  Overview 202
  11.1 Extensions of OLS 202
  11.2 Being Smart with Dummy Independent Variables in OLS 203
    11.2.1 Using Dummy Variables to Test Hypotheses about a Categorical Independent Variable with Only Two Values 203
    11.2.2 Using Dummy Variables to Test Hypotheses about a Categorical Independent Variable with More Than Two Values 207
## 11.3 Testing Interactive Hypotheses with Dummy Variables 210

### 11.4 Dummy Dependent Variables 212

- **11.4.1 The Linear Probability Model** 212
- **11.4.2 Binomial Logit and Binomial Probit** 215
- **11.4.3 Goodness-of-Fit with Dummy Dependent Variables** 219

### 11.5 Outliers and Influential Cases in OLS 220

- **11.5.1 Identifying Influential Cases** 221
- **11.5.2 Dealing with Influential Cases** 224

### 11.6 Multicollinearity 225

- **11.6.1 How Does Multicollinearity Happen?** 226
- **11.6.2 Detecting Multicollinearity** 227
- **11.6.3 Multicollinearity: A Simulated Example** 228
- **11.6.4 Multicollinearity: A Real-World Example** 230
- **11.6.5 Multicollinearity: What Should I Do?** 232

### 11.7 Being Careful with Time Series 233

- **11.7.1 Time-Series Notation** 233
- **11.7.2 Memory and Lags in Time-Series Analysis** 234
- **11.7.3 Trends and the Spurious Regression Problem** 236
- **11.7.4 The Differenced Dependent Variable** 239
- **11.7.5 The Lagged Dependent Variable** 241

### 11.8 Wrapping Up 242

**Concepts Introduced in This Chapter** 243

## 12 Multiple Regression Models III: Applications . . . . . . . . . . . . . . . . . . . . . 244

### Overview 244

- **12.1 Why Controlling for \textit{Z} Matters** 244
- **12.2 Example 1: The Economy and Presidential Popularity** 245
- **12.3 Example 2: Politics, Economics, and Public Support for Democracy** 248
- **12.4 Example 3: Competing Theories of How Politics Affects International Trade** 251
- **12.5 Conclusions** 253

**Concepts Introduced in This Chapter** 254

**Exercises** 254

**Appendix A. Critical Values of \textit{\chi}^2** 255

**Appendix B. Critical Values of \textit{t}** 256

**Appendix C. The \textit{A} Link Function for BNL Models** 257

**Appendix D. The \textit{\Phi} Link Function for BNP Models** 259

**Bibliography** 261

**Index** 265
Figures

1.1 The road to scientific knowledge  
1.2 From theory to hypothesis  
1.3 What would you expect to see based on the theory of economic voting?  
1.4 What would you expect to see based on the theory of economic voting? Two hypothetical cases  
1.5 What would you expect to see based on the theory of economic voting?  
1.6 What would you expect to see based on the theory of economic voting? Two hypothetical cases  
2.1 Presidential approval, 1995–2005  
2.2 Military spending in 2005  
2.3 Gross U.S. government debt as a percentage of GDP, 1960–2004  
2.4 Women as a percentage of members of parliament, 2004  
3.1 The path to evaluating a causal relationship  
3.2 Theoretical causes of the number of parties in legislatures  
3.3 Nazi vote and the number of parties winning seats in Weimar Republic elections, 1919–1933  
3.4 Number of parties winning seats in German Bundestag elections, 1949–2002  
4.1 The possibly confounding effects of a healthy lifestyle on the aspirin–blood-pressure relationship  
5.1 Reliability, validity, and hypothesis testing  
5.2 Polity IV score for Pakistan  
6.1 Pie graph of religious identification, NES 2004  
6.2 Bar graph of religious identification, NES 2004  
6.3 Example output from Stata’s “summarize” command with “detail” option  
6.4 Box-whisker plot of incumbent-party presidential vote percentage, 1880–2004  
6.5 Histogram of incumbent-party presidential vote percentage, 1880–2004
Figures

