Statistics in Corpus Linguistics

Do you use language corpora in your research or study, but find that you struggle with statistics? This practical introduction will equip you to understand the key principles of statistical thinking and apply these concepts to your own research, without the need for prior statistical knowledge. The book gives step-by-step guidance through the process of statistical analysis and provides multiple examples of how statistical techniques can be used to analyse and visualize linguistic data. It also includes a useful selection of discussion questions and exercises which you can use to check your understanding. The book comes with a companion website, which provides additional materials (including answers to exercises, datasets, advanced materials, teaching slides etc.) and Lancaster Stats Tools online (http://corpora.lancs.ac.uk/stats), a free click-and-analyse statistical tool for easy calculation of the statistical measures discussed in the book.

VACLAV BREZINA is a senior lecturer at the Department of Linguistics and English Language, Lancaster University. He specializes in corpus linguistics, statistics and applied linguistics, and has designed a number of different tools for corpus analysis.
Statistics in Corpus Linguistics
A Practical Guide

VACLAV BREZINA
Lancaster University
To Anna, Olinka and Jan, who share my passion for numbers.
Contents

List of Figures .......................... page x
List of Tables .......................... xiv
About This Book ......................... xvii
Acknowledgements ..................... xix

1 Introduction: Statistics Meets Corpus Linguistics .......................... 1
  1.1 What Is This Chapter About? ..................................... 1
  1.2 What Is Statistics? Science, Corpus Linguistics and Statistics .......... 1
  1.3 Basic Statistical Terminology .................................. 5
  1.4 Building of Corpora and Research Design ......................... 15
  1.5 Exploring Data and Data Visualization .......................... 22
  1.6 Application and Further Examples: Do Fiction Writers Use More Adjectives than Academics? ................. 30
  1.7 Exercises ............................................... 32
  Things to Remember ..................... 36
  Advanced Reading ....................... 36

2 Vocabulary: Frequency, Dispersion and Diversity .......................... 38
  2.1 What Is This Chapter About? ..................................... 38
  2.2 Tokens, Types, Lemmas and Lexemes .......................... 38
  2.3 Words in a Frequency List ...................................... 42
  2.4 The Whelk Problem: Dispersion .................................. 46
  2.5 Which Words Are Important? Average Reduced Frequency .......... 53
  2.6 Lexical Diversity: Type/Token Ratio (TTR), STTR and MATTR .......... 57
  2.7 Application and Further Examples: Do the British Talk about Weather All the Time? ......................... 59
  2.8 Exercises ............................................... 62
  Things to Remember ..................... 64
  Advanced Reading ....................... 65

3 Semantics and Discourse: Collocations, Keywords and Reliability of Manual Coding .................................. 66
  3.1 What Is This Chapter About? ..................................... 66
  3.2 Collocations and Association Measures .......................... 66
  3.3 Collocation Graphs and Networks: Exploring Cross-associations ....... 75
  3.4 Keywords and Lockwords ...................................... 79
  3.5 Inter-rater Agreement Measures .................................. 87
# CONTENTS

3.6 Application and Further Examples: What Do Readers of British Newspapers Think about Immigration?  92  
3.7 Exercises  96  
Things to Remember  100  
Advanced Reading  101

4 Lexico-grammar: From Simple Counts to Complex Models  102  
4.1 What Is This Chapter About?  102  
4.2 Analysing a Lexico-grammatical Feature  103  
4.3 Cross-tabulation, Percentages and Chi-squared Test  108  
4.4 Logistic Regression  117  
4.5 Application: That or Which?  130  
4.6 Exercises  134  
Things to Remember  137  
Advanced Reading  138

5 Register Variation: Correlation, Clusters and Factors  139  
5.1 What Is This Chapter About?  139  
5.2 Relationships between Variables: Correlations  139  
5.3 Classification: Hierarchical Agglomerative Cluster Analysis  151  
5.4 Multidimensional Analysis (MD)  160  
5.5 Application: Registers in New Zealand English  170  
5.6 Exercises  177  
Things to Remember  181  
Advanced Reading  182

