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 Vytlacil
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 Millmet
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

emerald PUBLISHING

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

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

First edition 2019

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)

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

This page intentionally left blank

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

This page intentionally left blank

PART I
SAMPLING DESIGN

CAN INTERNET MATCH HIGH-QUALITY TRADITIONAL SURVEYS? COMPARING THE HEALTH AND RETIREMENT STUDY AND ITS ONLINE VERSION

Marco Angrisani, Brian Finley, and Arie Kapteyn

Center for Economic and Social Research, University of Southern California, USA

ABSTRACT

We examine sample characteristics and elicited survey measures of two studies, the Health and Retirement Study (HRS), where interviews are done either in person or by phone, and the Understanding America Study (UAS), where surveys are completed online and a replica of the HRS core questionnaire is administered. By considering variables in various domains, our investigation provides a comprehensive assessment of how Internet data collection compares to more traditional interview modes. We document clear demographic differences between the UAS and HRS samples in terms of age and education. Yet, sample weights correct for these discrepancies and allow one to satisfactorily match population benchmarks as far as key socio-demographic variables are concerned. Comparison of a variety of survey

The Econometrics of Complex Survey Data: Theory and Applications

Advances in Econometrics, Volume 39, 3–33

ISSN: 0731-9053/doi:10.1108/S0731-905320190000039001

outcomes with population targets shows a strikingly good fit for both the HRS and the UAS. Outcome distributions in the HRS are only marginally closer to population targets than outcome distributions in the UAS. These patterns arise regardless of which variables are used to construct post-stratification weights in the UAS, confirming the robustness of these results. We find little evidence of mode effects when comparing the subjective measures of self-reported health and life satisfaction across interview modes. Specifically, we do not observe very clear primacy or recency effects for either health or life satisfaction. We do observe a significant social desirability effect, driven by the presence of an interviewer, as far as life satisfaction is concerned. By and large, our results suggest that Internet surveys can match high-quality traditional surveys.

Keywords: Online survey; survey methods; weighting; survey mode effects; face-to-face interviews; online interviews

1. INTRODUCTION

The collection of high-quality data on households and individuals tends to be labor intensive, costly and slow. When adopting traditional survey modes like face-to-face or telephone interviewing, typically several years elapse from the moment a survey is designed to final data availability. The Internet, with its promise of real-time results and looming ubiquity, provides a tempting alternative for faster and more cost-effective data collection. Online surveys, however, differ from more traditional surveys in several respects which may affect both sample representativeness and data quality.

First, Internet coverage is still not entirely pervasive, especially among more economically disadvantaged groups and the elderly. Data from the Pew Research Center showed that only 51% of Americans aged 65 or older had a home broadband connection in 2016, while the fraction of home broadband owners was about 77% among 18–29 year olds. Likewise, home broadband coverage was only 53% for Americans with incomes of less than $30,000 and 93% for those with incomes of $75,000 or greater.[1] As a result, the representativeness of online surveys may be jeopardized (Schonlau et al., 2009). Telephone surveys, however, face similar difficulties with the widespread adoption of voice mail and cell phones (Blumberg, Luke, & Cynamon, 2004). Second, even with complete coverage of the population, individual characteristics are bound to influence the likelihood of completing an online survey versus a face-to-face or phone survey, thereby introducing relevant selectivity issues and nonresponse biases that may vary by interview mode (Couper, 2011). Third, mode effects need to be considered, as the same question may be answered differently in person, by phone or over the Internet (Schwarz & Sudman, 1992). Face-to-face and phone interviews leave more room for clarification and offer more control of who is actually answering the questionnaire. On the other hand, Web surveys offer more privacy and could thereby encourage more accurate and honest reporting on personal and sensitive matters, while the presence of an interviewer in face-to-face and telephone interviews may induce interviewer effects.

Chang and Krosnick (2009) compare sample representativeness and data quality of Internet-based surveys and phone-based surveys. They conclude that as long as Internet data are collected from a probability-based sample, these exhibit higher accuracy than data collected by phone. Their study, however, is limited to a rather specific topic, namely politics.

