

Regression Inside Out

Linear regression analysis, with its many generalizations, is the predominant quantitative method used throughout the social sciences and beyond. The goal of the method is to study relations among variables. In this book, Schoon, Melamed, and Breiger turn regression modeling inside out to put the emphasis on the cases (people, organizations, and nations) that comprise the variables. By reanalyzing influential published research, they reveal new insights and present a principled way to unlock a set of more nuanced interpretations than has previously been attainable. The emphasis is on intuition and examples that can be reproduced using the code and datasets provided. Relating their contributions to methodologies that operate under quite different philosophical assumptions, the authors advance multimethod social science and help to bridge the divide between quantitative and qualitative research. The result is a modern, accessible, and innovative take on extracting knowledge from data.

Eric W. Schoon is Associate Professor of Sociology at The Ohio State University. His research interests include case-oriented and relational methods, sociological theory, and cultural dimensions of contentious politics. His work has appeared in outlets including *American Sociological Review*, *Journal of Politics*, *Social Forces*, and *Social Problems*.

David Melamed is Professor of Sociology and Translational Data Analytics at The Ohio State University. He is currently Co-Editor of *Sociological Methodology*. His research interests include the emergence of stratification and cooperation in complex systems. His work has appeared in *American Journal of Sociology*, *American Sociological Review*, and interdisciplinary venues.

Ronald L. Breiger is Regents Professor of Sociology at the University of Arizona. His interests include social network theory and methods and measurement issues in cultural and institutional analysis. He is the recipient of distinguished career awards from (respectively) the Methodology and Mathematical Sociology Sections of the American Sociological Association.

Strategies for Social Inquiry

Editors

Colin Elman, *Maxwell School of Syracuse University*

John Gerring, *University of Texas, Austin*

James Mahoney, *Northwestern University*

Editorial Board

Bear Braumoeller[†], David Collier, Francesco Guala, Peter Hedström, Theodore Hopf, Uskali Maki, Rose McDermott, Charles Ragin, Theda Skocpol, Peter Spiegler, David Waldner, Lisa Wedeen, Christopher Winship

This book series presents texts on a wide range of issues bearing upon the practice of social inquiry. Strategies are construed broadly to embrace the full spectrum of approaches to analysis, as well as relevant issues in philosophy of social science.

Published Titles

John Gerring and Jason Seawright, *Finding Your Social Science Project: The Research Sandbox*

Jennifer Widner, Michael Woolcock and Daniel Ortega Nieto (eds.), *The Case for Case Studies: Methods and Applications in International Development*

Colin Elman, John Gerring and James Mahoney (eds.), *The Production of Knowledge: Enhancing Progress in Social Science*

John Boswell, Jack Corbett and R. A. W. Rhodes, *The Art and Craft of Comparison*

John Gerring, *Case Study Research: Principles and Practices*, 2nd edition

Jason Seawright, *Multi-Method Social Science: Combining Qualitative and Quantitative Tools*

Peter Spiegler, *Behind the Model: A Constructive Critique of Economic Modeling*

James Mahoney and Kathleen Thelen (eds.), *Advances in Comparative-Historical Analysis*

Diana Kapiszewski, Lauren M. MacLean and Benjamin L. Read, *Field Research in Political Science: Practices and Principles*

Andrew Bennett and Jeffrey T. Checkel (eds.), *Process Tracing: From Metaphor to Analytic Tool*

Nicholas Weller and Jeb Barnes, *Finding Pathways: Mixed-Method Research for Studying Causal Mechanisms*

Thad Dunning, *Natural Experiments in the Social Sciences: A Design-Based Approach*

Carsten Q. Schneider and Claudius Wagemann, *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*

Michael Coppedge, *Democratization and Research Methods*

John Gerring, *Social Science Methodology: A Unified Framework*, 2nd edition

Regression Inside Out

Eric W. Schoon

The Ohio State University

David Melamed

The Ohio State University

Ronald L. Breiger

University of Arizona



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press & Assessment
978-1-108-84110-8 — Regression Inside Out
Eric W. Schoon, David Melamed, Ronald L. Breiger
Frontmatter
[More Information](#)



CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781108841108

DOI: 10.1017/9781108887205

© Eric W. Schoon, David Melamed and Ronald L. Breiger 2024

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press & Assessment.

First published 2024

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication Data

Names: Schoon, Eric W., author. | Melamed, David, author. | Breiger, Ronald L., author.

