Robustness Tests for Quantitative Research

The uncertainty researchers face in specifying their estimation models threatens the validity of their inferences. In regression analyses of observational data the "true model" remains unknown and researchers face a choice between plausible alternative specifications. Robustness testing allows researchers to explore the stability of their main estimates to plausible variations in model specifications. This highly accessible book presents the logic of robustness testing, provides an operational definition of robustness that can be applied in all quantitative research and introduces readers to diverse types of robustness tests. Focusing on each dimension of model uncertainty in separate chapters, the authors provide a systematic overview of existing tests and develop many new ones. Whether it be uncertainty about the population or sample, measurement, the set of explanatory variables and their functional form, causal or temporal heterogeneity, effect dynamics or spatial dependence, this book provides guidance and offers tests that researchers from across the social sciences can employ in their own research.

ERIC NEUMAYER is Professor of Environment and Development and Pro-Director Faculty Development at the London School of Economics and Political Science (LSE).

THOMAS PLÜMPER is Professor of Quantitative Social Research at the Vienna University of Economics and Business.

Both authors have published in highly ranked journals including the *American Journal of Public Health*, *Annals of the American Association of Geographers*, *International Organization*, *Political Analysis* and *World Development*.

Methodological Tools in the Social Sciences

The Methodological Tools in the Social Sciences series is comprised of accessible, stand-alone treatments of methodological topics encountered by social science researchers. The focus is on practical instruction for applying methods, for getting the methods right. The authors are leading researchers able to provide extensive examples of applications of the methods covered in each book. The books in the series strike a balance between the underlying theory and the implementation of the methods. They are accessible and discursive, and make technical code and data available to aid in replication and extension of the results, as well as enabling scholars to apply these methods to their own substantive problems. They also provide accessible advice on how to present results obtained from using the relevant methods.

Robustness Tests for Quantitative Research

Eric Neumayer
London School of Economics and Political Science

Thomas Plümper
Vienna University of Economics and Business

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

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

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

4843/24, 2nd Floor, Ansari Road, Daryaganj, Delhi – 110002, India

79 Anson Road, #06–04/06, Singapore 079906

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/9781108415392
DOI: 10.1017/9781108233590

First published 2017

Printed in the United Kingdom by Clays, St Ives plc

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

Library of Congress Cataloging-in-Publication Data
Names: Neumayer, Eric, 1970– author. | Plümper, Thomas, author.
Title: Robustness tests : causal inference with observational data / Eric Neumayer, London School of Economics and Political Science, Thomas Plümper.
Description: Cambridge, United Kingdom ; New York, NY, USA : University Printing House, [2017] | Includes bibliographical references and index.
Identifiers: LCCN 2017012398| ISBN 9781108415392 (hardback : alk. paper) | ISBN 9781108401388 (paperback : alk. paper)
Subjects: LCSH: Robust statistics. | Social sciences – Methodology.
Classification: LCC HA31.5 .N48 2017 | DDC 519.5/4–dc23
LC record available at https://lccn.loc.gov/2017012398

ISBN 978-1-108-41539-2 Hardback
ISBN 978-1-108-40138-8 Paperback

Contents

Figures

Tables

Robustness Tests

Population Definition and Sample Tests		
Name	**Action**	**Pages**
Population boundary test	a) Includes observations that may not belong to population	94
	b) Excludes observations that may belong to population	
	c) Structured permutation variants	
Bootstrap test	Resamples with replacement	95
Jackknife test	Drops one or more observations at a time	95
Core group test	Includes only cases known to be in the population	97
Outlier elimination test	Drops "outliers"	97
Cross-validation test	Draws new sample	100
Selection test	Assumes cases are selected	101
Stratification test	a) Over-samples under-represented cases	103
	robustness limit variant	
	b) Under-samples over-represented cases	
	robustness limit variant	
Interpolation test	Replaces missings by interpolated values	105
Out-of-sample prediction test	Replaces missings by theory-based out of sample predictions	106
Multiple imputation test	Replaces missings by multiply imputed values	106

Concept Validity and Measurement Tests		
Name	**Action**	**Pages**
Alternative proxy test	Replaces one proxy variable with another	118
Principal component test	Combines multiple proxies into single principal component	118
Randomized components test	Randomizes weights of components of composite variable	119
Principal component jitter test	Adds "jitter" to weights of principal component	120
Rescaling test	Changes the scale of a variable	124
Measurement error injection test	Adds "artificial" measurement error	124
	Robustness limit variant	
Re-categorization test	Changes the assigned categories of categorical variables	127

Explanatory and Omitted Variables Tests		
Name	**Action**	**Pages**
Explanatory variables test	Adds and/or removes right-hand-side variables	133
Between-variation test	Stepwise reduces the between-variation in the data	136
Groupwise fixed-effects test	Reduces the between-variation by including group dummies for similar cases	138
	Robustness limit variant	
Correlated artificial variable test	Adds artificial variable with defined properties	140
Spatial-error test	Adds spatial-error variable	141

Functional Form Tests		
Name	**Action**	**Pages**
Higher degree polynomial test	Relaxes the functional form by estimating a higher-degree polynomial model	149
Semi-parametric test	Relaxes the functional form by estimating a semi-parametric model	154
Functional form break test	Allows for different effect strength at break point	155