In view of the existing literature, the contribution of this paper is twofold. First, we 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. Second, we focus on a diverse set of outcomes, ranging from home ownership and labor force status to self-reported health and life satisfaction. The aforementioned sources of differences between Web surveys and face-to-face or telephone surveys may affect each of these outcomes in different ways. Thus, by considering variables in various domains, our investigation provides a more robust and comprehensive assessment of how Internet data collection compares to more traditional interview modes. Moreover, while our analysis is performed at a time when Internet coverage has increased substantially in the population, we focus (because of data availability and comparability issues) on the subgroup of individuals aged of 55 and older. Within this segment of the population, barriers to adoption of new technology may still imply significant selectivity issues and limit study generalizability, as recently pointed out by Remillard et al. (2014). The extent to which Internet data are comparable to data collected with more traditional interview modes is then of particular scientific interest in research concerned with this subpopulation.

2. METHODS AND OUTLINE

We consider and compare sample characteristics and elicited survey measures of two studies, the Health and Retirement Study (HRS), where interviews are done either in person or by phone, and the Understanding America Study (UAS), where surveys are completed online. Both the UAS and HRS maintain a panel of American households, but while the HRS focuses for the most part on its core questionnaire issued every two years, the UAS issues a wide variety of surveys to its panel. Included among the surveys administered in the UAS is a replica of the HRS core questionnaire (with some adaptation to accommodate differences in format between verbal and self-administered interviews). Thus, by examining responses to this questionnaire in the two studies, we can investigate not only differences in sample composition across these two studies but also differences in survey outcomes potentially stemming from different interview modes. In what follows, we will refer to the HRS core questionnaire simply as the HRS, and similarly refer to its replica in the UAS.

Whenever possible, we contrast HRS and UAS survey outcomes with comparable measures in the Current Population Survey (CPS), which we view as population benchmarks. The goal is to assess the extent to which HRS and UAS survey outcomes match population benchmarks and identify possible channels which observed discrepancies may stem from.

Section 3 briefly sketches the general features of the HRS and the UAS panels. Like other probability-based Internet surveys, the UAS suffers from low recruitment rates. Hence, weighting becomes critical to matching the underlying population characteristics. Section 3.3, therefore, goes into detail about the sampling and weighting procedures adopted by the UAS. Section 4 follows with a comparison of surveys' basic demographics to each other and to the reference population, as represented by the CPS. Section 5 performs similar comparisons focusing on survey outcomes such as home ownership, health insurance coverage and labor force status, as well as self-reported health and subjective well-being. By using a diverse array of measures, for which biases induced by representativeness issues and survey modes may differ, we aim to provide a fairly complete picture of the surveys' ability to match each other and the reference population. This analysis also compares different weighting procedures for the UAS to examine the importance of the choice of post-stratification variables on the quality of weighting. Section 5.1 focuses on the comparison of mode effects for two subjective variables – self-reported health and satisfaction with life. Section 6 concludes.

3. HRS AND UAS DESCRIPTIONS

3.1. The Health and Retirement Study

The HRS is a multipurpose, longitudinal household survey representing the US population over the age of 50. Since 1992, the HRS has surveyed age-eligible respondents and their spouses every two years to track transitions from work into retirement, to measure economic wellbeing in later life and to monitor changes in health status as individuals age. Starting in 2006, study participants 80 years or younger have been randomly assigned to either a phone or an "enhanced" face-to-face interview. In the former case, the HRS questionnaire is administered via computer-assisted telephone interviewing. In the latter case, the questionnaire is administered in person by an interviewer using computer-assisted personal interviewing technology and is complemented with a set of physical performance measures, collection of biomarkers and a survey on psychosocial topics. Respondents over the age of 80 are only interviewed face to face.

Initially, the HRS consisted of individuals born between 1931 and 1941 and their spouses, but additional cohorts have been added in 1998, 2004 and 2010, the youngest cohort to date comprising individuals born between 1954 and 1959. Once added, a cohort is indefinitely administered the HRS questionnaire on the same two-year cycle as previously existing panel members. Because of refresher samples over the years, the HRS is representative of households in which at least one member is 51 years old in 1998, 2004 and 2010, when new cohorts were added to the survey. In 2000, 2006 and 2012, the HRS represents households with members 53 or older; in 2002, 2008 and 2014, it represents households with members 55 and older.

