Spatial Analysis for the Social Sciences

Many theories in the social sciences predict spatial dependence or the similarity of behaviors at neighboring locations. *Spatial Analysis for the Social Sciences* demonstrates how researchers can diagnose and model this spatial dependence and draw more valid inferences as a result. The book is structured around the well-known Galton's problem and presents a step-by-step guide to the application of spatial analysis. The book examines a variety of spatial diagnostics and models through a series of applied examples drawn from the social sciences. These include spatial lag models that capture behavioral diffusion between actors, spatial error models that account for spatial dependence in errors, and models that incorporate spatial heterogeneity in the effects of covariates. *Spatial Analysis for the Social Sciences* also examines advanced spatial models for time-series cross-sectional data, categorical and limited dependent variables, count data, and survival data.

David Darmofal is an associate professor of political science at the University of South Carolina. His research focuses on spatial analysis and political geography and has appeared in a variety of journals including the *American Journal of Political Science*, *Journal of Politics*, and *Political Geography*. He has received best article awards from the *Journal of Politics* and *Political Research Quarterly*. He teaches regularly in the Inter-University Consortium for Political and Social Research (ICPSR) Summer Program in Quantitative Methods of Social Research.

Analytical Methods for Social Research

Analytical Methods for Social Research presents texts on empirical and formal methods for the social sciences. Volumes in the series address both the theoretical underpinnings of analytical techniques as well as their application in social research. Some series volumes are broad in scope, cutting across a number of disciplines. Others focus mainly on methodological applications within specific fields such as political science, sociology, demography, and public health. The series serves a mix of students and researchers in the social sciences and statistics.

Series Editors:

R. Michael Alvarez *California Institute of Technology*
Nathaniel L. Beck *New York University*
Stephen L. Morgan *Johns Hopkins University*
Lawrence L. Wu *New York University*

Other Titles in the Series:

Time Series Analysis for the Social Sciences, by Janet M. Box-Steffensmeier, John R. Freeman, Matthew P. Hitt, and Jon C. W. Pevehouse
Counterfactuals and Causal Inference, *Second Edition*, by Stephen L. Morgan and Christopher Winship
Statistical Modeling and Inference for Social Science, by Sean Gailmard
Formal Models of Domestic Politics, by Scott Gehlbach
Counterfactuals and Causal Inference: Methods and Principles for Social Research, by Stephen L. Morgan and Christopher Winship
Data Analysis Using Regression and Multilevel/Hierarchical Models, by Andrew Gelman and Jennifer Hill
Political Game Theory: An Introduction, by Nolan McCarty and Adam Meirowitz
Essential Mathematics for Political and Social Research, by Jeff Gill
Spatial Models of Parliamentary Voting, by Keith T. Poole
Ecological Inference: New Methodological Strategies, edited by Gary King, Ori Rosen, and Martin A. Tanner
Event History Modeling: A Guide for Social Scientists, by Janet M. Box-Steffensmeier and Bradford S. Jones

Spatial Analysis for the Social Sciences

DAVID DARMOFAL
University of South Carolina

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

32 Avenue of the Americas, New York, NY 10013-2473, USA

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9780521716383

First published 2015

Printed in the United States of America by Sheridan Books, Inc.

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication data
Darmofal, David, 1968–
Spatial analysis for the social sciences / David Darmofal.
pages cm. – (Analytical methods for social research)
Includes bibliographical references and index.
ISBN 978-0-521-88826-4 (hardback) – ISBN 978-0-521-71638-3 (paperback)
1. Spatial analysis (Statistics) 2. Social sciences–Research–Methodology. I. Title.
HA30.6.D37 2015
001.4′22–dc23 2015016112

ISBN 978-0-521-88826-4 Hardback
ISBN 978-0-521-71638-3 Paperback

Contents

List of Figures *page* xi
List of Tables xii
Preface xv

PART I GENERAL TOPICS 1

1 The Social Sciences and Spatial Analysis 3
1.1 *Introduction* 3
1.2 *Spatial Lag and Spatial Error Models* 5
1.3 *Outline of the Book* 6
1.4 *For Further Reading* 8

