Investigating the Bias Properties of Alternative Statistical Inference Methods in Mixed-Mode Surveys

by

Zeynep Tuba Suzer Gurtekin

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Survey Methodology)
in the University of Michigan
2013

Doctoral Committee:

Research Professor Richard L. Valliant, Co-Chair
Research Scientist Steven G. Heeringa, Co-Chair
Research Professor Mick P. Couper
Professor Trivellore Raghunathan
Assistant Research Scientist Sunghee Lee

Dedications

I would like to dedicate my dissertation to my family who has paved my way and supported me in every possible way to complete this phase of my life.

My dear grandparents, Bedia and Ali Rıza Belül, you put my future before anything else. I have no words to describe my love and gratitude for you. You are always in my thoughts.

My parents, Emine and Erkan Suzer, thank you for being my conscience and my biggest supporters. Growing up with no nonsense work ethics helps to succeed in the academia.

My sisters, Tüvana and Tuğberk, nephew and nieces, Sarp Kaya, Doğasu, Çağıl, and my brother-in-law, Kaan, you bring joy to my life. Being part of your successes has inspired me complete this thesis. I am looking forward to share many joyful events with you.

My husband, Levent, our promise for worse or better has been sealed. Your patience and support made this possible.

Acknowledgements

I would like to extend my gratitude to my dissertation advisors, Richard Valliant and Steve Heeringa, as well as my committee members, Mick Couper, Trivellore Raghunathan, and Sunghee Lee, for their support. In particular, I would like to thank my advisors, Rick and Steve, for their guidance and assistance in completing my dissertation work. I am very grateful that I was provided the flexibility to work on my own areas of research interest. Rick and Steve very patiently guided me in my research, proposal writing, conference papers and research prospects. They made it possible for me to receive a well-rounded academic experience and to become a productive researcher.

In addition to my dissertation work, I also had very interesting research opportunities during my tenure in the program. My research work with Rick provided me an opportunity to observe and practice how academic thinking could be applied to various practical problems. His inquisitive and supportive teaching and management style also served as an example for me. I am also grateful that I had an opportunity to work closely with James Lepkowski, who has been very influential in my decision to pursue a degree in survey methodology. I had stimulating opportunities to work on cutting-edge topics as a continuation of existing theory. The fact that he was so accessible provided a rich platform from which to learn about his editorial approach in scholarly writing, constructing experimental and sampling designs. I would like to thank Sunghee Lee for her support and the opportunity to work with her. Her solid research ethics have been inspirational in shaping my own research agenda.

The personal and professional touch of the program staff lessen the stress of the entire dissertation process. I would like to thank Jill Esau, Patsy Gregory, Sumi Raj, Jodi Holbrook, Nancy Oeffner, and Elisabeth Schneider for making us a part of a big family.

My colleagues and my friends gave invaluable feedback on my research proposals, manuscript drafts, and struggles. Here I would like to thank Niki Matusko, Mahmoud El Kasabi and Chan Zhang in particular for their continuous support and vivid discussion. I also would like to extend my gratitude to Christine Feak whose teaching has been the foundation of my dissertation writing.

I also would like to thank Arbitron and Surveys of Consumers for their financial support during my time in the program. Partial funding for this research was also provided by the DDIG - Doctoral Dissertation Improvement Grants (NSF) (Award Number SES 1238612).

I also would like to thank my family who has been my continuous inspiration and source of strength in my whole life.

Table of Contents

Dedications ii

Acknowledgements iii

List of Figures viii

List of Tables xii

Abstract xiv

Chapter 1 Introduction 1

1.1. Mixed-Mode Surveys 1

1.2. Mode Effects 6

1.3. Mode Choice and Mode Effects 9

1.4. Statistical Error Models and Mixed-Mode Survey Inference 14

1.5. Proposed Mixed-Mode Survey Inference Method 14

1.6. Empirical and Simulation Evaluations 15

Chapter 2 Statistical Error Models for Mixed-Mode Survey Estimators 17

2.1. Mode Effects and Statistical Error Models 17

2.2. Statistical Error Models 21

2.2.1. General model for a single-mode survey estimate 22

2.2.2. General model for a single-frame mixed-mode survey estimate 24

2.3. Existing Mode Effect Evaluation Methods 27

2.3.1. Regression Model Methods 28

2.3.2. Mixture Distribution Method 30

2.3.3. Propensity Score Matching 37

2.4. Existing Mode Effect Adjustment Methods 38

2.4.1. Response Mode in Calibration Estimator 38

Chapter 3 Proposed Mixed-Mode Survey Inference Methods 43

3.1. Bias Properties of Mixed-Mode Survey Mean Estimator 43

3.2. Evaluation of Mode Effects in Mixed-Mode Surveys 44

3.3. Multiple Imputation Methods 46

3.3.1. Response and Choice Models in Mixed-Mode Surveys 46

3.3.2. Expectations of Response and Choice Models in Mixed-Mode Surveys ... 47

3.3.2.1. Mean Estimator Ignoring Mode Effect 47
3.3.2.2. Proposed Mean Estimator 48
3.3.2.3. Estimation Errors for Alternative Estimators............................ 50
3.3.3. Imputation models 56
3.3.3.1. Ignorable Mode Choice Imputation Models 56
3.3.3.2. Nonignorable Mode Choice Imputation Models....................... 58
3.3.4. Special Cases of α : Alternative Empirical Combination Methods 63
Chapter 4 Modeling of Mode Choice and Mode Effects in the 1973 Current Population Survey (CPS) Match and 2012 CPS March Data 65
4.1. Introduction.. 65
4.2. Data Description 66
4.2.1. 1973 CPS Match Data.. 66
4.2.2. 2012 CPS March Data 78
4.2.3. Descriptive Statistics and Comparisons of 1973 and 2012......................... 78
4.3. Regression Models for Mode Choice and Measurement Discrepancy in the 1973 CPS Match Data 81
4.3.1. Model of Mode Choice 81
4.3.2. Model of Response Differences.. 83
4.3.3. Model for Total Family Income based on 1973 CPS Match Data.............. 88
4.4. Regression Models for CPS March 2012 Data 89
4.4.1. Models for Mode Choice based on CPS 2012 data 89
4.4.2. Models for Personal Income based on CPS 2012 data 91
4.4.3. Models for Health Insurance Coverage based on CPS 2012 data 95
Chapter 5 Empirical Evaluations of Mixed-Mode Survey Inference Methods 99
5.1. Introduction.. 99
5.2. Current Population Survey, 1973, and Social Security Records: Exact Match Data 100
5.3. Evaluation of Proposed Mixed-Mode Methods of Survey Inference 103
5.4. Results... 106
5.5. Discussion and Extensions.. 118
Chapter 6 Simulation Evaluations of Mixed-Mode Survey Inference Methods 120