6 Sociolinguistics and Stylistics: Individual and Social Variation  183  
6.1 What Is This Chapter About?  183  
6.2 Individual Style and Social Variation: Where Does a Sociolinguistic Variable Start?  183  
6.3 Group Comparison: T-Test, ANOVA, Mann–Whitney U Test, Kruskal–Wallis Test  186  
6.4 Individual Style: Correspondence Analysis  199  
6.5 Linguistic Context: Mixed-Effects Models  207  
6.6 Application: Who Is This Person from the White House?  211  
6.7 Exercises  215  
Things to Remember  217  
Advanced Reading  218

7 Change over Time: Working Diachronic Data  219  
7.1 What Is This Chapter About?  219  
7.2 Time as a Variable: Measuring and Visualizing Time  219
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.3</td>
<td>Finding and Interpreting Differences: Percentage Change and the Bootstrap Test</td>
<td>229</td>
</tr>
<tr>
<td>7.4</td>
<td>Grouping Time Periods: Neighbouring Cluster Analysis</td>
<td>235</td>
</tr>
<tr>
<td>7.5</td>
<td>Modelling Changes in Discourse: Peaks and Troughs and UFA</td>
<td>241</td>
</tr>
<tr>
<td>7.6</td>
<td>Application: Colours in the Seventeenth Century</td>
<td>247</td>
</tr>
<tr>
<td>7.7</td>
<td>Exercises</td>
<td>251</td>
</tr>
<tr>
<td></td>
<td>Things to Remember</td>
<td>255</td>
</tr>
<tr>
<td></td>
<td>Advanced Reading</td>
<td>256</td>
</tr>
<tr>
<td>8</td>
<td>Bringing Everything Together: Ten Principles of Statistical Thinking, Meta-analysis and Effect Sizes</td>
<td>257</td>
</tr>
<tr>
<td>8.1</td>
<td>What Is This Chapter About?</td>
<td>257</td>
</tr>
<tr>
<td>8.2</td>
<td>Ten Principles of Statistical Thinking</td>
<td>257</td>
</tr>
<tr>
<td>8.3</td>
<td>Meta-analysis: Statistical Synthesis of Research Results</td>
<td>267</td>
</tr>
<tr>
<td>8.4</td>
<td>Effect Sizes: A Guide for Meaningful Use</td>
<td>275</td>
</tr>
<tr>
<td>8.5</td>
<td>Exercises</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>Things to Remember</td>
<td>282</td>
</tr>
<tr>
<td></td>
<td>Advanced Reading</td>
<td>282</td>
</tr>
<tr>
<td></td>
<td>Final Remarks</td>
<td>283</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>285</td>
</tr>
<tr>
<td></td>
<td>Index</td>
<td>294</td>
</tr>
</tbody>
</table>
Figures

1.1 The relationship between the relative frequency of adjectives and verbs
page 4
1.2 Process of statistical analysis
6
1.3 Example of a dataset
7
1.4 The distribution of the first-person pronoun in the Trinity
Lancaster Corpus
9
1.5 Standard normal distribution
9
1.6 Dispersion of adjective frequencies in 11 corpus files
11
1.7 Confidence intervals: two situations
14
1.8 Research designs in corpus linguistics
21
1.9 Bar chart: variable \( x \) in three corpora
24
1.10 Boxplot: variable \( x \) in three corpora
24
1.11 Error bars: variable \( x \) in three corpora
25
1.12 Histogram: the definite article in BE06
26
1.13 Histogram: the \( f \)-word in BNC64
26
1.14 Scatterplot: the and \( I \) in BNC64
27
1.15 Scatterplot: the, \( I \) and \( you \) in BNC64
28
1.16 Top ten places connected with ‘going’ or ‘travelling’ in the BNC
28
1.17 Other types of visualizations
29
1.18 The use of adjectives by fiction and academic writers:
boxplot
31
1.19 The use of adjectives by fiction and academic writers: error bars
32
1.20 Great Britain: main island
33
2.1 Distribution of word frequencies in the BNC
45
2.2 Example corpus: calculation of \( SD \)
49
2.3 Distribution of words \( w_1 \) and \( w_2 \)
55
3.1 Frequency and exclusivity scale
74
3.2 Collocation graph: ‘love’ in BE06 (10a – log Dice (7), L3–R3, C5–NC5)
76
3.3 Collocation networks: concept demonstration
77
3.4 Third-order collocates of time in LOB (3a–MI(5), R5–L5, C4–NC4; no filter applied)
78
3.5 Collocation network of ‘university’ based on BE06 (3b–MI(3), L5–R5, C8–NC8)
79
3.6 Collocation networks around ‘immigrants’ in the *Guardian* 
(3a–MI(6), R5–L5, C10–NC10; no filter applied) 94