2 Defining Neighbors via a Spatial Weights Matrix 10
2.1 *The Importance of Space in the Social Sciences* 10
2.2 *Types of Spatial Data* 11
2.3 *Defining Neighbors in Areal Data* 14
2.4 *Misspecification of the Weights Matrix* 23
2.5 *Estimating a Spatial Weights Matrix* 23
2.6 *Application: Demographic Change* 24
2.7 *Additional Topics* 26
2.8 *Conclusion* 28

3 Spatial Autocorrelation and Statistical Inference 30
3.1 *Defining Spatial Autocorrelation* 31
3.2 *Spatial versus Temporal Autocorrelation* 32
3.3 *Monte Carlo Analysis* 34

	3.4	*Additional Topics*	37
	3.5	*Conclusion*	41
4	Diagnosing Spatial Dependence		43
	4.1	*Global Measures of Spatial Autocorrelation*	43
	4.2	*Local Measures of Spatial Autocorrelation*	52
	4.3	*Spatial Heterogeneity*	63
	4.4	*Additional Topics*	65
	4.5	*Conclusion*	67
5	Diagnosing Spatial Dependence in the Presence of Covariates		68
	5.1	*Unfocused Diagnostics for Spatial Dependence in OLS Regressions*	70
	5.2	*Focused Diagnostics for Spatial Dependence in OLS Regressions*	77
	5.3	*Decision Rule for Standard and Robust LM Diagnostics*	87
	5.4	*Application: Diagnosing Spatial Dependence in Poverty Rates Using LM Diagnostics*	88
	5.5	*Application: Diagnosing Spatial Dependence in Roll-Call Voting Using LM Diagnostics*	91
	5.6	*Application: Diagnosing Spatial Dependence in Turnout Using LM Diagnostics*	91
	5.7	*Conclusion*	95
6	Spatial Lag and Spatial Error Models		96
	6.1	*Maximum Likelihood Spatial Lag Estimation*	97
	6.2	*ML Spatial Error Estimation*	102
	6.3	*Estimation for Large Numbers of Observations*	106
	6.4	*Interpreting Substantive Effects in the Spatial Lag Model*	107
	6.5	*Application: State Spending on Higher Education*	108
	6.6	*Goodness of Fit Statistics*	111
	6.7	*Additional Topics*	114
	6.8	*Conclusion*	117
7	Spatial Heterogeneity		119
	7.1	*Spatial Heterogeneity in Parameters*	119
	7.2	*Spatial Random Coefficients Models*	119
	7.3	*Spatial Switching Regressions*	120
	7.4	*Application: Spatial Switching Regression Model of Spatial Heterogeneity in the Sources of Voting Behavior during the New Deal Realignment*	122

7.5 *Spatial Expansion Models* 128
7.6 *Geographically Weighted Regression Models* 131
7.7 *Application: GWR Model of Spatial Heterogeneity in Voting during the New Deal Realignment* 132
7.8 *Conclusion* 137

PART II ADVANCED TOPICS 139

8 Time-Series Cross-Sectional and Panel Data Models 141
8.1 *Space–Time Models* 142
8.2 *Fixed Effects Spatial Models* 143
8.3 *Random Effects Spatial Models* 144
8.4 *Spatial Error Component Model with Spatial and Temporal Autocorrelation* 146
8.5 *TSCS Spatial Lag Model with a Temporal Lag* 147
8.6 *Appropriateness of Random Effects Models for Spatial Areal Data* 148
8.7 *Spatial Hausman Test* 148
8.8 *Nonparametric Covariance Matrix Estimation for Space–Time Models* 150
8.9 *Lagrange Multiplier Tests for Space–Time Models* 151
8.10 *Application: Government Ideology and Representation* 154
8.11 *Additional Topics* 156
8.12 *Conclusion* 157

9 Advanced Spatial Models 158
9.1 *Spatial Binary Dependent Variable Models* 158
9.2 *Diagnostics for Spatial Dependence in Binary Dependent Variable Models* 169
9.3 *Application: Spatial Probit Model of Immigrant Demographics* 170
9.4 *Spatial Multinomial Models* 171
9.5 *Spatial Count Models* 176
9.6 *Spatial Survival Models* 178
9.7 *Application: Spatial Survival Models of NAFTA Position Announcements* 180
9.8 *Conclusion* 198