6.1. Introduction 120
6.2. Simulation Study 120
6.2.1. Motivation 120
6.2.2. Creation of Hypothetical Populations: Models used to Generate the Hypothetical Populations 121
6.2.3. Hypothetical Populations Characteristics 124
6.2.4. Estimation Methods 126
6.2.5. Simulation Study Results 127
6.3. Empirical Comparison Analysis: Application of the methods to CPS March 2012 Data 130
6.3.1. Motivation 130
6.3.2. Creation of Sample Replicates 131
6.3.3. ANOVA for Repeated Measurements 131
6.3.4. Estimation Methods 132
6.3.5. Empirical Comparison Analysis Results 133
6.3.5.1. Personal Income 133
6.3.5.2. Health Insurance Coverage 136
6.4. Discussion 138
Chapter 7 Conclusions 140
7.1. Theoretical Framework 140
7.2. Bias of Alternative Estimators 141
7.3. Empirical and Simulation Studies 141
7.4. Limitations 142
7.5. Future Research 143
References 144

List of Figures

Figure 1.1 – Percentage of Householders by Interview Mode by Month in Sample, Current Population Survey (CPS), March 2012 4

Figure 1.2 – Unweighted Respondent Household Characteristics by Interview Mode, Current Population Survey (CPS), March 2012 10

Figure 1.3 – Unweighted Respondent Householder Characteristics by Interview Mode, Current Population Survey (CPS), March 2012 11

Figure 1.4 – Schematic Chart for the Proposed Mixed-Mode Survey Inference Method 15

Figure 2.1 – Survey Life Cycle for Mixed-mode Surveys from a Quality Perspective (adapted from Groves et al., 2009) 19

Figure 4.1 – Family Income Construct Correlations. CPS Imputations for missing items are excluded. Incomes are top-coded at 50,000 USD. The CPS constructed total family income measure includes income from all family members age 15 and over....... 70

Figure 4.2 – Scatterplot of Modified Total Family Income versus IRS AGI (Form 1040) in 1973 CPS Match Data 72

Figure 4.3 – Boxplots of Differences in Modified Total Family Income and IRS AGI by Response Mode in 1973 CPS Match Data.. 73

Figure 4.4 – Distribution Differences in Householder Characteristics between 1973 CPS Match Data and 2012 Weighted CPS Data... 80

Figure 4.5 – Marginal Mean Absolute Difference for Welfare Receipt Status Controlling for all the Covariates in the Final Model (Table 4.6) by Response Mode (Personal=1: In-person, Personal=0: Telephone). The corresponding bars represent 95% confidence intervals. 85

Figure 4.6 – Marginal Mean Absolute Difference for Age Controlling for all the Covariates in the Final Model (Table 4.6) by Response Mode (Personal=1: In-person, Personal=0: Telephone). The corresponding bars represent 95% confidence intervals. .. 86

Figure 4.7 – Marginal Mean Absolute Difference for Income Type Controlling for all the Covariates in the Final Model (Table 4.6) by Response Mode (Personal=1: In-person, Personal=0: Telephone). The corresponding bars represent 95% confidence intervals.. 87

Figure 5.1 – Under the ignorable mode choice imputation model relative differences ($\text{RelDiff}_{CM} = \left(\bar{y}_{CM_i} - \bar{y}_{IRS} / \bar{y}_{IRS}\right)$, where CM_i for $i = 1, 2, 3, 4$ is the combination method) in 50 samples of estimates of wage and salary income mean with the four alternative methods of estimation (CM1, CM2, CM3, and CM4=Standard Combination Method) and mode-specific imputed data (Telephone and In-person) by item missing treatment procedure. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean relative difference. 110

Figure 5.2 – Under the ignorable mode choice imputation model absolute relative differences ($\text{AbsRelDiff}_{CM} = \left(\left|\bar{y}_{CM_i} - \bar{y}_{IRS}\right| / \bar{y}_{IRS}\right)$, where CM_i for $i = 1, 2, 3, 4$ is the combination method) in 50 samples of estimates of wage and salary income mean with the four alternative methods of estimation (CM1, CM2, CM3, and CM4=Standard Combination Method) and mode-specific imputed data (Telephone and In-person) by item missing treatment procedure. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean absolute relative difference. 111

Figure 5.3 – Under the nonignorable mode choice imputation model relative differences ($\text{RelDif}_{CM} = \left(\bar{y}_{CM_i} - \bar{y}_{IRS} / \bar{y}_{IRS}\right)$, where CM_i for $i = 1, 2, 3, 4$ is the combination method) in 50 samples of estimates of wage and salary income mean with the four alternative methods of estimation (CM1, CM2, CM3, and CM4=Standard Combination Method) and mode-specific imputed data (Telephone and In-person) by item missing treatment procedure. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean relative difference. 112

Figure 5.4 – Under the nonignorable mode choice regression model absolute relative differences ($\text{AbsRelBias}_{CM} = \left(\left|\bar{y}_{CM_i} - \bar{y}_{IRS}\right| / \bar{y}_{IRS}\right)$, where CM_i for $i = 1, 2, 3, 4$ is the combination method) in 50 samples of estimates of wage and salary income mean with the four alternative methods of estimation (CM1, CM2, CM3, and CM4=Standard Combination Method) and mode-specific imputed data (Telephone and In-person) by

item missing treatment procedure. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean absolute relative difference. 113

Figure 5.5 – Under the ignorable mode choice imputation model relative differences ($\text{RelDif}=\bar{y}-\bar{y}_{IRS}/\bar{y}_{IRS}$) in 50 samples of estimates of wage and salary income mean with the CM4 (Standard Combination Method) and mode-specific completed (Telephone and In-person) data by item missing treatment procedure. Telephone.O and Inperson.O data represent relative differences for mean wage and salary income based on the observed data. Telephone.I and Inperson.I data represent relative differences for mean wage and salary income based on the imputed data. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean relative difference. 114