3.7 Collocation networks around ‘immigrants’ in the *Daily Mail* 
(3a–MI(6), R5–L5, C20–NC20; no filter applied) 94

3.8 Selected collocation networks 97

4.1 The definite and indefinite articles in BNC subcorpora 104

4.2 The *vs a(n)* dataset: linguistic feature design (an excerpt) 105

4.3 A mosaic plot: article type by contextual determination 109

4.4 Logistic regression: a basic schema 119

4.5 Article use in English: a dataset (an excerpt) 122

4.6 A sentence from this book corrected for ‘grammar’ 130

4.7 Visualization of the relationship between which and that and 
a separator 132

4.8 *Must, have to and need to* in British English (BE06) 135

5.1 Nouns and adjectives in BE06 140

5.2 Verbs and adjectives in BE06 140

5.3 Pronouns and coordinators in BE06 141

5.4 Correlation: five data points 143

5.5 Correlation: covariance 143

5.6 Statistically significant (p <0.05) Pearson’s correlations 
in relation to the number of observations 145

5.7 Multi-panel scatterplot: nouns, adjectives, verbs, pronouns 
and coordinators 149

5.8 Correlation matrix: nouns, adjectives, verbs, pronouns and 
coordinators 150

5.9 Colour terms in the BNC 152

5.10 Creating clusters: Steps 1–4 155

5.11 Creating clusters: final result 156

5.12 Colour terms: a tree plot (dendrogram) – z-score$^2$ normalized, 
Euclidean distance, SLINK method 156

5.13 Tree plot: SLINK method 157

5.14 Tree plot: CLINK method 157

5.15 Tree plot: average linkage method 158

5.16 Tree plot: Ward’s method 159

5.17 A dataset for multidimensional analysis (a small extract) 164

5.18 Data reduction: ten variables into two factors 165

5.19 Promax factor rotation 166

5.20 Factor extraction: scree plot 167

5.21 Mean scores of registers placed on Dimension 1: Involved vs 
Informational 169

5.22 Correlation matrix: 44 variables 173

5.23 Correlation between mean word length and contractions: 
register clusters 174
LIST OF FIGURES

5.24 Cluster plot: registers in New Zealand English 175
5.25 Dimension 1: New Zealand English – full MD analysis 177
5.26 Dimension 2: New Zealand English – full MD analysis 177
5.27 Relationship between mean word length (number of characters) and mean sentence length (number of words) in BNC 178
5.28 Relationship between the use of the past and the present tense in BE06 178
5.29 Relationship between the use of adjectives and colour terms in BE06 179
5.30 Relationship between text length (tokens) and type–token ratio (TTR) in BNC 179
5.31 Dimension 3 181
5.32 Dimension 4 181
6.1 Distribution of personal pronouns in BNC64 female speakers 188
6.2 ANOVA calculation: between-group variance (top), within-group variance (bottom) 193
6.3 Dataset from BNC64 – relative frequencies and ranks: use of personal pronouns 195
6.4 Distribution of ain’t in BNC64 speakers: social-class effect 198
6.5 Ain’t in BNC64: 95% CI 198
6.6 A correspondence plot: word classes in the speech of individual speakers 201
6.7 Speaker (row) profiles: Euclidean distance 204
6.8 Speaker (row) profiles: chi-squared distance 205
6.9 Sociolinguistic dataset: internal and external factors (an excerpt) 208
6.10 Mixed-effects models: output 209
6.11 Correspondence analysis: use of word classes by White House press secretaries 214
6.12 Correspondence analysis: use of epistemic markers in BNC64 216