Title: Regression inside out / Eric W. Schoon, The Ohio State University, David Melamed, The Ohio State University, Ronald L. Breiger, University of Arizona.

Description: Cambridge, United Kingdom ; New York, NY, USA : Cambridge University Press, 2024. | Series: Strategies for social inquiry |

Includes bibliographical references and index.

Identifiers: LCCN 2023035626 | ISBN 9781108841108 (hardback) |

ISBN 9781108887205 (ebook)

Subjects: LCSH: Regression analysis.

Classification: LCC QA278.2 .S35 2024 | DDC 519.5/36–dc23/eng/20231103

LC record available at <https://lccn.loc.gov/2023035626>

ISBN 978-1-108-84110-8 Hardback

ISBN 978-1-108-74488-1 Paperback

Cambridge University Press & Assessment has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

	<i>List of Figures</i>	<i>page</i> ix
	<i>List of Tables</i>	xi
	<i>Acknowledgments</i>	xiii
1	Introduction	1
	1.1 A Case-Oriented Approach to Regression Models	3
	1.2 A Methodological Gateway	7
	1.3 Understanding versus Improving Models	9
	1.4 Plan of the Book	10
	1.5 Reproducibility and Replication Materials	14
<hr/>		
Part I	Turning Regression Inside Out	
2	OLS Inside Out	17
	2.1 Regression by the Book	18
	2.2 Turning OLS Inside Out	21
	2.2.1 Simple Regression	21
	2.2.2 Multiple Regression Inside Out	25
	2.3 Geometry of the Regression Model	30
	2.3.1 Visualizing the Model Space	33
	2.3.2 Clusters of Cases	36
	2.4 Regression Coefficients and Fitted Values inside a Field of Cases and Variables	37
	2.4.1 Multiple Regression Coefficients	39
	2.4.2 Fitted Values	41
	2.4.3 Positionings of Cases and Variables	43
	2.5 A Matrix of Profile Similarity	46
	2.6 Summary	48

vi	Contents	
3	Generalizing Regression Inside Out	50
	3.1 Generalizing RIO	50
	3.2 Logistic Regression	52
	3.3 Poisson and Negative Binomial Models	55
	3.4 Random Intercept Mixed Model	59
	3.5 Fixed Effects Models with a Continuous Outcome	60
	3.6 Summary	61
4	Turning Variance Inside Out (Coauthor: Eunsung Yoon)	62
	4.1 Inference and Uncertainty	62
	4.2 Challenges in Calculating Case Contributions to Standard Errors	64
	4.3 The Squared Residual Approach	66
	4.4 The Leave-One-Out Approach	69
	4.5 Choosing between the Squared Residuals and Leave-One-Out	76
	4.6 Summary	77

Part II Action and Interaction

5	Action Detection	81
	5.1 Regression as Storytelling	83
	5.2 Example 1: Inequality and Regional Integration	85
	5.3 Example 2: Poverty, Inequality, and Life Expectancy	96
	5.4 Example 3: Dog Ownership in the United States	105
	5.5 Summary	111
6	Interaction Detection	113
	6.1 Decomposing by Predictor States or Clusters	114
	6.2 Regressing on Rows of P	120
	6.3 Relation to Existing Approaches to Interaction Detection	122
	6.4 Summary	123

