

THE ECONOMETRICS OF
COMPLEX SURVEY DATA:
THEORY AND APPLICATIONS

ADVANCES IN ECONOMETRICS

Series editors: Thomas B. Fomby,
R. Carter Hill, Ivan Jeliazkov,
Juan Carlos Escanciano,
Eric Hillebrand, Daniel L. Millimet,
Rodney Strachan, David T. Jacho-Chávez,
Alicia Rambaldi

Previous Volumes

- Volume 21: Modelling and Evaluating Treatment Effects in Econometrics – Edited by Daniel L. Millimet, Jeffrey A. Smith and Edward Vytlačil
- Volume 22: Econometrics and Risk Management – Edited by Jean-Pierre Fouque, Thomas B. Fomby and Knut Solna
- Volume 23: Bayesian Econometrics – Edited by Siddhartha Chib, Gary Koop, Bill Griffiths and Dek Terrell
- Volume 24: Measurement Error: Consequences, Applications and Solutions – Edited by Jane Binner, David Edgerton and Thomas Elger
- Volume 25: Nonparametric Econometric Methods – Edited by Qi Li and Jeffrey S. Racine
- Volume 26: Maximum Simulated Likelihood Methods and Applications – Edited by R. Carter Hill and William Greene
- Volume 27A: Missing Data Methods: Cross-Sectional Methods and Applications – Edited by David M. Drukker
- Volume 27B: Missing Data Methods: Time-Series Methods and Applications – Edited by David M. Drukker
- Volume 28: DSGE Models in Macroeconomics: Estimation, Evaluation and New Developments – Edited by Nathan Balke, Fabio Canova, Fabio Milani and Mark Wynne
- Volume 29: Essays in Honor of Jerry Hausman – Edited by Badi H. Baltagi, Whitney Newey, Hal White and R. Carter Hill
- Volume 30: 30th Anniversary Edition – Edited by Dek Terrell and Daniel Millimet
- Volume 31: Structural Econometric Models – Edited by Eugene Choo and Matthew Shum
- Volume 32: VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims – Edited by Thomas B. Fomby, Lutz Kilian and Anthony Murphy
- Volume 33: Essays in Honor of Peter C. B. Phillips – Edited by Thomas B. Fomby, Yoosoon Chang and Joon Y. Park
- Volume 34: Bayesian Model Comparison – Edited by Ivan Jeliazkov and Dale J. Poirier
- Volume 35: Dynamic Factor Models – Edited by Eric Hillebrand and Siem Jan Koopman
- Volume 36: Essays in Honor of Aman Ullah – Edited by Gloria Gonzalez-Rivera, R. Carter Hill and Tae-Hwy Lee
- Volume 37: Spatial Econometrics – Edited by Badi H. Baltagi, James P. LeSage, and R. Kelley Pace
- Volume 38: Regression Discontinuity Designs: Theory and Applications – Edited by Matias D. Cattaneo and Juan Carlos Escanciano

ADVANCES IN ECONOMETRICS VOLUME 39

THE ECONOMETRICS OF COMPLEX SURVEY DATA: THEORY AND APPLICATIONS

EDITED BY

KIM P. HUYNH
Bank of Canada, Canada

DAVID T. JACHO-CHÁVEZ
Emory University, USA

GAUTAM TRIPATHI
University of Luxembourg, Luxembourg



United Kingdom – North America – Japan
India – Malaysia – China

Emerald Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2019

Copyright © 2019 Emerald Publishing Limited

Reprints and permissions service

Contact: permissions@emeraldinsight.com

No part of this book may be reproduced, stored in a retrieval system, transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without either the prior written permission of the publisher or a licence permitting restricted copying issued in the UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center. Any opinions expressed in the chapters are those of the authors. Whilst Emerald makes every effort to ensure the quality and accuracy of its content, Emerald makes no representation implied or otherwise, as to the chapters' suitability and application and disclaims any warranties, express or implied, to their use.

British Library Cataloguing in Publication Data

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

ISBN: 978-1-78756-726-9 (Print)

ISBN: 978-1-78756-725-2 (Online)

ISBN: 978-1-78756-727-6 (Epub)

ISSN: 0731-9053 (Series)



ISOQAR certified
Management System,
awarded to Emerald
for adherence to
Environmental
standard
ISO 14001:2004.