6.6 Histograms of incumbent-party presidential vote percentage, 1880–2004, depicted with 2 and then 10 blocks 117
6.7 Kernel density plot of incumbent-party presidential vote percentage, 1880–2004 117
7.1 The normal probability distribution 123
7.2 The 68–95–99 rule 124
7.3 Frequency distribution of 600 rolls of a die 124
8.1 Box-whisker plot of Government Duration for majority and minority governments 147
8.2 Kernel density plot of Government Duration for majority and minority governments 147
8.3 Scatter plot of change in GDP and incumbent-party vote share 151
8.4 Scatter plot of change in GDP and incumbent-party vote share with mean-delimited quadrants 152
8.5 What is wrong with this table? 156
9.1 Scatter plot of change in GDP and incumbent-party vote share 162
9.2 Three possible lines 163
9.3 OLS regression line through scatter plot with mean-delimited quadrants 165
9.4 Stata results for two-variable regression model of VOTE = α + β × GROWTH 166
9.5 Venn diagram of variance and covariance for X and Y 168
10.1 Venn diagram in which X, Y, and Z are correlated 191
10.2 Venn diagram in which X and Y are correlated with Z, but not with each other 192
11.1 Stata output when we include both gender dummy variables in our model 204
11.2 Regression lines from the interactive model 207
11.3 Regression lines from the interactive model 212
11.4 Three different models of Bush vote 218
11.5 Stata lvr2plot for the model presented in Table 11.8 222
11.6 OLS line with scatter plot for Florida 2000 223
11.7 Venn diagram with multicollinearity 226
11.8 The growth of golf and the decline of the American family, 1947–2002 237
11.9 The growth of the U.S. economy and the decline of the family, 1947–2002 238
11.10 First differences of the number of golf courses and percentage of married families, 1947–2002 240
12.1 A simple causal model of the relationship between the economy and presidential popularity 245
12.2 A revised model of presidential popularity 246
### Tables

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Example of cross-sectional data</td>
<td>80</td>
</tr>
<tr>
<td>4.2</td>
<td>Example of time-series data</td>
<td>81</td>
</tr>
<tr>
<td>6.1</td>
<td>Frequency table for religious identification in the 2004 NES</td>
<td>109</td>
</tr>
<tr>
<td>6.2</td>
<td>Values of incumbent vote ranked from smallest to largest</td>
<td>113</td>
</tr>
<tr>
<td>6.3</td>
<td>Median incomes of the 50 states, 2004–2005</td>
<td>119</td>
</tr>
<tr>
<td>8.1</td>
<td>Variable types and appropriate bivariate hypothesis tests</td>
<td>135</td>
</tr>
<tr>
<td>8.2</td>
<td>Union households and vote in the 2004 U.S. presidential election</td>
<td>140</td>
</tr>
<tr>
<td>8.3</td>
<td>Gender and vote in the 2004 U.S. presidential election: Hypothetical scenario</td>
<td>141</td>
</tr>
<tr>
<td>8.4</td>
<td>Gender and vote in the 2004 U.S. presidential election: Expectations for hypothetical scenario if there were no relationship</td>
<td>142</td>
</tr>
<tr>
<td>8.5</td>
<td>Gender and vote in the 2004 U.S. presidential election</td>
<td>142</td>
</tr>
<tr>
<td>8.6</td>
<td>Gender and vote in the 2004 U.S. presidential election: Calculating the expected cell values if gender and presidential vote are unrelated</td>
<td>142</td>
</tr>
<tr>
<td>8.7</td>
<td>Gender and vote in the 2004 U.S. presidential election</td>
<td>142</td>
</tr>
<tr>
<td>8.8</td>
<td>Gender and vote in the 2004 U.S. presidential election</td>
<td>143</td>
</tr>
<tr>
<td>8.9</td>
<td>Government type and government duration</td>
<td>149</td>
</tr>
<tr>
<td>8.10</td>
<td>Contributions of individual election years to the covariance calculation</td>
<td>153</td>
</tr>
<tr>
<td>8.11</td>
<td>Covariance table for economic growth and incumbent-party presidential vote, 1880–2004</td>
<td>154</td>
</tr>
<tr>
<td>9.1</td>
<td>Measures of total residuals for three different lines</td>
<td>164</td>
</tr>
<tr>
<td>10.1</td>
<td>Three regression models of U.S. presidential elections</td>
<td>193</td>
</tr>
<tr>
<td>10.2</td>
<td>Bias in $\hat{\beta}_1$ when the true population model is $Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i + u_i$ but we leave out $Z$</td>
<td>201</td>
</tr>
<tr>
<td>11.1</td>
<td>Two models of the effects of gender and income on Hillary Clinton Thermometer scores</td>
<td>205</td>
</tr>
<tr>
<td>11.2</td>
<td>Religious Identification in the 1996 NES</td>
<td>208</td>
</tr>
</tbody>
</table>
Tables