10 Conclusion 200
10.1 *Theoretical Development on Geographic Influences in the Social Sciences* 201

10.2 *Development of Spatial Methods for Large Samples in the Social Sciences* 202
10.3 *Use of Diagnostics for Spatial Models* 203
10.4 *Conclusion* 203

PART III APPENDICES ON IMPLEMENTING SPATIAL ANALYSES 205
Appendix A Getting Data Ready for a Spatial Analysis 207
Appendix B Spatial Software 209
Appendix C Web Resources for Spatial Analysis 215

Glossary 219
Bibliography 223
Index 241

List of Figures

2.1 Irregular areal data – Illinois block groups in the 2000 census. *page* 12
2.2 John Snow's map of the 1854 London cholera epidemic. 13
2.3 Rook neighbor definition. 16
2.4 Bishop neighbor definition. 16
2.5 Queen neighbor definition. 17
2.6 Percentage change in the Hispanic population, 2000–2010. 25
3.1 Bias of the OLS estimator with omitted spatially lagged dependent variable. 36
4.1 Internal conflict, 1991–2004. 46
4.2 Moran scatterplot – county-level margin of victory in the 1984 presidential election. 57
4.3 Local spatial autocorrelation – county-level margin of victory in the 1984 presidential election. 58
4.4 Local spatial autocorrelation – county-level margin of victory in the 2000 presidential election. 59
4.5 Seating chart – 26th Congress, 1839–1841. 62
4.6 Significant local Moran's *I*'s for roll-call voting, 26th Congress, 1839–1841. 64
7.1 GWR coefficients for model of changes in the Democratic vote, 1928–1932. 136
9.1 Nonspatial Cox state-level frailties for model of the timing of NAFTA position announcements. 195
9.2 Spatial Cox state-level frailties for model of the timing of NAFTA position announcements. 195

List of Tables

3.1 Bias of the OLS Estimator with Omitted Spatially Lagged Dependent Variable *page* 37
3.2 Ratio of OLS Standard Error to True Standard Error with Spatial Error Dependence 37
4.1 Join Count Results for Civil Wars, 1991–2004 48
4.2 Global Moran's *I* Results for Census Tract-Level Poverty Rates in the United States 52
4.3 Global Moran's *I* Results for County-Level Margins of Victory in the 1984 and 2000 Presidential Elections 56
5.1 Global Moran's *I* Results for California Poverty Rates 89
5.2 OLS Estimates for California Poverty Rates 90
5.3 Spatial Diagnostics, California Poverty Rates 90
5.4 OLS Estimates for Roll-Call Voting, Side-by-Side Analysis 92
5.5 Spatial Diagnostics for Roll-Call Voting, Side-by-Side Analysis 92
5.6 OLS Estimates for Turnout in the 1828 Presidential Election 94
5.7 Diagnostics for Spatial Dependence in OLS Model of Turnout in the 1828 Presidential Election 94
6.1 OLS and ML Spatial Lag Estimates for Roll-Call Voting, Side-by-Side Analysis 101
6.2 OLS and ML Estimates for California Poverty Rates 104
6.3 OLS and ML Spatial Error Estimates for Turnout in the 1828 Presidential Election 105
6.4 Estimates for Spatial Lag Model of Higher Education Spending 109
6.5 Substantive Effects from Spatial Lag Model of Higher Education Spending 110
6.6 Direct, Indirect, and Total Effects of Covariates in Spatial Lag Model of Higher Education Spending 111
7.1 OLS Estimates for Change in Democratic Vote, 1928–1932 125

