Random Sets in Econometrics

Random set theory is a fascinating branch of mathematics that amalgamates techniques from topology, convex geometry, and probability theory. Social scientists routinely conduct empirical work with data and assumptions that reveal a set to which the parameter of interest belongs, but not its exact value. Random set theory provides a coherent mathematical framework to conduct identification analysis and statistical inference in this setting and has become a fundamental tool in econometrics and finance. This is the first book dedicated to the use of the theory in econometrics written to be accessible for readers without a background in pure mathematics. Molchanov and Molinari define the basics of the theory and illustrate the mathematical concepts by their application in the analysis of econometric models. The book includes sets of exercises to accompany each chapter as well as examples to help readers apply the theory effectively.

Ilya Molchanov is Professor of Probability at the University of Bern, Switzerland, having previously worked in Germany, the Netherlands, and Scotland. His research and publications focus on probability theory, spatial statistics, and mathematical finance, with the main emphasis on stochastic geometry and the theory of random sets.

Francesca Molinari is the H. T. Warshow and Robert Irving Warshow Professor of Economics at Cornell University. She is a specialist in econometric theory, with emphasis on partial identification. She has authored numerous papers in this area, including empirical ones on estimation of risk preferences using property insurance data.

Econometric Society Monograph Series

The Econometric Society is an international society for the advancement of economic theory in relation to statistics and mathematics. The Econometric Society Monograph series is designed to promote the publication of original research contributions of high quality in mathematical economics and theoretical and applied econometrics.

Books in the Series

B. Honore, A. Pakes, M. Piazzesi, & L. Samuelson (eds.), *Advances in Economics and Econometrics: Eleventh World Congress, Vols. I & II*, 2017
S. Maurer, *On the Shoulders of Giants: Colleagues Remember Suzanne Scotchmer's Contributions to Economics*, 2017
C. P. Chambers & F. Echenique, *Revealed Preference Theory*, 2016
J.-F. Mertens, S. Sorins, & S. Samir, *Repeated Games*, 2015
C. Hsiao, *Analysis of Panel Data, 3rd ed.*, 2014
C. Cameron & P. Trivedi, *Regression Analysis of Count Data, 2nd ed.*, 2013
A. Harvey, *Dynamic Models for Volatility and Heavy Tails, with Applications to Financial and Economic Time Series*, 2013
D. Acemoglu, M. Areilano, & E. Dekel (eds.), *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress, Vols. I, II, & III*, 2013
M. Fleurbaey and F. Maniquet, *A Theory of Fairness and Social Justice*, 2011
R. Vohra, *Mechanism Design: A Linear Programming Approach*, 2011
K. Samphantharak & R. Townsend, *Households as Corporate Firms: An Analysis of Household Finance Using Integrated Household Surveys and Corporate Financial Accounting*, 2009
I. Gilboa, *Theory of Decision under Uncertainty*, 2009
F. Vega-Redondo, *Complex Networks*, 2007
R. Blundell, W. Newey, & T. Persson, *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Vols. I, II, & III*, 2006
J. Roemer, *Democracy, Education, and Equality*, 2006
C. Blackorby, W. Bossert, & D. Donaldson, *Population Issues in Social Choice Theory, Welfare Economics and Ethics*, 2005
R. Koenker, *Quantile Regression*, 2005
C. Hsiao, *Analysis of Panel Data, 2nd ed.*, 2003
M. Dewatripont, L. P. Hausen, & S. J. Turnovsky (eds.), *Advances in Economics and Econometrics:Theory and Applications, Eighth World Congress, Vols. I, II, & III*, 2003
E. Ghysels, N. Swanson, & M. Watson (eds.), *Essays in Econometrics: Collected Papers of Clive W. J. Granger, Vols. I & II*, 2001.
S. Strøm (ed.), *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*, 1999.
A. C. Cameron & P. K. Trivedi, *Regression Analysis of Count-Data*, 1998.
D. Jacobs, E. Kalai, & M. Kamien (eds.), *Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures*, 1998.
D. M. Kreps & K.F. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress, Vols. I, II, & III*, 1997
R. Guesnerie, *A Contribution to the Pure Theory of Taxation*, 1995
C. Sims (ed.), *Advances in Econometrics, Sixth World Congress, Vols. I & II*, 1994
H. White *Inference, Estimation and Specification Analysis*, 1994