Certificate Number 1985
ISO 14001



INVESTOR IN PEOPLE

CONTENTS

LIST OF CONTRIBUTORS	vii
INTRODUCTION	ix

PART I SAMPLING DESIGN

CAN INTERNET MATCH HIGH QUALITY TRADITIONAL SURVEYS? COMPARING THE HEALTH AND RETIREMENT STUDY AND ITS ONLINE VERSION <i>Marco Angrisani, Brian Finley, and Arie Kapteyn</i>	3
EFFECTIVENESS OF STRATIFIED RANDOM SAMPLING FOR PAYMENT CARD ACCEPTANCE AND USAGE <i>Christopher S. Henry and Tamás Ilyés</i>	35

PART II VARIANCE ESTIMATION

WILD BOOTSTRAP RANDOMIZATION INFERENCE FOR FEW TREATED CLUSTERS <i>James G. MacKinnon and Matthew D. Webb</i>	61
VARIANCE ESTIMATION FOR SURVEY-WEIGHTED DATA USING BOOTSTRAP RESAMPLING METHODS: 2013 METHODS-OF-PAYMENT SURVEY QUESTIONNAIRE <i>Heng Chen and Q. Rallye Shen</i>	87

PART III ESTIMATION AND INFERENCE

MODEL-SELECTION TESTS FOR COMPLEX SURVEY SAMPLES

Iraj Rahmani and Jeffrey M. Wooldridge 109

INFERENCE IN CONDITIONAL MOMENT RESTRICTION MODELS WHEN THERE IS SELECTION DUE TO STRATIFICATION

Antonio Cosma, Andrei V. Kostyrka, and Gautam Tripathi 137

NONPARAMETRIC KERNEL REGRESSION USING COMPLEX SURVEY DATA

Luc Clair 173

NEAREST NEIGHBOR IMPUTATION FOR GENERAL PARAMETER ESTIMATION IN SURVEY SAMPLING

Shu Yang and Jae Kwang Kim 209

PART IV APPLICATIONS IN BUSINESS, HOUSEHOLD, AND CRIME SURVEYS

IMPROVING RESPONSE QUALITY WITH PLANNED MISSING DATA: AN APPLICATION TO A SURVEY OF BANKS

Geoffrey R. Gerdes and Xuemei Liu 237

DOES SELECTIVE CRIME REPORTING INFLUENCE OUR ABILITY TO DETECT RACIAL DISCRIMINATION IN THE NYPD'S STOP-AND-FRISK PROGRAM?

Steven F. Lehrer and Louis-Pierre Lepage 259

SURVEY EVIDENCE ON BLACK MARKET LIQUOR IN COLOMBIA

*Gustavo J. Canavire-Bacarreza, Alexander L. Lundberg, and
Alejandra Montoya-Agudelo* 287

INDEX 315

LIST OF CONTRIBUTORS

<i>Marco Angrisani</i>	University of Southern California, USA
<i>Gustavo J. Canavire-Bacarreza</i>	Inter-American Development Bank, USA
<i>Heng Chen</i>	Bank of Canada, Canada
<i>Luc Clair</i>	University of Winnipeg, Canada and Canadian Centre for Agri-Food Research in Health and Medicine, Canada
<i>Antonio Cosma</i>	University of Luxembourg, Luxembourg
<i>Brian Finley</i>	University of Southern California, USA
<i>Geoffrey R. Gerdes</i>	Federal Reserve Board of Governors, USA
<i>Christopher S. Henry</i>	Bank of Canada, Canada School of Economics/CERDI, University of Auvergne, France
<i>Tamás Ilyés</i>	Magyar Nemzeti Bank, Hungary
<i>Arie Kapteyn</i>	University of Southern California, USA
<i>Andreï V. Kostyrka</i>	University of Luxembourg, Luxembourg
<i>Jae Kwang Kim</i>	Iowa State University, USA
<i>Steven F. Lehrer</i>	Queen's University, Canada
<i>Louis-Pierre Lepage</i>	University of Michigan, USA
<i>Xuemei Liu</i>	Federal Reserve Board of Governors, USA
<i>Alexander L. Lundberg</i>	West Virginia University, USA
<i>James G. MacKinnon</i>	Queen's University, Canada
<i>Alejandra Montoya-Agudelo</i>	Universidad EAFIT, Colombia
<i>Matthew D. Webb</i>	Carleton University, Canada
<i>Iraj Rahmani</i>	Nazarbayev University, Kazakhstan
<i>Q. Rallye Shen</i>	Bank of Canada, Canada
<i>Gautam Tripathi</i>	University of Luxembourg, Luxembourg
<i>Jeffrey M. Wooldridge</i>	Michigan State University, USA
<i>Shu Yang</i>	North Carolina State University, USA

INTRODUCTION

The assumption of simple random sampling is widely used in applied research in the social, behavioral and biomedical sciences, as well as in empirical public policy analysis. However, this assumption is seldom true in practice. Stratified and cluster sampling are routinely used by most statistical agencies in the world, and because of budgetary reasons, the actual sampling process may be even more complicated. Correct statistical analysis therefore requires a careful consideration of these complex survey designs when performing estimation and inference.

The papers in this volume of *Advances in Econometrics* were presented at the “Econometrics of Complex Survey Data: Theory and Applications” conference organized by the Bank of Canada, Ottawa, Canada, from October 19 to 20, 2017. The editors would like to acknowledge the generous financial support provided by the Bank of Canada.