Figure 5.6 – Under the ignorable mode choice imputation model absolute relative differences ($\text{AbsRelDif}=|\bar{y}-\bar{y}_{IRS}|/\bar{y}_{IRS}$) in 50 samples of estimates of wage and salary income mean with the CM4 (Standard Combination Method) and mode-specific completed (Telephone and In-person) data by item missing treatment procedure. Telephone.O and Inperson.O data represent absolute relative differences for mean wage and salary income based on the observed data. Telephone.I and Inperson.I data represent absolute relative differences for mean wage and salary income based on the imputed data. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean absolute relative difference. 115

Figure 5.7 – Under the nonignorable mode choice imputation model relative differences ($\text{RelDif}=\bar{y}-\bar{y}_{IRS}/\bar{y}_{IRS}$) in 50 samples of estimates of wage and salary income mean with the CM4 (Standard Combination Method) and mode-specific completed (Telephone and In-person) data by item missing treatment procedure. Telephone.O and Inperson.O data represent relative differences for mean wage and salary income based on the observed data. Telephone.I and Inperson.I data represent relative differences for mean wage and salary income based on the imputed data. Sample sizes are 400 and 800 each

for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean relative difference. 116

Figure 5.8 – Under the ignorable mode choice imputation model absolute relative differences ($\text{AbsRelDif}=|\bar{y}-\bar{y}_{IRS}|/\bar{y}_{IRS}$) in 50 samples of estimates of wage and salary income mean with the CM4 (Standard Combination Method) and mode-specific completed (Telephone and In-person) data by item missing treatment procedure. Telephone.O and Inperson.O data represent absolute relative differences for mean wage and salary income based on the observed data. Telephone.I and Inperson.I data represent absolute relative differences for mean wage and salary income based on the imputed data. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean absolute relative difference. 117

Figure 6.1 – Density Plots of Survey Response Data for Telephone and In-person Hypothetical Populations, and AGI data. beta=(0.97, 0.7) for telephone and in-person, respectively, in $Y^{*}_{pj}=\beta^{(AGI)}_{p}Y^{(AGI)}_{j}$, and $c\in(1,0.5)$ in $\hat{Y}_j=X^{(Y)}_j\hat{\beta}^{(Y)}+e_j$, where $e_j\overset{iid}{\sim}N(0,\sigma^2/c)$ 125

Figure 6.2 – Telephone and In-person Hypothetical Population Income Means for Varying Beta and Constant by Response Mode. The vertical axis shows combinations of $\beta^{(AGI)}_{I}$ and $c\in(0.5,10,15)$. 126

Figure 6.3 – Relative Biases for Alternative Inference Methods under Ignorable Mode Choice Imputation Model 128

Figure 6.4 – Relative Biases for Alternative Inference Methods under Nonignorable Mode Choice Imputation Model 129

Figure 6.5 – 95% Confidence Intervals for Personal Income by Method under the Ignorable Mode Choice Imputation Model... 135

Figure 6.6 – 95% Confidence Intervals for Personal Income by Method under the Nonignorable Mode Choice Imputation Model 136

List of Tables

Table 1.1 – Reporting Patterns for a Three Phase Sequential Mixed-mode Survey Design (R : Reporters (subscripts $P = M, T, I$ correspond to Mail, Telephone, In-person modes, NR_M : Nonreporting units by mail, NR_T : Nonreporting units by telephone, NR_I : Nonreporting units by in-person mode) ... 3

Table 1.3 – Reporting Patterns for a CPS-like Mixed-mode Rotating Panel Survey Design (R : Reporters (subscripts $P = T, I$ correspond to Telephone, In-person modes, NR_T : Nonreporting units by telephone, NR_I : Nonreporting units by in-person mode) 5

Table 2.1 – Selection and Mode Effects as defined by Vannieuwenhuyze et al. (2010, 2012) ... 32

Table 4.1 – Householder Covariate Percentages in the Special 1973 CPS Match File and 2012 CPS Data (*1973 and 2012 categories are not comparable.) ... 75

Table 4.2 – Household Covariate Frequencies and Percentages in the 1973 CPS Match Data ... 77

Table 4.3 – Response Mode Distribution by Month in Sample ... 82

Table 4.4 – Mode Choice Logistic Model Type III Tests, 1973 CPS Match Data 82

Table 4.5 – Mode Choice Logistic Model including Interactions Type III Tests, 1973 CPS Match Data ... 83

Table 4.6 – Type III tests for the Factors in the Linear Regression Model of Absolute Difference, 1973 CPS Match Data ... 84

Table 4.7 – Response Mode % by Age x Work Status ... 89

Table 4.8 – Mode Choice Logistic Model Type III Tests, CPS March 2012 Respondents ... 90

Table 4.9 – Covariates included in the Response Imputation Models (Y-Model) 92

Table 4.10 – Beta Coefficient Estimates and Standard Errors for the Response Imputation Model for the natural logarithm of Personal Income ... 92

Table 4.11 – Covariates included in the Response Imputation Models (Y-Model) ... 97

Table 4.12– Unadjusted Means for Variables of Interest in three studies ... 98

Table 5.1 – Response Mode Distribution by Month in Sample ... 102

Table 5.2 – Sample Quintiles of Reported Wage and Salary Income 103
Table 5.3 – Sample Percentage of Item Missing in Reported Wage and Salary Income.. 103
Table 5.4 – Covariates included in the Mode Choice Imputation Models (R-Model).. 104
Table 5.5 – Covariates included in the Response Imputation Models (Y-Model) 105
Table 5.6 – Number of Significant Differences at 95% confidence level between Telephone and In-person Mode-specific Estimates... 109
Table 6.1 – Covariates included in the Response Models (Y-Model)................ 123
Table 6.2 – Covariates included in the Mode Choice Models (R-Model).......... 124
Table 6.3 – F statistics for Repeated Measurement ANOVA for Personal Income (n_b = 200) 134
Table 6.4 – Mean and Standard Errors for Personal Income by Method and Imputation Model (n_b = 200) 134
Table 6.5 – 95% Confidence Intervals for Personal Income by Method and Imputation Model.. 134
Table 6.6 – F statistics for Repeated Measurement ANOVA for Health Insurance Coverage (n_b = 200)... 138
Table 6.7 – Means and Standard Errors for Health Insurance Coverage by Method and Imputation Model (n_b = 200) 138