(Continued after the index)

Cambridge University Press
978-1-107-12120-1 — Random Sets in Econometrics
Ilya Molchanov , Francesca Molinari
Frontmatter
More Information

Random Sets in Econometrics

Ilya Molchanov
University of Bern

Francesca Molinari
Cornell University

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

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, 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/9781107121201
DOI: 10.1017/9781316392973

First published 2018

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

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

Library of Congress Cataloging-in-Publication Data
Names: Molchanov, Ilya S., 1962- author. | Molinari, Francesca, author.
Title: Random sets in econometrics / Ilya Molchanov, University of Bern,
Francesca Molinari, Cornell University.
Description: Cambridge, United Kingdom ; New York, NY : Cambridge University
Press, 2018. | Includes bibliographical references and index.
Identifiers: LCCN 2017055335| ISBN 9781107121201 (hbk : alk. paper) | ISBN
9781107548732 (pbk : alk. paper)
Subjects: LCSH: Econometrics. | Random sets.
Classification: LCC HB139 .M64 2018 | DDC 330.01/5192–dc23 LC record available at
https://lccn.loc.gov/2017055335

ISBN 978-1-107-12120-1 Hardback
ISBN 978-1-107-54873-2 Paperback

Cambridge University Press
978-1-107-12120-1 — Random Sets in Econometrics
Ilya Molchanov , Francesca Molinari
Frontmatter
More Information

To Liza, Lucine, Niccolò, and Ivan

Contents

Preface

"...there is enormous scope for fruitful inference using data and assumptions that partially identify population parameters."

C. F. Manski [104, p. 2]

"...this will be a basic book for the future since the notion of a set is the cornerstone of mathematics."

G. S. Watson, Preface to G. Matheron's book [109]

Random set theory is concerned with the development of a coherent mathematical framework to study random objects whose realizations are sets.[1] Such objects appeared a long time ago in statistics and econometrics in the form of confidence regions, which can be naturally described as random sets. The first idea of a general random set in the form of a region that depends on chance appears in Kolmogorov [89], originally published in 1933. A systematic development of the theory of random sets did not occur until a while later, stimulated by the study in general equilibrium theory and decision theory of correspondences and nonadditive functionals, as well as the needs in image analysis, microscopy, and material science, of statistical techniques to develop models for random sets, estimate their parameters, filter noisy images, and classify biological images.

These and other related applications of set-valued random variables induced the development of statistical models for random sets, furthered the understanding of their distributions, and led to the seminal contributions of Choquet [39], Aumann [11], and Debreu [46] and to the first self-contained treatment of the theory of random sets given by Matheron [109]. Since then, the theory expanded in several directions, developing its relationship with convex geometry and providing various limit theorems for random sets and set-valued processes, and more. A detailed account of the modern mathematical theory

[1] This preface is based largely on one of our published articles, Molchanov and Molinari [119].

of random sets is provided by Molchanov [117]; we systematically refer to the second edition of this monograph which is now available.

More recently, the development within econometrics of partial identification analysis on one side, and financial models with transaction costs on the other, have provided a new and natural area of application for random set theory.

Partially identified econometric models appear when the available data and maintained assumptions do not suffice to uniquely identify the statistical functional of interest, whether finite- or infinite-dimensional, even as data accumulate; see Tamer [150] for a review and Manski [104] for a systematic treatment. For this class of models, partial identification proposes that econometric analysis should study the set of values for the statistical functional which are observationally equivalent: these are the parameter values that could generate the same distribution of observables as the one in the data, for some data-generating process consistent with the maintained assumptions. In this book, this set of values is referred to as the functional's sharp identification region. The goals of the analysis are to obtain a tractable characterization of the sharp identification region, to provide methods for estimating it, and to conduct tests of hypotheses and make confidence statements about it.