Below is a brief overview of the papers accepted in this volume, grouped into the following four categories: (1) sampling design; (2) variance estimation; (3) estimation and inference and (4) business, household and crime surveys.

SAMPLING DESIGN

“Can Internet Match High Quality Traditional Surveys? Comparing the Health and Retirement Study and Its Online Version” by Marco Angrisani, Brian Finley and Arie Kapteyn revisit the question of comparability of online and more traditional interview modes by studying differences across Internet-based, face-to-face and phone-based surveys. They find little evidence of mode effects when comparing various outcomes providing support for internet-based surveys.

“Effectiveness of Stratified Random Sampling for Payment Card Acceptance and Usage” by Christopher S. Henry and Tamás Ilyés uses the universe of merchant cash registers in Hungary to assess the effect of stratified random sampling on estimates of payment card acceptance and usage. It compares county, industry, and store size stratifications to mimic the usual stratification criteria for standard merchant surveys. By doing this, they can quantify the effect on estimates of card acceptance for different sample sizes.

VARIANCE ESTIMATION

“Wild Bootstrap Randomization Inference for Few Treated Clusters” by James G. MacKinnon and Matthew D. Webb proposes a bootstrap-based alternative to randomization inference, which mitigates problems of over- or under-rejection in t tests in pure treatment or difference-in-differences settings when the number of clusters is very small.

“Variance Estimation for Survey-weighted Data Using Bootstrap Resampling Methods: 2013 Methods-of-Payment Survey Questionnaire” by Heng Chen and Q. Rallye Shen proposes a bootstrap-resampling method to estimate variability when sampling units are selected through an approximate stratified two-stage sampling design. Their proposed method allows for randomness from both the sampling design and the raking procedure.

ESTIMATION AND INFERENCE

“Model Selection Tests for Complex Survey Samples” by Iraj Rahmani and Jeffrey M. Wooldridge extends Vuong’s model selection test (“Likelihood Ratio Tests for Model Selection and Non-Nested Hypothesis,” *Econometrica*, 1989) to allow for complex survey samples. By using an M-estimation setting, their test applies to general estimation problems including linear and nonlinear least squares, Poisson regression and fractional response models. With cluster samples and panel data, they show how to combine the weighted objective function with a cluster-robust variance estimator, thereby expanding the scope of their test.

“Inference in Conditional Moment Restriction Models When There is Selection Due to Stratification” by Antonio Cosma, Andreï V. Kostyrka and Gautam Tripathi shows how to use a smoothed empirical likelihood approach to conduct efficient semiparametric inference in models characterized as conditional moment equalities when data are collected by variable probability sampling.

“Nonparametric Kernel Regression Using Complex Survey Data” by Luc Clair derives the asymptotic properties of a design-based nonparametric kernel-based regression estimator under a combined inference framework involving multivariate mixed data. It also proposes a least squares cross-validation procedure for selecting the bandwidth for both continuous and discrete variables. Simulation results show that the estimator is consistent and that efficiency gains can be achieved by weighting observations by the inverse of their inclusion probabilities if the sampling is endogenous.

“Nearest Neighbor Imputation for General Parameter Estimation in Survey Sampling” by Shu Yang and Jae Kwang Kim studies the asymptotic properties of the nearest neighbor population imputation estimator of population parameters

when handling item nonresponse in survey sampling. When estimating a variance, the authors propose a replication variance estimator.

BUSINESS, HOUSEHOLD AND CRIME SURVEYS

Last but not least, “Improving Response Quality with Planned Missing Data: An Application to a Survey of Banks” by Geoffrey R. Gerdes and Xuemei Liu reports a “random blocking” approach to shortening the questionnaires for individual respondents when collecting data on noncash payments by type, cash withdrawals and deposits, and related information in a survey of a population of depository institutions in the United States. Their approach is a special case of multiple matrix sampling and an extension of a split questionnaire or planned missing value design. They find that the proposed blocking approach helped increase unit-level and item-level response for smaller institutions.

“Does Selective Crime Reporting Influence Our Ability to Detect Racial Discrimination in the NYPD’s Stop-and-frisk Program?” by Steven F. Lehrer and Louis-Pierre Lepage uses data from the New York City’s Stop-and-Frisk program to assess the presence of crime type heterogeneity in racial bias and police officer decisions of reported crime type. They find evidence that differences in biases across crime types are substantial while accounting for sample-selection which may arise from conditioning on crime type.

“Survey Evidence on Black Market Liquor in Colombia” by Gustavo J. Canavire-Bacarreza, Alexander L. Lundberg and Alejandra Montoya-Agudelo uses a unique national survey on illegal liquor commissioned by the Colombian government to estimate the determinants of the demand for smuggled and adulterated liquor. To address unit and item nonresponse, they implement a multiple imputation procedure with chained equations.

Kim P. Huynh

David T. Jacho-Chávez

Gautam Tripathi