7.1 Modals in the Brown family corpora 220
7.2 Modals in the Brown family corpora: an alternative interpretation 223
7.3 Google n-gram viewer: ‘man’ and ‘woman’ 224
7.4 Modals in the Brown family corpora: original (top) and rescaled (bottom) 225
7.5 Modals in British English: (a) boxplots; (b) 95% CI error bars 227
7.6 Candlestick plot: the development of individual modals 1931–2006 228
7.7 Bootstrapping: demonstration of the concept 231
7.8 Example of a dataset for the bootstrap test: its in EEBO 233
7.9 Data points over time: an invented example 235
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.10</td>
<td>Two clustering principles: (a) hierarchical agglomerative clustering; (b) variability-based neighbour clustering</td>
<td>237</td>
</tr>
<tr>
<td>7.11</td>
<td>Dendrograms: (a) hierarchical agglomerative clustering; (b) variability-based neighbour clustering</td>
<td>238</td>
</tr>
<tr>
<td>7.12</td>
<td>Dendrogram: use of the possessive pronoun <em>its</em> in the seventeenth century</td>
<td>239</td>
</tr>
<tr>
<td>7.13</td>
<td>Scree plot: use of the possessive pronoun <em>its</em> in the seventeenth century</td>
<td>240</td>
</tr>
<tr>
<td>7.14</td>
<td>Resulting peaks and troughs graphs: settings as indicated</td>
<td>244</td>
</tr>
<tr>
<td>7.15</td>
<td>Results of UFA for <em>war</em> 1940–2009 (3a–MI(3), L5–R5, C10relative–NC10relative; AC1)</td>
<td>246</td>
</tr>
<tr>
<td>7.16</td>
<td>Frequency of colour terms in the seventeenth century</td>
<td>248</td>
</tr>
<tr>
<td>7.17</td>
<td>Candlestick plot: colours in the seventeenth century</td>
<td>249</td>
</tr>
<tr>
<td>7.18</td>
<td>Results of UFA for <em>red</em> 1600–99 (3a–MI(3), L5–R5, C10relative–NC10relative; AC1)</td>
<td>250</td>
</tr>
<tr>
<td>7.19</td>
<td>VNC: <em>red</em> in the seventeenth century</td>
<td>251</td>
</tr>
<tr>
<td>7.20</td>
<td>Number of tweets related to an episode of the UK <em>X-Factor</em> (16/11/2014, 7–11pm)</td>
<td>252</td>
</tr>
<tr>
<td>7.21</td>
<td>Development of frequencies of <em>handsome</em>, <em>pretty</em> and <em>beautiful</em> followed by a male (M) or female (F) person in the seventeenth century</td>
<td>252</td>
</tr>
<tr>
<td>7.22</td>
<td>Development of frequencies of the possessive pronoun <em>its</em> in the seventeenth century</td>
<td>253</td>
</tr>
<tr>
<td>7.23</td>
<td>Four frequency change scenarios</td>
<td>254</td>
</tr>
<tr>
<td>7.24</td>
<td><em>Handsome</em> in the seventeenth century</td>
<td>254</td>
</tr>
<tr>
<td>7.25</td>
<td><em>Pretty</em> in the seventeenth century</td>
<td>255</td>
</tr>
<tr>
<td>8.1</td>
<td>Overview of genres in BE06 (Baker 2009)</td>
<td>260</td>
</tr>
<tr>
<td>8.2</td>
<td>Past tense in different written genres of BE06</td>
<td>260</td>
</tr>
<tr>
<td>8.3</td>
<td>Past tense (a) and present tense (b) in different written genres of BE06: boxplot rendition</td>
<td>265</td>
</tr>
<tr>
<td>8.4</td>
<td>Finding the Globe</td>
<td>268</td>
</tr>
<tr>
<td>8.5</td>
<td>Forest plot: meta-analysis of four studies</td>
<td>274</td>
</tr>
<tr>
<td>8.6</td>
<td>Comparison of two subcorpora</td>
<td>278</td>
</tr>
<tr>
<td>8.7</td>
<td>Forest plot: example 1</td>
<td>281</td>
</tr>
<tr>
<td>8.8</td>
<td>Forest plot: example 2</td>
<td>281</td>
</tr>
</tbody>
</table>
# Tables