7.2 Diagnostics for Spatial Dependence in OLS Model of Change in Democratic Vote, 1928–1932 125
7.3 OLS and IV Spatial Lag Estimates for Change in Democratic Vote, 1928–1932 126
7.4 GWR Estimates for Change in the Democratic Vote, 1928–1932 135
8.1 ML Spatial Lag Estimates for State Government Ideology, 1961–2010 156
9.1 ML Probit and GMM Spatial Linearized Probit Estimates for Models of Immigrant Demographics 171
9.2 Model Choice Statistics for Models of the Timing of NAFTA Position Announcements 192
9.3 Posterior Summaries for Cox Models of the Timing of NAFTA Position Announcements 194
9.4 Posterior Summaries for Weibull Models of the Timing of NAFTA Position Announcements 197

Preface

This book is designed for social scientists who seek to model the spatial dimension of behavior that is inherent in substantive social science theories. The social sciences are united by their focus on subjects that are inherently social and interdependent. Unlike more individuated concerns, the units studied by social scientists often interact with each other and affect each other's behavior as a consequence. This interaction is promoted by spatial proximity – units that are more geographically proximate are more likely to interact with each other and influence each other as a consequence. Shared concerns combine with spatial proximity to promote familiarity. Just as familiarity has long been observed to breed contempt, so also does it promote cooperation. Even when spatially proximate units do not interact with each other, they often exhibit similar behaviors as a result of shared environmental influences. In short, there is a strong spatial component to many of the behaviors studied by social scientists and this is reflected in many of our theories in the social sciences.

It might surprise social scientists who have not seriously considered the spatial dimension of behavior to learn that all social science data are, in fact, spatial data. The behaviors, processes, and events of interest to social scientists occur at specific geographic locations. The past two decades have seen an explosion in social science data that are geocoded – coded to include the geographic locations of the observations. This increase in the availability of geocoded data has been matched by advances in geographic information system (GIS) software, such as ESRI's (Environmental Systems Research Institute) popular ArcGIS package. At the same time, significant advancements have been made in the development of spatial diagnostics and estimators, many of which are now included in dedicated spatial software such as GeoDa as well as standard statistical packages such as R, Stata, and WinBUGS. The time has never been better for scholars wishing to model spatial relationships in their data.

This book is organized around Galton's problem, a common reference point across many of the social sciences. In the late nineteenth century, well before

modern spatial diagnostics and estimators had even been imagined, Galton raised an important concern that would shape subsequent research in the social sciences. Galton noted that behaviors should not be assumed to be independent across units of observation. Instead, units may share similar behaviors as a result of behavioral diffusion. Alternatively, units may exhibit similarity in their behaviors as a result of shared exposure to common factors that influence behavior. Galton's problem and the two alternative sources of spatial similarity (and dissimilarity) in behaviors serve as organizing principles for the book's discussion of how social scientists can diagnose and model spatial dependence in their data.

The book is designed to guide researchers through the sequential steps of diagnosing and modeling spatial dependence in their data. The book begins by introducing social scientists to the principal forms of spatial data. Next, the critical question of how "neighboring" observations are defined is discussed. The book then examines the principal differences between spatial dependence and the more familiar temporal dependence in time series analysis. The book next presents diagnostics that researchers can employ to diagnose spatial dependence in their data. The book discusses alternative approaches to modeling this dependence. The latter chapters are devoted to specialized, advanced models for spatial dependence. The book concludes by examining software and web resources for diagnosing and modeling spatial dependence.

The methods discussed in this book are demonstrated through a variety of examples. Among these are applications to demographic change, poverty rates, immigrant demographics, civil wars, partisan voting, legislative roll-call voting behavior, voter turnout, state spending on higher education, the New Deal realignment, government ideology and representation, and the timing of position announcements on the North American Free Trade Agreement (NAFTA). These applications employ many different types of areal units, including census tracts, counties, congressional districts, states, and countries. A variety of spatial methods are demonstrated through these applications. These spatial topics include the diagnosis of spatial dependence at the global and local levels in the absence of covariates, spatial lag and spatial error models, models for spatial time-series cross-sectional (TSCS) data, geographically weighted regressions (GWR), spatial binary dependent variable models, and spatial survival models. The data and code for the applications in this book are available at http://thedata.harvard.edu/dvn/dv/David.