Abstract

Early in the history of survey research, mixed-mode surveys were proposed to decrease non-observational survey errors under certain survey budgets (Hansen & Hurwitz, 1946; Hochstim, 1967). Recently, pressing issues of increasing non-observational survey error and survey costs influenced survey researchers to adapt many variations of mixed-mode surveys (Brick & Lepkowski, 2008; Couper, 2011; De Leeuw, 2005). The statistical inference in the earlier studies implicitly assumed ignorable mode effects; that is, all survey modes generate values close to true values for all the members of the population. Later, theoretical frameworks were developed to discuss the factors that may yield nonignorable mode effects for different subgroups in the population (De Leeuw, 1992, 2005; Groves et al., 2009; Schwarz, Strack, Hippler, & Bishop, 1991; Tourangeau, Rips, & Rasinski, 2000a). But empirical work could only study parts of the frameworks and was conditional on specific survey designs. With a few exceptions (Aquilino, 1994; Beland & St-Pierre, 2008; Soulakova, Hartman, Gibson, & Davis, 2009), the focus of the empirical work was mostly on estimates of full population quantities. The theory and the empirical results emphasized the possible differences between the self- and interviewer-administered surveys, the audio and visual channel dependent presentations and the variable dependent nature of mode effects. Inference in later mixed-mode survey designs, generally adopted the early assumption that mode effects could be ignored and did not challenge that assumption with any empirical work. In sequential or concurrent mixed-mode survey inference, in which data are collected via multiple response modes, responses from multiple modes have been combined without adjusting for any measurement error.

In practice, survey modes are not randomly assigned in mixed-mode surveys. This nonrandom assignment establishes a challenge to evaluate mode effects directly in mixed-mode surveys. This dissertation defines this

nonrandom assignment as mode choice. Recent methods have been developed to unconfound the mode choice and the mode effects (Camillo & D'Attoma, 2011; Jäckle, Roberts, & Lynn, 2010; Lugtig, Lensvelt-Mulders, Frerichs, & Greven, 2011; Vannieuwenhuyze, Loosveldt, & Molenberghs, 2010, 2012). These methods challenge the general notion of ignorable mode effects in the mixed-mode surveys and motivate a more systematic approach to evaluate mode effects. Buelens and Van den Brakel (2011) also propose a mode calibrated method for estimating changes in means over time.

This dissertation also proposes an alternate method that evaluates and adjusts for mode effects. The respondent data for a given mode and phase are used to create completed data sets for a given sample. Then, the completed data sets are used to compute mode-specific survey means. The survey means are then combined to produce one survey estimate. The ways in which the mean estimates can be combined are (1) a simple average, (2) a minimum variance combination, and (3) a minimum mean square error combination. The last of these requires some measure of true values that are unaffected by mode effects. The dissertation includes conceptual work and empirical/simulation evaluation of inference methods. The conceptual work includes extension of a single survey mode statistical error model to a mixed-mode survey context. The bias properties of the standard method of estimation, which ignores mode effects, and proposed methods, which adjust for mode effects under a simple measurement model, are investigated.

The dissertation work includes three studies. Two studies use a special type of data that include hypothetical true values at the person level. The data, 1973 Current Population (CPS) Match Data, include both survey and Internal Revenue Service (IRS) data. The first empirical study focuses on a variable of interest, wage and salary income, for which measurement complexities are minimal. The following simulation study augments the data to include cases with more complicated measurement properties. Varying degrees of mode effects were simulated based on the observed data to evaluate the proposed methods under more complicated situations. Since both studies include benchmark values, which may not be the usual case, a third study conducts an empirical comparison analysis for both personal income and health insurance coverage for which no benchmark values are available.

In the first empirical study, the variable of interest is the wage and salary income, which is a continuous variable. The corresponding person level data allowed computation of relative differences for the standard method, the alternative combination methods and the mode–specific estimates relative to a benchmark. The first set of empirical evaluations focused on a set in which mode effects are ignorable and variances of the mode-specific estimates are equal. Ignorable mode effects and equal variance properties yielded a special case of the combination weights in alternative (3) above that minimizes the mean square error of combined estimator. As a result, performance differences were not significant between the alternative combination methods. On the other hand, they all outperformed the standard method as expected.

In the second set of simulation studies, hypothetical populations were created by varying the severity of mode effects and the model fit as defined by the error variance of the underlying regression model. The variable of interest is total family income which includes other components of income in addition to wages and salaries. Results were again in the expected direction. As the severity of the mode effects increased, relative to the population values, differences for the alternative estimators were smaller compared to the relative differences of the standard method. More importantly, a poor model fit diluted potential improvements of the alternative methods. Two imputation models were tested — one in which mode choice was ignorable and another in which mode choice was nonignorable. The performance of these models depended on how well the imputation models fit the data. The nonignorable mode choice model that assumes that mode choice also depends on the variable of interest distribution and helped to compensate for the poor model fit.

In the third set of empirical evaluations, 2012 CPS March Data were used. In this dataset benchmark values were not available. Both personal income and health insurance coverage are variables of interest. Sensitivity results showed that applying the proposed imputation method may yield differences in the mean personal income and health insurance coverage. Although the sensitivity analysis cannot address the source of the differences, it addresses the further need to investigate the mode effects systematically.

Given the special subset of 1973 CPS Match data, the first study addressed two research questions in particular: (1) can mode effects for wage and salary income be ignored for estimation? and (2) can improved estimators be developed that account for the possibility that modes might have different biases? Related to the first question, under the ignorable mode choice imputation model, relative and absolute relative differences for in-person mode-specific means were on average greater than those for the telephone mode-specific means. The difference in relative and absolute relative difference between in-person and telephone mode specific estimates was eliminated under the nonignorable mode choice imputation model. In principle, telephone and in-person mode effects should be studied separately under randomized experimental conditions in which the true values of the measured construct are known. Tourangeau, Rips, & Rasinski (2000b) discuss such designs to assess the accuracy of survey reports. A mixed-mode survey data structure does not provide such conditions. Instead only average differences in mode effects can be evaluated under measurement error and imputation models. Related to the second research question, smaller relative and absolute relative differences suggested that improved estimators can be developed that account for the possibility that modes might have different biases.

The second study creates a situation in which mode effects and goodness of model fit are controlled explicitly. Results show that a better performing estimator in terms of relative bias is possible, i.e. estimators that weight mode-specific means by the inverse of variances and inverse of mean square errors can outperform the standard estimator that ignores mode effects. But weighting by the inverse of variances may yield greater relative bias in a case in which lower quality data has lower variance. Also the estimator that weights by inverse of mean square errors is not generally feasible. Therefore more research needs to be done to test the properties of the empirically optimal estimator. Furthermore, results show that modeling assumptions are crucial and models need to be guided by theoretical frameworks.