Conceptually, partial identification predicates a shift of focus from single valued to set-valued objects, which renders it naturally suited for the use of random set theory as a mathematical framework to conduct identification analysis and statistical inference, and to unify a number of special results and produce novel general results. The random sets approach complements the more traditional one, based on mathematical tools for (single-valued) random vectors, that has proved extremely productive since the beginning of the research program in partial identification.

While the traditional approach has provided tractable characterizations of the sharp identification region in many econometric applications of substantial interest (see, e.g., the results in Manski [104]), there exist many important problems in which such a characterization is difficult to obtain. This is the case, for example, when one is interested in learning the identified features of best linear predictors (ordinary least squares) in the presence of missing or interval-valued outcome and covariate data, or in learning the identified features of payoff functions in finite games with multiple pure strategy Nash equilibria. These difficulties have proven so severe that, until the introduction of random set methods in econometrics, researchers had turned to characterizing regions in the parameter space that include all the parameter values that may have generated the observables, but may include other (infeasible) parameter values as well. These larger regions are called "outer regions." The inclusion in the outer regions of parameter values that are infeasible may weaken the researchers' ability to make useful predictions and to test for model misspecification.

Turning to statistical inference, the traditional approach in partial identification based on laws of large numbers and central limit theorems for single-valued random vectors provided general procedures that are applicable for a wide range of econometric models. In certain cases, however, it is possible to use random set methods to characterize directly the asymptotic distribution of set-valued estimators, in particular by working with their boundary structure, thereby obtaining inference procedures that can be simpler to apply.

The connection between partial identification and random set theory stems from the fact that a lack of point identification can generally be traced back to a collection of random variables that are consistent with the available data and maintained assumptions. Examples include interval data in regression models and multiple equilibria in game theoretic models. In many cases, this collection of random variables is equal to the family of selections of a properly specified random closed set and random set theory can be applied to describe their distribution and to derive statistical properties of estimators that rely upon them.

In order to fruitfully apply random set theory for identification and inference, the econometrician needs to carry out three fundamental steps. First, she needs to define the random closed set that is relevant for the problem under consideration using all information given by the available data and maintained assumptions. This is a delicate task, but one that is typically carried out in identification analysis regardless of whether random set theory is applied. Second, she needs to determine how the observable random variables relate to this random closed set. Often, one of two cases occurs: either the observable variables determine a random set to which the (unobservable) variable of interest belongs with probability one or the (expectation of the) (un)observable variable belongs to (the expectation of) a random set determined by the model. Finally, the econometrician needs to determine which tool from random set theory should be utilized. To date, new applications of random set theory to econometrics have fruitfully exploited (Aumann) selection expectations and their support functions, (Choquet) capacity functionals, and laws of large numbers and central limit theorems for random sets.

In finance it is possible to represent the range of prices (which are always non-unique in case of transaction costs) as random sets. In the univariate case, this set is a segment, with the end-points being bid and ask prices. The no-arbitrage property of the dynamic model with discrete time means that a trading strategy that, starting from no investment, leads to a non-trivial non-negative outcome with probability one is impossible. Since the prices change with time, they can be represented as a set-valued process in discrete time. Then the no-arbitrage property holds (in the univariate case) if and only if there exists a martingale with respect to an equivalent probability measure that evolves inside the set-valued process. In case of several assets, it is typical to work with conical random sets that represent all solvent positions on several assets, where negative amounts in some of the assets are compensated by the positive amounts in the others (see Kabanov and Safarian [81]).

The goal of this book is to introduce the theory of random sets from the perspective of applications in econometrics. Our view is that the instruction of random set theory could be fruitfully incorporated into Ph.D.-level field courses in econometrics on partial identification and in microeconomics on decision theory. Important prerequisites for the study of random set theory include measure theory and probability theory; good knowledge of convex analysis and general topology is beneficial but not essential.