Part III RIO as a Gateway

7	RIO as a Gateway to Case Selection	127
	7.1 Case Studies and Case Selection	129
	7.1.1 Case Selection Strategies	129

	7.2 Regression Inside Out as a Gateway to Case Selection	134
	7.3 The Effects of Welfare Policies on Poverty	135
	7.3.1 Algorithmic Case Selection	136
	7.3.2 Turning the Regression Model Inside Out	141
	7.3.3 Italy	144
	7.3.4 Ireland	151
	7.3.5 Netherlands	155
	7.3.6 France and the United States	156
	7.3.7 Australia and Canada	157
	7.3.8 Beyond Most-Similar: Differences and Diversity	159
	7.4 Determinants of Rebellion by Excluded Groups	160
	7.4.1 Turning the Regression Model Inside Out	163
	7.4.2 Deviant Cases	163
	7.4.3 Most-Similar	168
	7.5 Summary	171
8	RIO as a Gateway to Configurational Comparative Analysis	172
	8.1 Comparing QCA and Regression	173
	8.2 RIO as a Gateway to Configurational Comparative Methods	176
	8.2.1 Equifinality and Multifinality in Regression	178
	8.2.2 Asymmetric Causality in Regression	180
	8.3 What Is Configurational Comparative Regression Good For?	182
	8.4 Education-Based Participatory Inequality: Reanalyzing Schneider and Makszin (2014)	184
	8.4.1 Overview of the Original Study	184
	8.4.2 Turning Education-Based Participatory Inequality Inside Out	186
	8.4.3 Conclusions	192
	8.5 The Correlates of Poverty: Reanalyzing Ragin and Fiss (2017)	193
	8.5.1 Overview of the Original Study	193
	8.5.2 Turning the Correlates of Poverty “Inside Out”	194
	8.5.3 Conclusions	200
	8.6 Summary	200
9	RIO as a Gateway to Field Theory	202
	9.1 The Formalization of Social Fields	203
	9.2 Social Logics: A Decomposition of Leverage in the Hat Matrix	205
	9.3 The Co-constitution of Institutions and Properties	210
	9.4 A Network Approach to Field Analysis of Regression	211
	9.5 Summary	218

viii	Contents	
10	Conclusion	220
	10.1 Opportunities for Further Methodological Development	224
	10.2 Summary	226
	<i>Appendix A A Brief Introduction to Matrices and Matrix Multiplication</i>	227
	<i>Appendix B Computation of the Singular Value Decomposition (SVD)</i>	230
	<i>Appendix C Variance for Binomial and Count Outcomes</i>	233
	<i>Appendix D Compositional Effects in Using RIO to Detect Statistical Interactions</i>	235
	<i>Appendix E Monte Carlo Simulation Detecting Interactions by Regressing on Rows of \mathbf{P}</i>	247
	References	254
	Index	265

Figures

2.1 Scatterplot of \mathbf{Z}_Y and adjusted \mathbf{Z}_X . (Circles proportional to size of contribution of each country to the regression coefficient)	page 23
2.2 Scatter plot of $\widehat{\mathbf{Z}}_Y$ and trans.resid	35
2.3 Plot of the overall model space, as represented by \mathbf{US}^* and \mathbf{V}^*	39
2.4 Plot of the overall model space, as represented by \mathbf{US}^* and \mathbf{V}^* , coefficients as distances to the origin	40
2.5 Plot of the overall model space, as represented by \mathbf{US}^* and \mathbf{V}^* , $\widehat{\mathbf{Z}}_Y$ values as distances to the origin	42
2.6 Plot of the overall model space, as represented by \mathbf{US}^* and \mathbf{V}^* , k -means three-cluster solution	44
2.7 Depiction of $\cos(\mathbf{UU}^T)$	48
3.1 Visual representation of the logistic regression model in Table 3.1	54
3.2 Visual representation of the Poisson regression model in Table 3.2	57
3.3 Visual representation of the negative binomial regression model in Table 3.3	58
5.1 Biplot of fixed effects model using standardized variable	89
5.2 Biplots of Model 2, Table 5.2 for Denmark, Spain, Great Britain, and Italy	90
5.3 Cook’s distance for Model 1, Table 5.2	93
5.4 Biplots of Model 4 for Denmark, Spain, Great Britain, and Italy	95
5.5 Cook’s distance for clusters from Model 3, Table 5.2	96
5.6a–b Turning Model 1 (Inequality) inside out	99
5.7 Cook’s distance for Table 5.4, Model 1	101
5.8a–b Turning Model 2 (Poverty) inside out	103
5.9 Sums of squared errors for k -means cluster solutions	108
6.1 Marginal means from Model 2 in Table 6.1 illustrating the interaction between race and gender on education	116
6.2 Histogram of education from the GSS	117
6.3 Marginal means from Model 2 in Table 6.2 illustrating the interaction between education and paternal occupational prestige on respondent occupational prestige	119