The third study is an empirical comparison study that tests the differences in mode-specific and combined estimates. The results show a sensitivity to modeling assumptions. The significant differences in mode-specific means for both personal income and health insurance coverage motivate future research to investigate mode effects for these two variables of interest.

Chapter 1 Introduction

1.1. Mixed-Mode Surveys

Mixed-mode designs have been widely used by large government and scientific surveys in the last five decades. There are many possible design variations to meet multiple methodological goals such as decreasing nonresponse and coverage bias, reducing survey costs, and improving timeliness and measurement quality.

This study focuses on one specific mixed-mode design in which multiple response modes are used to decrease nonresponse under certain budget constraints. There are two general mixed-mode survey designs: (1) sequential mixed-mode surveys an example of which is the American Community Survey (ACS) and (2) panel surveys that use a mix of survey modes such as the Current Population Survey (CPS). In this dissertation, the second design is considered in the illustrations, empirical investigations and simulation studies.

Traditionally in combining data from multiple response modes, statistical inference methods assume that mode effects are ignorable. With that assumption, data obtained using different modes are combined for estimation with no special adjustments made for the possibility that one mode may, in some sense, yield more accurate observations than another. This dissertation discusses the implications of this assumption conceptually and evaluates alternative inference methods that adjust for estimated mode effects in combining data from multiple response modes.

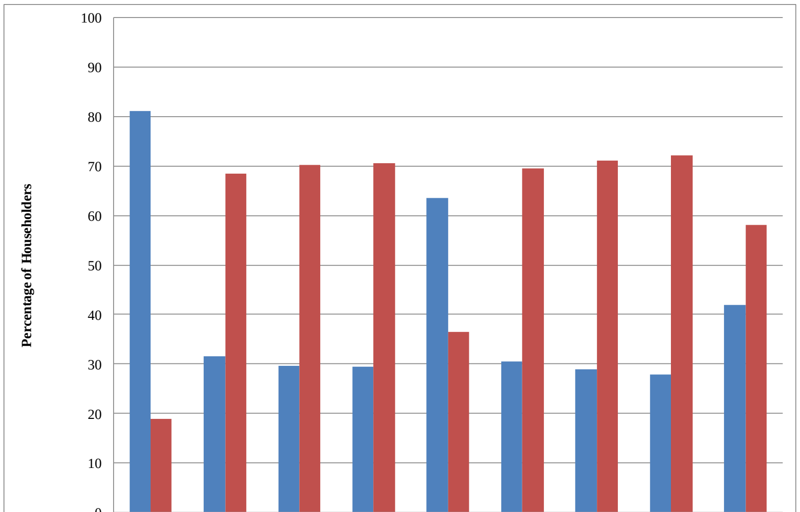
For example, in the U.S. one of the most prominent surveys, the American Community Survey (ACS), conducted by the U.S. Census Bureau, uses a sequential mixed-mode survey design to minimize data collection costs and decrease survey nonresponse. The ACS collects critical socioeconomic data to help communities determine where to locate services and allocate resources (Davis & Alexander, Jr., 1997). For each of the monthly ACS samples, three phases of data collection are conducted over a 3-month period. In the first phase, a mail survey collects responses from households residing at a probability sample of housing unit addresses. The mail survey nonrespondent and unmailable postal addresses for which telephone numbers are available are then followed up by computer-assisted telephone interviewing (CATI) in the second month. In the third month, computer-assisted personal interviewing (CAPI) responses are collected from a subsample of the housing units that have not yet responded or been contacted. Mail returns continue to be accepted during this entire period. Beginning in late 2012, the Census Bureau also announced that, the ACS will offer an Internet response option in addition to mail response option starting at the first contact (The Census Bureau, 2012).

Table 1.1 illustrates a simplified data structure for an ACS-like sequential mixed-mode survey. The design is based on a probability sample of households that is selected from a list frame that ideally contains both housing unit addresses and telephone numbers for all units. In most applications, as in the ACS, the telephone numbers can be available for only a subset of the addresses on the frame. For example, telephone numbers are not available for 60% of the nonresponding households to mail contact in the ACS (Diffendal, 2001). For this subgroup, only mail and in-person nonresponse follow-up phases are applicable. While the sample and, correspondingly, the interview follow-up unit is a household, the unit of analysis for this project is the person. The sequential mixed-mode data collection starts with a mail survey contact and nonrespondents at this initial mail phase are followed up in subsequent "phases" by telephone and in-person contacts. The columns in Table 1.1 capture the mail, the telephone and the in-person survey data decomposition by reporting status. This conceptualization considers univariate vectors of data for one variable in particular. Each data vector denoted by Y_M , Y_T and Y_I corresponds to one phase—mail, telephone or in-person.

Table 1.1 – Reporting Patterns for a Three Phase Sequential Mixed-mode Survey Design (R : Reporters (subscripts $P = M, T, I$ correspond to Mail, Telephone, In-person modes, NR_M : Nonreporting units by mail, NR_T : Nonreporting units by telephone, NR_I : Nonreporting units by in-person mode)

Y_M (Mail)	Y_T (Telephone)	Y_I (In-person)
R_M	R_M	R_M
NR_M	R_T	R_T
	NR_T	R_I
		NR_I

Additionally, panel surveys that offer multiple response options in subsequent waves are one of the many possible mixed-mode survey designs. To reduce survey costs and respondent burden, panel surveys may offer alternative survey response options such as telephone, and web in the waves after the first wave (De Leeuw, 2005). For example, the Current Population Survey (CPS), a monthly rotating panel survey, implements a mixed-mode survey design. The CPS rotating panel design employs a 4-8-4 cycle for a selected household. Interviews are conducted for two sets of four consecutive waves that are eight months apart. A majority of the CPS interviews are conducted by telephone, except for the first and fifth wave interviews. For these two waves that begin a sequence of four months of interviews in the 4-8-4 cycle, the dominant mode is in-person, as shown in Figure 1.1. Table 1.2 illustrates a data structure for a CPS-like mixed-mode panel survey. As shown in Table 1.2, for a given mode each phase is composed of reporting and nonreporting units, which are reporting units for the alternative mode. This data structure includes month in sample decomposition because of its rotating panel survey nature. In this dissertation this data structure is conceptualized as an example of a mixed-mode survey design in which multiple modes are available for a given phase. Month in sample is a sampling design factor that denotes the rotating panel for a given survey period. The data from multiple rotating panels are considered to compose a cross-sectional data. Month in sample is incorporated into the modeling of mode-specific Y vectors in the empirical and the simulation studies described later in this dissertation.