Causal Heterogeneity and Context Conditionality Tests		
Name	**Action**	**Pages**
Bootstrap test	Tests the homogeneity assumption by resampling with replacement	162
Sample split test	Tests the homogeneity assumption by splitting the sample	162
Groupwise jackknife test	Tests the homogeneity assumption by groupwise dropping cases	163
Unit-specific effect test	Relaxes the homogeneity assumption by estimating case-specific effects	163
Group-specific effect test	Relaxes the homogeneity assumption by estimating group-specific effects	163
Random coefficients test	Allows random variation in effects	165
Multilevel test	Allows clustered variation in effects	165
Chow test of group heterogeneity	Relaxes the homogeneity assumption by interacting variables with group dummies	165
	Randomized permutation variant	
Conditionality test	Relaxes the homogeneity assumption by estimating an interaction effects model	168
Non-linear conditionality test	Relaxes the linear-symmetry assumption of interaction effects model	170

Temporal Heterogeneity Tests		
Name	**Action**	**Pages**
Trended effect test	Interacts variables with measure of time	182
Period fixed-effects test	Allows for unobserved time-varying heterogeneity	183
Temporal splines test	Allows for unobserved time-varying heterogeneity	183
Chow test of temporal heterogeneity	Interacts variables with dummies for shock or break periods	185
	Structured permutation variant	

Effect Dynamics Tests		
Name	**Action**	**Pages**
Temporal aggregation test	Changes the definition of periods	196
Dynamic specification test	Replaces static baseline with dynamic specification	197
Dynamic constraints test	Relaxes dynamic constraints	198
Alternative dynamic specification test	Replaces baseline dynamic specification	198
Effect onset test	Changes the assumption on lags and leads	199
Effect duration test	Changes the assumption on the duration of an effect	199
Temporal function form test	Changes the evolution of effect strength over periods	199
Random coefficients dynamics test	Allows for random variation in dynamics across units	203
Unit-specific dynamics test	Allows for unit-specific dynamics	203
Group-specific dynamics test	Allows for group-specific dynamics	203
Dynamic interaction effect test	Models conditional dynamics	203

Spatial Correlation and Dependence Tests		
Name	**Action**	**Pages**
Spatial specification test	Allows for spatial correlation	214
Spatial-error test	Controls for spatially correlated error processes	220
Alternative connectivity test	Varies the assumed mechanism causing spatial dependence	221
Connectivity functional form test	Varies the functional form of the connectivity variable	223
Heterogeneous exposure test	Allows for heterogeneity in exposure to spatial stimulus	223
Heterogeneous responsiveness test	Allows for heterogeneity in response to spatial stimulus	226
Spatial placebo variable test	Includes a placebo spatial-effect variable with randomized weights	226

Acknowledgments

This book provides a systematic foundation and justification for robustness testing as a way to improve the validity of causal inferences based on regression analysis of observational data. We define and operationalize robustness, and we make it measureable on a scale that ranges from 0 to 1. Not every social scientist will like the fact that we institutionalize a term, and many will disagree with our definition of robustness. This is very much intended. Definitions are not correct, they can only be useful. We believe in the usefulness of our definition. The same goes for the large number of robustness tests that we propose in part 2 of the book.

We have come a long way since we began the robustness project in 2012. We were helped along by the fact that the UK's Economic and Social Research Council (ESRC) did not have a program of funding risky but potentially transformative research agendas and that corresponding research proposals were not likely to be accepted by the ESRC's standard review process. As a consequence, the ESRC established a funding scheme for potentially "transformative research" and invented a fundamentally different refereeing process in which UK universities were only allowed to submit up to two proposals. Our project was successfully submitted through the University of Essex, for which we thank its then Pro-Vice Chancellor Research, Professor David Sanders.

We presented parts and preliminary findings of the project on various occasions: the annual conferences of the European Political Science Association in 2012 (Berlin), 2013 (Barcelona), 2014 (Edinburgh), 2015 (Vienna), and 2016 (Brussels), the annual meetings of the American Political Science Association in 2014 (Washington D.C.) and 2015 (San Francisco), the PSA Quantitative Network Conference at the University of Oxford in 2015 (thanks to Steve Fisher and Andreas Murr for organizing the meeting and inviting our presentation). Our incredible friends in Texas were so kind as to give us the opportunity to present central concepts of the book at the University of Texas at Austin's Research Methodology Conference organized and hosted by Chris Wlezien. We thank Harold Clarke, Guy Whitten, Stephen Jessee, Bob Luskin, Tse-Min Lin, Chris Wlezien, Marianne Stewart, Zach Elkins, Mike Findley, Randy Stephenson, and Robert Moser for comments and stimulating discussions. We also thank Guy Whitten for

organizing the annual "come to Texas when it's cold in Europe" workshop at Texas A&M. Guy gave us the opportunity to present central aspects of the book in a special workshop.

The ESRC transformative research grant gave us the opportunity to organize a workshop on robustness testing at the University of Cologne. Many thanks to Andre Kaiser for hosting the conference, and to Adel Daoud, Alexander Schmidt-Catran, Andre Kaiser, Cameron Wimpy, Daniel Stegmüller, Guy Whitten, Holger Döring, Katja Möhring, Laron Williams, Luke Keele, Mark Kayser, Ole Ostendorf, Philip Manow, Ray Duch, Richard Traunmüller, Simon Hug, Thomas Gschwend, Rene Lindstädt, and Thomas König for presenting papers at the conference and for stimulating debates. Finally, we wish to thank John Haslam and Stephanie Taylor at Cambridge University Press for their support.

Numerous colleagues have read and commented on various parts at various stages of the project. In alphabetical order we want to mention: Dominik Hangartner, Dirk Junge, Bob Luskin, Alex Quiroz, Tom Scotto, Daniel Stegmüller, and Richard Traunmüller. Special thanks to Dirk Junge who helped in programming some of the tests proposed in this book.