6.4	Marginal means from Model 3 in Table 6.2 illustrating the interaction between paternal occupational prestige and gender on respondent occupational prestige	119
7.1	Residuals plot for Table 7.2	137
7.2	Cook's distance results for Table 7.2	137
7.3	DFBETA for (A) Government transfers, (B) GDP, and (C) Pretax/pretransfer poverty from Table 7.2	138
7.4	Diagonal values on the Hat matrix for Table 7.2	139
7.5	Deviations from the mean value on government transfers	140
7.6	OLS of effects of social welfare policies on poverty	142
7.7	Cosine similarity between Italy and all other cases	145
7.8	OLS of effects of social welfare policies on poverty by welfare regime type	149
7.9	Poverty by residualized government transfers	153
7.10a–b	Residualized GDP (a) and pretransfer poverty (b) by posttransfer poverty	154
7.11	OLS of effects of social welfare policies on poverty, most-similar	158
7.12	Model space for analysis of rebellion by excluded, (a) Variables only, (b) variables and cases, and (c) variables and observations for USA only	164
7.13	Plot of model space with cases aggregated	165
7.14	Plot of model space with cases aggregated, Kosovo included	166
7.15	Average residuals by country from Table 7.8	167
7.16	Logistic regression of rebellion by excluded, similar cases	169
8.1	Zones of data and no data	181
8.2	Contributions of cases more in versus out of sets representing three welfare types	192
9.1	The matrix $\cos(\mathbf{UU}^T)$	214
9.2	Blocked matrix $\cos(\mathbf{UU}^T)$	215
9.3	The blockmodel for the partition shown in Figure 9.2	216
9.4	Residuals from the model of Figure 9.3	217
D.1	Differences between subsets with interaction included	238
D.2	Differences between subsets without interaction	239
D.3	Differences between subsets with interaction	240
D.4	Differences between subsets without interaction	242
D.5	Differences between subsets with interaction	243
D.6	Differences between subsets without interaction	244

Tables

2.1 Independent and dependent variables, taken from Kenworthy (1999)	page 19
2.2 Contributions to regression coefficient, simple regression	22
2.3 OLS regression of Z_Y on the variables in Z_X	26
2.4 Predicted \widehat{Z}_Y for OLS regression of Z_Y on the variables in Z_X	26
2.5 OLS regression coefficients as sums across cases	29
2.6a–c SVD matrices for Z_X (Kenworthy 1999)	31
2.7 Transpose of matrix $VS^{-1}U^T$ for OLS of posttransfer poverty	32
2.8 Residualized transfers	34
2.9 Singular value decomposition of a 15×3 matrix consisting of transfers (scaled), pre-pov (scaled), and the \widehat{Z}_Y from Table 2.4	38
2.10 Distances of nations' projections	42
2.11 Values above or below the mean (+1 or -1), for each variable, with nations grouped by the three clusters	45
3.1 Summary of a logistic regression predicting whether the respondent has had an extramarital affair	53
3.2 Summary of a Poisson regression predicting the count of extramarital affairs	56
3.3 Summary of a negative binomial regression predicting the count of extramarital affairs	58
4.1 The results of OLS regression	67
4.2 Raw data (Kenworthy 1999) with squared residuals, MSE_j , and case variances	68
4.3 The 15-country data set (Kenworthy 1999)	73
4.4 The results of the OLS regression	73
5.1 Summary of examples	82
5.2 Replication of Beckfield's (2006) results	88
5.3 Contributions of Great Britain and Italy versus other countries	91
5.4 FE with and without Great Britain (GBR) and Italy (ITA)	92
5.5 FE without Italy, without Great Britain	94

xii

List of Tables

5.6	Replicating Rambotti and Breiger (2020)	98
5.7	Results for Model 1, Table 5.6 with countries omitted	102
5.8	Correlates of dog ownership	106
5.9	Descriptive statistics by cluster	109
5.10	Logistic regression coefficients by cluster	110
6.1	Summary of two OLS models predicting respondent education	115
6.2	Summary of three OLS models predicting occupational prestige	117
6.3	Summary of an OLS regression model predicting the effect of gender on education	121
7.1	Case selection methods	131
7.2	Kenworthy's (1999: 1131) Table 4, government transfers	136
7.3	Best matched pairs of countries based on Mahalanobis distance calculated on independent variable matrix for Table 7.2	141
7.4	Cases representing each selection strategy for Kenworthy (1999)	142
7.5	Coordinates for points in Figure 7.7	143
7.6	Contributions to regression model by welfare regime type	150
7.7	Case contributions to regression coefficients from Table 7.2	151
7.8	Logistic regression of rebellion by excluded	161
8.1	Summary of Schneider and Makszin's (2014) three welfare types	186
8.2a–b	Matrices V and US for labor market expenditures and social expenditures	188
8.3	Replication of Schneider and Makszin's (2014) Table 1, Model 4	189
8.4	Contributions of cases more in versus out of sets representing three welfare types	190
8.5	Logistic regression of likelihood of a person being in poverty	195
8.6	Contributions of respondents by poverty status	197
8.7	Contributions of respondents by poverty status and race	198
9.1	Matrix Z_X (three variables) and Z_Y from Kenworthy (1999)	206
9.2	SVD for matrix Z_X of Table 9.1	207
9.3	Decomposing leverage (h_i) by squares of cells in UV^T	209
9.4	Mean density within each of 4×4 submatrices of Figure 9.2	216
B.1	Signs of the entries in matrix V	232