	1	2	3	4	5	6	7	8	Overall
In-person %	81.1	31.5	29.7	29.4	63.5	30.5	28.9	27.9	41.9
Telephone %	18.9	68.5	70.3	70.6	36.5	69.5	71.1	72.1	58.1

Figure 1.1 – Percentage of Householders by Interview Mode by Month in Sample, Current Population Survey (CPS), March 2012

Table 1.2 – Reporting Patterns for a CPS-like Mixed-mode Rotating Panel Survey Design (R : Reporters (subscripts $P = T, I$ correspond to Telephone, In-person modes, NR_T : Nonreporting units by telephone, NR_I : Nonreporting units by in-person mode)

Month in Sample	Y_T (Telephone)	Y_I (In-person)
1	R_T	NR_I
	NR_T	R_I
2	R_T	NR_I
	NR_T	R_I
3	R_T	NR_I
	NR_T	R_I
4	R_T	NR_I
	NR_T	R_I
5	R_T	NR_I
	NR_T	R_I
6	R_T	NR_I
	NR_T	R_I
7	R_T	NR_I
	NR_T	R_I
8	R_T	NR_I
	NR_T	R_I

In both designs, as shown in Table 1.1 and Table 1.2, considering an underlying "true" distribution of Y and the corresponding parameters, mail, telephone and in-person interviews may produce response distributions that differ from both the "true" distribution of Y or the distribution of Y that would be observed if all the units have responded by one mode. In particular, different survey modes may produce different distribution parameters. For the purpose of this dissertation, the population parameter of interest is the mean $\bar{Y}$.

For example, personal income is a sensitive topic in most surveys conducted in the U.S. and elsewhere (Moore, Stinson, & Welniak, Jr., 2000). Given that personal income is a sensitive topic, mail, telephone and in-person modes may impose different social desirability bias conditions for personal income measurement in the surveys (Aquilino, 1994; Tourangeau & Smith, 1996; Tourangeau & Yan, 2007). In addition, in the in-person mode, it will be harder for the respondent to seriously misreport their income as the interviewer can observe some of the income associated wealth indicators. Also, different income structures (e.g. business owners, self-employed, investors) may imply differences in the complexity of the measurement (Körmendi, 1988; Moore et al., 2000). The complexity of the measurement is handled differently by different survey modes, which may yield differences in the response distributions by mode.

Despite possible differences in measuring income by different data collection modes, most current estimation practice in mixed-mode surveys, including the CPS, ignores the mode of data collection when the data are combined for estimation. For example, observations on Y obtained from the respondent sets, R_M, R_T, and R_I in Table 1.1 and R_T, and R_I in Table 1.2 are combined without adjusting for possible mode effects. This dissertation proposes to develop a mixed-mode survey estimation procedure that adjusts for mode measurement effects and, to the extent possible, produces estimators that are more nearly unbiased than methods that ignore the potential for mode effects.

1.2. Mode Effects

According to the Total Survey Error (TSE) framework, the term *mode effects* specifically refers to measurement error differences due to mode of survey administration, although this definition may take on different scope and meanings depending on the context (Groves et al., 2009). In the taxonomy of survey errors, coverage error, nonresponse error and sampling error are classified as *non-observational errors*. Measurement errors are classified as *observational errors* (Groves, 1989). Measurement error sources include: the respondent, the instrument, interviewers, and data collection modes. This taxonomy of survey errors omits processing error deliberately as the sources of processing error are different from the sources of measurement error. For the purpose of this study, the effects of processing error, if any, are considered to be a part of the mode effects, and there will be no attempt to disentangle the effects of measurement and processing errors.

Researchers discuss the factors related to mode effects under various frameworks (De Leeuw, 1992, 2005; Groves et al., 2009; Schwarz et al., 1991; Tourangeau et al., 2000a). These frameworks have their roots in communication, social and cognitive psychology theories but they have not been completely tested or incorporated into the statistical models that assess mode effects. These frameworks' scope is also somewhat limited in conceptualizing all the possible interaction effects of the features of data collection mode and the other sources of measurement error. Tucker and Lepkowski (2008) reemphasize that understanding the nature of mode effects most likely relies on the interactions between the mode, the interviewer, the respondent, and the instrument. In this dissertation, the comparisons of measurement errors include controls for mode and respondent interactions. The effects of the common survey instrument and question wording are also not isolated in this study. However, we assume that the questions for each mode are expected to be tested for validity. We assume that the effect of interviewers, if any, is the same for the telephone and the in-person modes (i.e. no interviewer by mode interaction). This is equivalent to treating interviewers as if they have been randomized to sample cases. These assumptions are incorporated in a simple measurement error model as presented in Chapters 2 and 3. A simple measurement error model as presented in equation (1.1) is explicitly assumed in deriving the expectations of estimation errors. Measurement error models are reviewed in detail by Biemer and Stokes (1991), Groves (1989,1991,1999), and Lessler and Kalsbeek (1992).

Factual versus attitudinal questions is another distinction that imposes different restrictions in the measurement error models. We assume that methods proposed here will apply to both factual and attitudinal questions, although only the former type of questions will be included in the analysis.

We will focus on two variables: (1) personal income, and (2) health insurance coverage. Both the income and insurance coverage constructs are measured by various Census Bureau surveys due to different levels of needs. The taxonomy of surveys and differences in methodologies are listed on http://www.census.gov/hhes/www/income/about/index.html and http://www.census.gov/hhes/www/hlthins/about/index.html for income and health insurance coverage, respectively.

Personal income and health insurance coverage data are important in many economic and health-policy analyses (Boudreaux, Ziegenfuss, Graven, Davern, & Blewett, 2011; Moore et al., 2000). Underreporting seems to be the dominant error type for both concepts (Boudreaux et al., 2011; Moore et al., 2000). However, the error sources for each concept differ. In their review, Moore, Stinson, and Welniak, Jr. (2000) focus on two sources of survey errors in income estimation: (1) nonresponse, and (2) measurement. These two components are also expected to vary by mode and subgroup (Greenlees, Reece, & Zieschang, 1982). In particular, social desirability bias on income measures is expected to vary by mode (Aquilino, 1994; Holbrook, Green, & Krosnick, 2003). Moreover, the physical presence of interviewers is also expected to impact item nonresponse and measurement errors. For example, in the ACS the item nonresponse data rates are on the higher end for the income and health insurance coverage questions compared to the item nonresponse rates in the other questions. Furthermore, compared to the telephone and in-person responses, the mail mode has higher item nonresponse data rates for income and health insurance coverage questions.

