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

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

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

ISOQAR
REGISTERED

Certificate Number 1985
ISO 14001

INVESTOR IN PEOPLE

CONTENTS

PART III
ESTIMATION AND INFERENCE

PART IV
APPLICATIONS IN BUSINESS,
HOUSEHOLD, AND CRIME SURVEYS

LIST OF CONTRIBUTORS

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

<div align="right">

Kim P. Huynh

David T. Jacho-Chávez

Gautam Tripathi

</div>