As with any project, this book has benefited from countless conversations. My work in spatial analysis began in graduate school. I was fortunate to be a graduate student in the Political Science Department at the University of Illinois at Urbana-Champaign, an engaging and supportive environment for graduate students. At the time I was working on a dissertation employing county-level data that presented some unique challenges – and opportunities – for analysis. One of my advisors, Wendy Tam Cho, was herself beginning to explore the effects of spatial dependence on political behavior and suggested that I also

explore the leverage that spatial analysis could provide for my research. As it turned out, Illinois was a fortuitous place to be, as one of the leading scholars in spatial econometrics, Luc Anselin, had recently joined the faculty. Like Wendy, Luc also became a valuable guide for my burgeoning interest in spatial analysis. I thank Wendy and Luc for their encouragement and support for me as a young scholar.

I also thank the series editors, R. Michael Alvarez, Neal Beck, Stephen Morgan, and Lawrence Wu, and my editor at Cambridge University Press, Robert Dreesen, for their guidance and support on this project. I thank the University of South Carolina for its support during the writing of this book and for providing an active and engaging academic environment in which to write it. Thanks also to Janet Box-Steffensmeier and the Political Science Department at Ohio State University, who provided me the opportunity to explore my thinking on spatial analysis further and to plan and teach my own graduate course on spatial analysis during a year as a postdoctoral fellow in the Program in Statistics and Methodology (PRISM) at Ohio State.

I also want to thank the many other scholars whose conversations, advice, and discussions over the years have shaped my understanding of spatial analysis. In addition to those already mentioned, these scholars include Brady Baybeck, Kyle Beardsley, Fred Boehmke, Jake Bowers, Sarah Brooks, Greg Caldeira, Mike Crespin, Chuck Finocchiaro, Rich Fording, Rob Franzese, Brian Gaines, Andrew Gelman, Elisabeth Gerber, Jeff Gill, Jim Gimpel, Kristian Gleditsch, Michael Greig, Jude Hays, Zaryab Iqbal, Jennifer Jerit, Aya Kachi, Luke Keele, Holger Kern, James Kuklinski, Marcus Kurtz, Andrew Lawson, Suzanna Linn, Arthur Lupia, Stephen A. Matthews, Scott McClurg, Daniel McMillen, James Monogan, Jacob Montgomery, Clayton Nall, Peter Nardulli, Brendan Nyhan, John Patty, Mark Peffley, Maggie Penn, Thomas Plümper, Kirk Randazzo, Jon Rogowski, Norman Schofield, Betsy Sinclair, Anand Sokhey, Sarah Wilson Sokhey, Lyndsey Young Stanfill, Harvey Starr, Tracy Sulkin, Rocio Titiunik, Craig Volden, Lee Walker, Mike Ward, Justin Wedeking, Alan Wiseman, Chris Witko, Jack Wright, and Chris Zorn. I apologize to any scholars who were inadvertently left off this list, as I have benefited from numerous conversations with scholars on spatial analysis over the years. I also thank seminar participants at the University of Illinois at Urbana–Champaign, the Quantitative Institute for Social Science at the University of Kentucky, the Center for Political Studies at the University of Michigan, the Methods Speaker Series at Ohio State University, the Biostatistics Forum in the Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, the Political Science Research Workshop at the University of South Carolina, the Center for New Institutional Social Sciences at Washington University in St. Louis, and faculty and students in the Interactive Television (ITV) Program in Advanced Political Methodology at Ohio State University, the University of Illinois, the University of Minnesota, and the University of Wisconsin for feedback on spatial presentations. I thank Lynn Shirley at the

University of South Carolina for help in identifying and accessing spatial data and Michael Fix and Adam Pernsteiner for helpful research assistance on this project. I also thank SAGE Publications for the permission to reprint portions of my *American Politics Research* article, "The Political Geography of the New Deal Realignment" in Chapter 7 and Blackwell Publishing for the permission to reprint portions of my *American Journal of Political Science* article, "Bayesian Spatial Survival Models for Political Event Processes" in Chapter 9.

Finally, I want to thank my mother. The past few years have been a rewarding time, as roles have reversed and I've been able to repay her for the support and encouragement she has provided throughout my life. These years have also reinforced the importance of spatial proximity for all of our relationships. This book is dedicated to her.