Despite a possible association with mode and item nonresponse data rates, the ACS and the CPS hot-deck imputation models do not incorporate indicator variables for the response mode. As will be stated later, our goal is to create a completed data set for each mode phase (e.g. mail, telephone, in-person) as if every person had responded using the same mode. Since the ACS and the CPS imputations ignore interview mode, using ACS- and CPS-imputed values in our models to create a completed data set (i.e. imputing the counterfactual measure for the nonobserved modes under the same model) may obscure the mode effect. To avoid this possibility, we use only actual reported data in the modeling for each phase. Thus, any reporter in a phase who has an imputed value for income or health insurance will be excluded or treated as a nonreporter for the purposes of modeling. In Chapter 5, one of the simulations includes imputed values for cases with item nonresponse data for the wage and salary income. A second simulation excludes any cases where the original CPS responses were missing. In Chapter 6 item nonrespondents on the dependent variables are excluded.

1.3. Mode Choice and Mode Effects

Mixed-mode survey designs such as the ACS and the CPS yield nonrandom mixes of survey modes conditioned on the survey design. When a choice is available, different types of respondents may choose different survey response modes. For example, this difference can be seen in the distributions of the selected household and householder characteristics by interview mode, which except for gender and metropolitan status, are significantly different ($p <= .0001$) in 2012 CPS March (see Figure 1.2 and Figure 1.3). For example, after controlling for other household and householder characteristics measured in the 2012 CPS March, respondents with higher education were more likely to respond in the telephone mode compared to respondents who have less than a 12th Grade education.

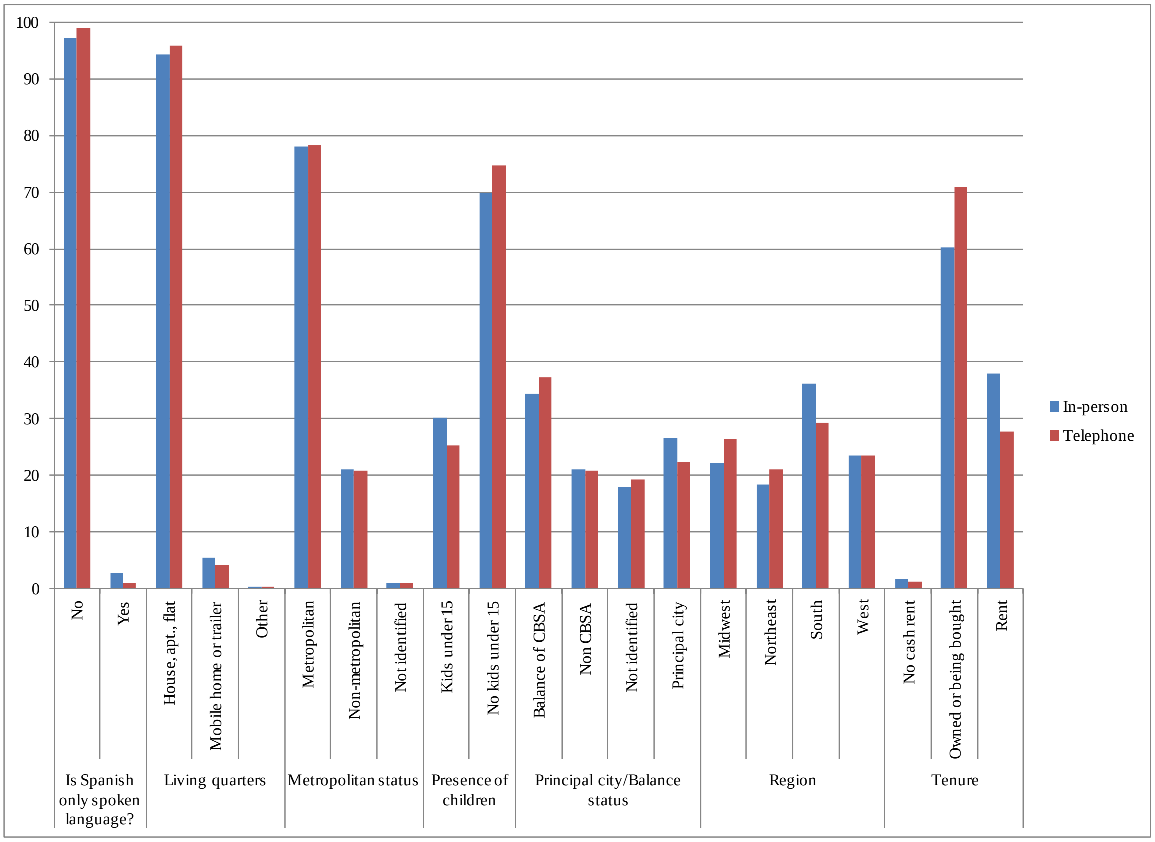


Figure 1.2 – Unweighted Respondent Household Characteristics by Interview Mode, Current Population Survey (CPS), March 2012