1.1 The effect size $r$ and its standard interpretation  
1.2 Brown family sampling frame  
1.3 Frequencies of selected words and expressions in three English corpora  
1.4 Different levels of analysis in corpus linguistics  
1.5 Subcorpora in mini-research  
2.1 Type, lemma and lexeme: advantages and disadvantages  
2.2 Top ten words in the BNC  
2.3 Example corpus: one million tokens  
2.4 Calculation of DP with the example corpus  
2.5 BE06  
2.6 Weather-related lemmas in BE06  
2.7 Ranks of weather-related lemmas in BE06  
2.8 BNC: distribution of four selected words  
3.1 Observed frequencies  
3.2 Expected frequencies: random occurrence baseline  
3.3 Association measures: overview  
3.4 Ranking of collocates of ‘new’ in BE06 (L3–R3)  
3.5 Collocation parameters notation (CPN)  
3.6 AmE06: American English keywords  
3.7 Decisions about keywords: BASIC options  
3.8 Comparison of selected lexical items in BE06 and AmE06  
3.9 American English keywords: different keyword identification procedures  
3.10 BE06: selected concordances for ‘religion’  
3.11 Double coding: concordances from the ‘Think about’ task  
3.12 Overview of inter-rater agreement measures  
3.13 Keywords  
3.14 Evaluations of ‘immigrant(s)’ in the GU and DM corpora  
3.15 Collocates of issue in BE06  
3.16 Keywords  
3.17 Examples for rating  
4.1 Examples of lexico-grammatical variables with a grammatical frame  
4.2 Cross-tabulation: article type by contextual determination  
4.3 Percentage options in cross-tabulation
List of Tables

4.4 Strong (semi-)modals in different genres of British and American English: cross-tabulation 112
4.5 Expected frequencies: article type by contextual determination 114
4.6 Interpretation of Cramer’s $V$ 115
4.7 Probabilities: article type by contextual determination 116
4.8 Probabilities: article type by noun type 121
4.9 Models: an overview 124
4.10 A part of the logistic regression output: large standard errors 124
4.11 Cross-tabulation: separator use with *which* and *that* relativizers 131
4.12 *Which* and *that* in different contextual situations: cross-tabulation 133
4.13 *Which* or *that*: logistic regression estimates 134
4.14 Appropriate research design 134
5.1 Ranks of nouns and adjectives in five texts from BE06 147
5.2 Correlation table (Pearson’s correlations): nouns, adjectives, verbs, pronouns and coordinators 148
5.3 The full set of Biber’s (1988) features based on Conrad & Biber (2001: 18–19) 162
5.4 Factor 1: loadings of individual variables 168
5.5 Registers in ICE-NZ 171
5.6 Results of factor analysis of NZ English: factor loadings 176
5.7 Results of factor analysis of NZ English: factor loadings of Factors 3 and 4 180
6.1 Cross-tabulation table: word classes in the speech of individual speakers 201
6.2 Cross-tabulation table: verbs and articles in the speech of two speakers 202
6.3 Row (speaker) profiles 203
6.4 WH corpus 213
6.5 Swearing and gender: BNC64 217
7.1 Comparison of two periods in the EEBO corpus: Commonwealth & Protectorate and Restoration 230
7.2 Comparison of two periods in the EEBO corpus: results of the bootstrap test 233
7.3 Final evaluation of the results: *its*, *must* and *pestilence* 234
7.4 Relative frequency (per million) of the possessive pronoun *its* in the seventeenth century 239
7.5 Relative frequencies (per million) of *war* 243
7.6 Differences between relative frequencies of *war* 243
7.7 Log transformed relative frequency (per million) of *war* 243
7.8 Collocate profiles of *war* 245
8.1 The use of the past and the present tense in different registers (original dataset) 258
<table>
<thead>
<tr>
<th></th>
<th>LIST OF TABLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.2</td>
<td>The use of the past and the present tense in different registers (research report)</td>
</tr>
<tr>
<td>8.3</td>
<td>Examples of use of the past tense in academic writing and mystery fiction</td>
</tr>
<tr>
<td>8.4</td>
<td>Examples of texts: academic writing and mystery fiction</td>
</tr>
<tr>
<td>8.5</td>
<td>Overview of answers</td>
</tr>
<tr>
<td>8.6</td>
<td>Studies reviewed for the meta-analysis</td>
</tr>
<tr>
<td>8.7</td>
<td>Input data for a simple meta-analysis</td>
</tr>
<tr>
<td>8.8</td>
<td>Effect size measures introduced in this book</td>
</tr>
<tr>
<td>8.9</td>
<td>Effect size transformation and extrapolation</td>
</tr>
<tr>
<td>8.10</td>
<td>Effect size: standard interpretation</td>
</tr>
<tr>
<td>8.11</td>
<td>Effect size measures: BNC validation</td>
</tr>
<tr>
<td>8.12</td>
<td>Effect size transformations</td>
</tr>
</tbody>
</table>
About This Book

What Is This Book About?