We use the 2014 wave of the HRS and rely on the RAND version (Bugliari et al., 2016) of the data, a large user-friendly subset of the HRS that combines data from all waves, adds information that may have been provided by the spouse to the respondent's record and has consistent imputation of financial variables.

As mentioned above, the 2014 HRS wave is representative of individuals aged 55 or older. Accordingly, we will select only individuals who are 55 or older in both the CPS and the UAS to proceed with our comparison exercise. For HRS waves 1992–2004, the CPS was used to establish population benchmarks for the post-stratification of sample weights. Starting from 2006, however, the American Community Survey (ACS) has served as the basis for post-stratification. Hence, population targets for the 2014 HRS wave, which is used in this study, have been computed off the ACS. Post-stratification in the HRS is based on gender, age, race/ethnicity (Hispanic, Black non-Hispanic, other non-Hispanic) and geography (Metropolitan Statistical Area (MSA) and non-MSA counties).

3.2. The Understanding America Study

The UAS is a nationally representative Internet panel of approximately 6,000 respondents. It began in 2014 and is managed by the Center for Economic and Social Research at the University of Southern California. The UAS is based on a probability sample drawn from the US population aged 18 and older. Panel members are selected through address-based sampling. After joining the panel, individuals are invited to take, on average, two surveys each month. Invitations to panel members to take surveys are sent by email and surveys are answered online. Respondents are typically compensated $20 for a 30-minute survey. Individuals who do not have Internet access are provided with both access and a tablet for completing surveys. This is a very important feature of the recruitment procedure, extending coverage to groups that would otherwise not be reached.

The UAS has an estimated recruitment rate of 15 to 20% which is comparable to or slightly higher than those of other probability-based Internet panels like the GfK KnowledgePanel or the RAND American Life Panel. Such low recruitment rates have led some researchers to argue that there is little practical difference between opting out of a probability sample and opting into a nonprobability (convenience) Internet panel (Rivers, 2013). In general, probability-based panels tend to better represent the underlying population in terms of demographic characteristics (Chang and Krosnick, 2009). Yet, nonprobability Internet panels may still be used as a basis of population norms as long as the data can be appropriately weighted to compensate for coverage errors and selection bias. Hays, Liu, and Kapteyn (2015) conclude that "No hard-and-fast rules determine when convenience panels are adequate for use in population inference or when response rates to probability Internet panels will be high enough to assume unbiased estimates." Our study does not try to address this issue: we will not compare the UAS – a probability-based Internet panel – to a convenience Internet panel. However, it is important to remember that the UAS shares low recruitment rates with other probability-based Internet panels. Because of this, weighting may become a crucial issue when the objective is to represent the underlying population. In view of this, the next section describes in some detail the weighting procedures used by the UAS.[2]

3.3. UAS Sampling and Weighting Procedures

An important feature of the UAS sampling procedure is that member recruitment is done in batches and that the first recruitment batch was sampled differently from subsequent batches.[3] The first batch was a simple random sample of addresses from the United States Postal Service (USPS) database. Subsequent batches were based on the sequential importance sampling (SIS) algorithm developed by Meijer (2014) and Angrisani et al. (2014).[4] This is a type of adaptive sampling (Groves & Heeringa, 2006; Tourangeau et al., 2017; Wagner et al., 2012) that generates unequal sampling probabilities with desirable statistical properties. Specifically, before sampling an additional batch, the SIS algorithm computes the unweighted distributions of specific demographic characteristics (e.g., sex, age, marital status and education) in the UAS at that point in time. It then assigns to each zip code a nonzero probability of being drawn, which is an increasing function of the degree of "desirability" of the zip code. The degree of desirability is a measure of how much, given its population characteristics, a zip code is expected to move the current distributions of demographics in the UAS toward those of the US population. For example, if at a particular point in time the UAS panel underrepresents females