The book is organized as follows. Chapter 1 provides basic notions of random set theory, including the definition of a random set and of the functional that characterizes its distribution. Chapter 2 focuses on the selections of random sets: these are the random elements that almost surely belong to the random set. The most important result in the chapter (from the perspective of applications in econometrics, and particularly in partial identification) is Theorem 2.13, which provides a necessary and sufficient condition characterizing selections in terms of a dominance property between their distribution and the distribution of the random set. This characterization leads to a natural sample analog that can be used for estimation and inference. Chapter 3 introduces the concept of selection expectation of a random set. If the random set is defined on a nonatomic probability space, its selection expectation is always convex. This means that the boundary of the selection expectation is uniquely characterized by its support function. This fact is used to provide necessary and sufficient conditions for the existence of selections with given moments, which again lead to natural sample analogs that can be used for estimation and inference. Chapter 4 introduces the Minkowski sum of random sets, which equals the set of sums of all their points or all their selections and can be equivalently defined using the arithmetic sum of the support functions of the random sets. Laws of large numbers and central limit theorems for Minkowski sums of random sets are derived building on existing results in functional spaces, exploiting the connection between convex sets and their support function. Chapter 5 discusses estimation and inference of sets of functionals defined via inequalities, with particular emphasis on inequalities involving the probability distribution of random sets. In each chapter, results from the theory of random sets are presented alongside applications in partial identification.

Throughout the book, we use the capital Latin letters A, B, K, L, M, F to denote deterministic (non-random) sets, and bold ones $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$, etc. to denote random sets. We use the lowercase Latin letters u, v to denote points and s, t to denote scalars. We use Greek letter ε to denote unobservable random vectors, and lowercase bold Latin letters $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$, etc. to denote observable ones. We denote parameter vectors and sets of parameter vectors, respectively, by θ and Θ, and for a given parameter θ we denote its sharp identification region by $\mathsf{H}[\theta]$.

The theory of random closed sets generally applies to the space of closed subsets of a locally compact Hausdorff second countable topological space $\mathfrak{X}$, which is often assumed to be the Euclidean space denoted by $\mathbb{R}^d$.

To ease the flow of exposition, we make no use of footnotes, but rather use end-of-chapter notes. We also postpone the vast majority of references to the existing literature to these chapter notes.

Acknowledgments

Many of the applications of random set theory to econometrics presented in this book are based on joint work of the authors with Arie Beresteanu. We are grateful to Arie for a very productive collaboration. The writing of this book was stimulated by a CeMMAP Masterclass on "Random Sets in Economics" delivered by Ilya Molchanov on April 18–19, 2012 at CeMMAP in London, UK, and later at the Universidad Carlos III de Madrid. We are grateful to Andrew Chesher and Adam Rosen for organizing the CeMMAP Masterclass and the associated conference on "New Developments in the Use of Random Sets in Economics." We are grateful to Adam Rosen for many stimulating conversations on the topics of this book and for organizing a second CeMMAP Masterclass on "Applications of Random Set Methods in Econometrics" delivered by Ilya Molchanov and Francesca Molinari on June 19–20, 2014 at CeMMAP in London, UK. We are grateful to Rosa Matzkin and an anonymous reviewer for detailed comments on an earlier version of this book, and to Matthew Thirkettle for detailed and constructive comments on a second draft. We thank Larry Blume, David Easley, Chuck Manski, and Whitney Newey for many conversations on various topics that appear in this book and for their interest and continued support of our research program on applications of random set theory in econometrics.

We also thank UCL, UC3M, and Oberwolfach for their hospitality, and the Swiss National Science Foundation Grants 200021_153597 and IZ73Z0_152292 (IM) and the USA National Science Foundation Grant SES-0922330 (FM) for financially supporting part of the efforts that have led to this book.