This book is a practical introduction to statistical procedures in corpus linguistics, a discipline that uses computers to analyse language, organized according to linguistic topics. These range from vocabulary and grammar to sociolinguistics, discourse analysis and historical investigations of language. The book offers an overview of the state-of-the-art methodologies of language analysis using corpora and introduces new techniques that have not previously been used in the field. No prior knowledge of statistics is assumed; instead, all necessary concepts and methods are explained in non-technical language. In addition, all procedures described in the book can be easily carried out using Lancaster Stats Tools online (see ‘How Should You Use This Book?’ below). Throughout the book, many examples (case studies) of the application of corpus statistics are provided and standard reporting of statistics is shown. The emphasis of the book on the practical aspects of statistical analysis of language is also reflected in its focus on research design and the implications of different ‘shapes’ of data for statistical analysis – for this reason, the companion website offers complete datasets used in this book for easy replication of the analyses. Corpus linguistics is an extremely versatile methodology of language analysis applicable in a wide range of contexts, in linguistics, social science, digital humanities and elsewhere – the book thus aims to facilitate meaningful use of corpora for as wide a range of users as possible.

Who Is This Book For?

The book is intended for anyone interested in corpus linguistics and quantitative analysis of language. This includes students and researchers in the field of linguistics, sociology, history, psychology, education etc. The main goal of the book is to help readers understand key principles of statistical thinking in order to be able to make informed decisions about the applications of particular statistical techniques. To facilitate this, in addition to the expository parts, the book also includes discussion questions (‘Think about…’) and exercises which the readers can use to better engage with the material and to check their comprehension of the subject matter; answers to the exercises are provided at the companion website (http://corpora.lancs.ac.uk/stats/materials.php).
How Should You Use This Book?

The book reflects the needs of students and researchers who are looking for a practical guidebook on corpus statistics, which is grounded in the current literature in the field and reflects the best practice. The book can be used as a course book or for independent study. After reviewing general statistical principles in Chapter 1, readers can follow their own path through the book according to the linguistic topics of their interest. Statistical techniques introduced in the book are cross-referenced and included in the Index at the end of the book.

The book comes with a companion website – Lancaster Stats Tools online – (http://corpora.lancs.ac.uk/stats), which not only provides additional examples, datasets, video tutorials and PowerPoint slides but also, more importantly, includes easy-to-use tools for calculating statistics and producing graphs discussed in the book. In fact, all procedures described in the book can be performed by the reader using Lancaster Stats Tools online. Readers thus don’t need to rely on commercial statistical packages such as *IBM SPSS* which are not easily affordable for users without institutional subscriptions. Neither will readers be required to learn the complex syntax of free statistical packages such as *R*. Instead, Lancaster Stats Tools online offers access to powerful statistical tools through a simple user interface, into which the data can be directly copied-pasted from a spreadsheet (e.g. Excel or Calc). I believe that statistics shouldn’t be a hurdle in our research – computers can and should do all the hard work of number crunching for us; statistics, instead, can be used as a very effective analytical tool – all that is needed is to understand the basic principles of statistical thinking and their application to language analysis. Let’s explore them together!
I would like to thank Tony McEnery and Dana Gablasova for their continued support and encouragement throughout the process of writing this book as well as for their detailed comments on each of the chapters. The work has also greatly benefited from the insightful points raised by Gabriele Pallotti who has read a large part of the manuscript. In addition, I would also like to thank the following colleagues for their helpful comments on different parts of the manuscript: Peter Diggle, Michael Gauthier, Andrew Hardie and Irene Marin Cervantes. I thank Gill Smith for her help with the formatting of the manuscript and Irene Marin Cervantes for preparing index entries. Thanks are also due to two anonymous reviewers for their very helpful and encouraging comments.

The writing of this book was supported by ESRC grants nos. EP/P001559/1 and ES/K002155/1.