Acknowledgments

This book is the result of a collaboration that began at the University of Arizona. Ronald Breiger began working on his innovative approach to recasting regression modeling in relational terms around 2005, and in 2010, he invited Dave Melamed and Eric Schoon (then graduate students) to work with him, first as research assistants, then as full collaborators. Our first conference publication that called for “turn[ing] the conventional regression modeling ‘inside out’ to reveal a network of relations among the cases on the basis of their attribute and behavioral similarity” was Breiger and colleagues (2011). Over the next dozen years, the three of us worked together to continue to develop, refine, and explore the implications of turning regression models inside out.

We would like to express special thanks to the direct collaborators and coauthors with whom we have been privileged to work in developing and refining aspects of regression inside out, including Andrew P. Davis, Alessandro Lomi, Simone Rambotti, and Eunsung Yoon (with whom we have coauthored Chapter 4 of this book). Thanks to Simone also for coining the acronym RIO for our shared aspiration to turn regression modeling inside out. Additionally, we would like to thank our collaborators and coauthors who have worked with us in exploring and developing substantive applications: Gary Ackerman, Victor Asal, H. Brinton Milward, Lauren Pinson, and R. Karl Rethemeyer.

As with any work that has developed over such a long time, it is difficult to acknowledge everyone who has offered support, advice, or feedback on the ideas presented here. We will likely miss some people who have influenced this work, and for that we are sorry. That being said, we express our thanks to the following individuals. In the early days, Ron was the beneficiary of exceptional advice, support, criticism, and encouragement from Omar Lizardo and Kieran Healy. Jim Moody has been an unflagging source of astute insight, questioning, and inspiration. We have received additional valuable feedback, criticism, and encouragement over the years from Elisa Bienenstock, Matthew Brashears, Carter Butts, Christina Diaz, Patrick Doreian, the late Scott Eliason, Jeremy Fiel, Lane Kenworthy, Balázs Kovács, Erin Leahey,

John Levi Martin, Charles Ragin, and A. Joseph West, for all of which we are grateful. Additionally, we are grateful to Jason Beckfield, Charles Ragin, and Carsten Schneider, who shared replications materials from their published work with us, thereby meaningfully advancing our efforts to engage with substantively important research of exceptional quality.

We have benefited greatly from questions, critiques, and suggestions made by participants in conferences, colloquia, and workshops, including the 2005 Sunbelt Social Networks Conference, the Columbia University Networks and Time Seminar Series, the Duke University Network Analysis Center, Emory University's program in Quantitative Theory and Methods, the MaxPo Center on Coping with Instability in Market Societies (Paris, France), the Networks in the Global World conference (St. Petersburg, Russia), the NYU-Abu Dhabi Program in Social Research and Public Policy, the Quantifying Social Fields Conference at the University of California, Berkeley, the Stanford University Center for Education Research, and the Yale University Institute for Network Science.

We are grateful to the editors of the *Strategies for Social Inquiry* series, Colin Elman, John Gerring, and James Mahoney, for the support and encouragement they provided us throughout this project. Special thanks to them for exposing us to the critical insights and inspiration of Adam Slez and the late Bear Braumoeller, who participated along with the editors in an intensive book workshop the editors organized in September 2021, while we were in the midst of writing this book, and whose comments led to substantial improvements in our manuscript. We would especially like to thank Jim Mahoney for shepherding us through the early stages of the proposal process.

We thank John Haslam and his team at Cambridge University Press for their professional support. We also thank Christopher Kleps for his research assistance during the completion of this manuscript.

Elements of this work have been supported through funding from the Defense Threat Reduction Agency (HDTRA1-10-1-0017; Ronald Breiger, PI, 2010-2016) and the Intelligence Advanced Research Projects Activity (2018-17121900006; Eric Schoon, PI, 2018-2019).¹

Finally, we would like to thank our families for the continual support and encouragement that made this book and the research that went into it possible. We dedicate this book to them.

¹ The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of DTRA, ODNI, IARPA, or the US government. The US government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation herein.