11.3 The same model of religion and income on Hillary Clinton Thermometer scores with different reference categories 209
11.4 The effects of gender and feelings toward the women’s movement on Hillary Clinton Thermometer scores 211
11.5 The effects of partisanship and performance evaluations on votes for Bush in 2004 213
11.6 The effects of partisanship and performance evaluations on votes for Bush in 2004: Three different types of models 217
11.7 Classification table from LPM of the effects of partisanship and performance evaluations on votes for Bush in 2004 219
11.8 Votes for Gore and Buchanan in Florida counties in the 2000 U.S. presidential election 222
11.9 The five largest (absolute-value) DFBETA scores for $\beta$ from the model presented in Table 11.8 224
11.10 Votes for Gore and Buchanan in Florida counties in the 2000 U.S. presidential election 225
11.11 Random draws of increasing size from a population with substantial multicollinearity 230
11.12 Pairwise correlations between independent variables 231
11.13 Model results from random draws of increasing size from the 2004 NES 232
11.14 Golf and the decline of the family, 1947–2002 238
11.15 GDP and the decline of the family, 1947–2002 239
12.1 Excerpts from the table of MacKuen, Erikson, and Stimson on the relationship between the economy and presidential popularity 246
12.2 Excerpts from the table of Evans and Whitefield on the relationship between the economy and support for democracy 250
12.3 Excerpts from the table of Morrow, Siverson, and Tabares on the political causes of international trade 252
An inevitable part of the production of a book like this is an accumulation of massive intellectual debts. We have been overwhelmed by both the quality and quantity of help that we have received from our professional (and even personal) contacts as we have gone through every stage of this project.

This book arose out of more than 20 years of combined teaching experience at Brown University, the University of California, Los Angeles, the University of Essex, the University of Minnesota, and Texas A&M University. We tried out most of the examples in this book on numerous classes of students before we refined them into their present state. We thus owe a debt to every student who raised his or her hand or showed us a furrowed brow as we worked our way through these attempts to explain the complicated processes of scientifically studying politics.

More immediately, this project came out of separate and skeptical conversations that each author had with Ed Parsons during his visit to Texas A&M in the spring of 2006. Without Ed’s perfect balance of candor and encouragement, this book would not have been started. At every stage in the process he has helped us immensely. He obtained three sets of superbly helpful reviews and seemed always to know the right times to be in and out of touch as we worked our way through them. It has been a tremendous pleasure to work with Ed on the book.

Throughout the process of writing this book, we got a steady stream of support, understanding, and patience from Christine, Deb, Abigail, and Elizabeth. We thank them for putting up with our crazy hours and for helping us to keep things in perspective as we worked on this project.

For both authors, the lines separating family, friends, and professional colleagues are pretty blurry. We relied on our combined networks quite heavily at every stage in the production of this book. Early in the process of putting the manuscript together, we received sage advice from Jeff Gill about textbook writing for social scientists and how to handle early versions of our chapters. Our fathers, Lyman A. “Bud” Kellstedt and David Kellstedt
G. Whitten, provided their own unique and valuable perspectives on early drafts of the book. In separate but related ongoing conversations, John Transue and Alan M. Brookhart engaged us in lengthy debates about the nature of experiments, quasi-experiments, and observational studies. Other colleagues and friends provided input that also improved this book, including Harold Clarke, Geoffrey Evans, John Jackson, Marisa Kellam, Eric Lawrence, Christine Lipsmeyer, Evan Parker-Stephen, David Peterson, James Rogers, Randy Stevenson, Georg Vanberg, Rilla Whitten, and Jenifer Whitten-Woodring.

Despite all of this help, we remain solely responsible for any deficiencies that persist in the book. We look forward to hearing about them from you so that we can make future editions of this book better.

Throughout the process of writing this book, we have been mindful of how our thinking has been shaped by our teachers at a variety of levels. We are indebted to them in ways that are difficult to express. In particular, Guy Whitten thanks the following, all from his days at the University of Rochester: Larry M. Bartels, Richard Niemi, G. Bingham Powell, Lynda Powell, William H. Riker, and David Weimer. Paul Kellstedt thanks Al Reynolds and Bob Terbog of Calvin College; Michael Lewis-Beck, Vicki Hesli, and Jack Wright at the University of Iowa; and Jim Stimson and John Freeman at the University of Minnesota.

Although we have learned much from the aforementioned professors, we owe our largest debt to our parents: Lyman A. “Bud” Kellstedt, Charmaine C. Kellstedt, David G. Whitten, and Jo Wright-Whitten. We dedicate this book to the four of them – the best teachers we ever had.
THE FUNDAMENTALS OF POLITICAL SCIENCE RESEARCH