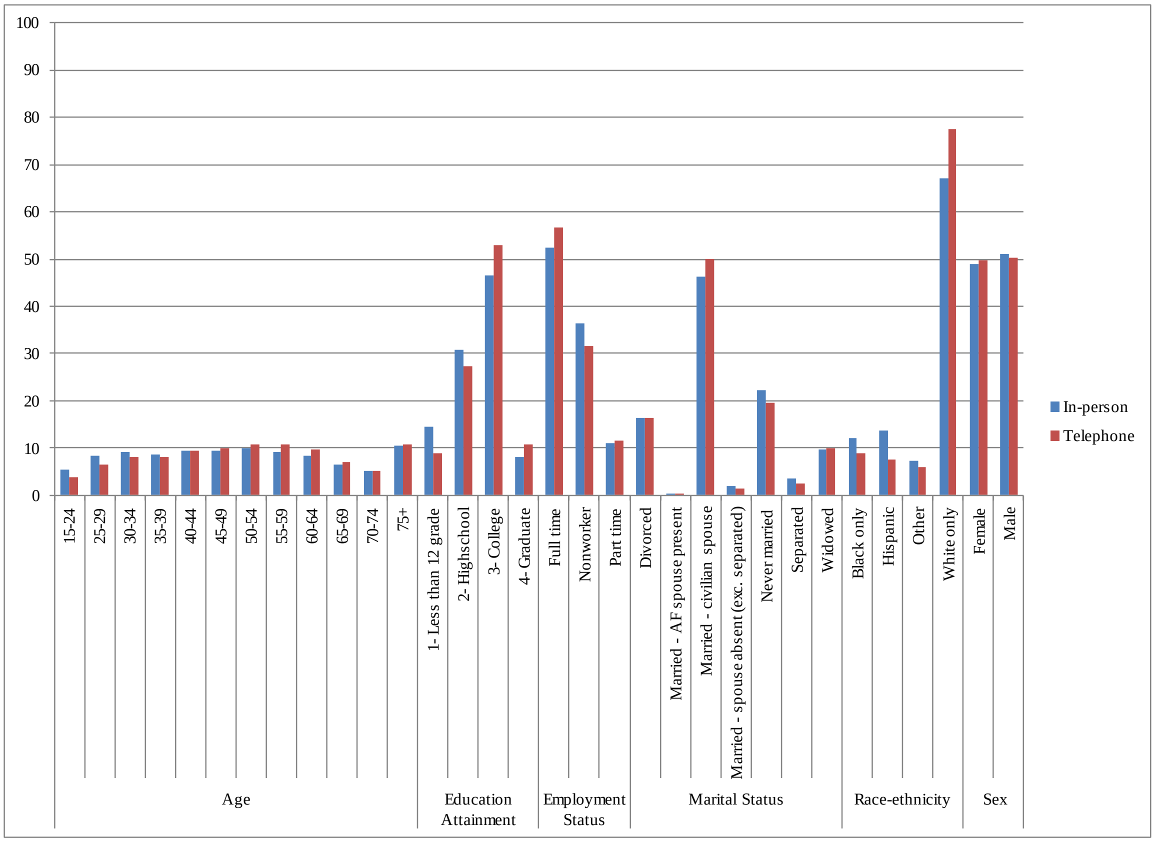


Figure 1.3 – Unweighted Respondent Householder Characteristics by Interview Mode, Current Population Survey (CPS), March 2012

In the remainder of this dissertation, this nonrandom assignment of survey response mode is referred to as mode choice. The term *selection effects* has been previously used in the literature to describe the effect of nonrandom assignment of survey response mode on the parameter of interest (Vannieuwenhuyze et al., 2010, 2012; Voogt & Saris, 2005). However, the term does not apply well to the mixed-mode survey designs of interest in this investigation. Instead, the term *mode choice* will be used to signify the respondents' decision-making process of choosing a response mode among the given alternatives. This term also motivates future research to extend decision-making theories to mixed-mode survey design and data investigations.

Mode choice is typically confounded with the mode effects in mixed-mode surveys. Confounding of mode choice and mode effects can be shown using a general formulation of a response model for the two response modes (in-person-I, telephone -T) used in the CPS:

$$y_j = \mu_j + R_{Tj}B_{Td} + R_{Ij}B_{Id} + \varepsilon_j \text{ , where:} \tag{1.1}$$

$j = 1, 2, 3, ...N$ indexes individual persons in the survey population,

$d = 1, 2, 3, ...D$ denotes groups defined by demographics or other characteristics related to mode effects,

μ_j can depend on X_j, a vector of covariates for person j,

Subscripts T and I correspond to telephone and in-person modes,

$B_{Td} =$ reporting error for group *d* who responds by telephone,

$B_{Id} =$ reporting error for group *d* who responds by in-person,

$$R_{Tj} = \begin{cases} 1 \text{ if population unit } j \text{ responds in telephone mode} \\ 0 \qquad\qquad\qquad\qquad \text{if otherwise} \end{cases},$$

$$R_{Ij} = \begin{cases} 1 \text{ if population unit } j \text{ responds in in-person mode} \\ 0 \qquad\qquad\qquad\qquad \text{if otherwise} \end{cases}$$

$$\varepsilon_j \overset{iid}{\sim} (0, \sigma^2).$$

More generally, this formulation can be extended to as many response modes as included in the mixed-mode design. For simplicity of presentation, only telephone and in-person modes are considered in this formulation. Also any errors associated with unit and item nonresponse are ignored. The simple response model in (1.1) assumes independence of residuals among all population members, i.e. $\text{cov}(e_j, e_{j*}) = 0$.

For illustration, consider a case where the covariate X_j is categorical and divides the population into $d = 1, 2, \ldots, D$ groups. The reporting errors for each person in group *d* are B_{Td} and B_{Id}. Also, assume for this illustration that $R_{Tj} = 1 - R_{Ij}$ and that the mode choice is deterministic, i.e., each person in the population will respond by either *T* or *I* and that the choice is fixed, not random. The finite population average of the model mean in (1.1) is

$$\bar{\mu} = \frac{1}{N}\sum_{d=1}^{D}\left[\sum_{j\in U_{Td}}\mu_j + \sum_{j\in U_{Id}}\mu_j\right]$$
$$= \sum_{d=1}^{D}\left[\mathrm{Prp}_{Td}\bar{\mu}_{Td} + \mathrm{Prp}_{Id}\bar{\mu}_{Id}\right]$$

where N is the number of persons in the finite population, U_{Td} and U_{Id} are the sets of population units in group d that respond by telephone and in-person, $\mathrm{Prp}_{Td} = N_{Td}/N$ with N_{Td} being the number of persons in d that respond by telephone, and $\mathrm{Prp}_{Id} = N_{Id}/N$ where N_{Id} is the number in d that respond in-person.

If we observe the entire finite population and use the simple mean,

$$\bar{Y} = N^{-1}\sum\nolimits_{j=1}^{N} y_j\ , \tag{1.2}$$

an estimate of $\bar{\mu}$, the model expectation of $\bar{Y}$ under (1.1) is

$$E_M\left(\bar{Y}\right) = \sum_{d}\left[\mathrm{Prp}_{Td}\left(\bar{\mu}_{Td} + B_{Td}\right) + \mathrm{Prp}_{Id}\left(\bar{\mu}_{Id} + B_{Id}\right)\right] \tag{1.3}$$

Consequently, the model bias can be expressed as

$$E_M\left(\bar{Y} - \bar{\mu}\right) = \sum_{d}\left[\mathrm{Prp}_{Td}B_{Td} + \mathrm{Prp}_{Id}B_{Id}\right] \tag{1.4}$$

If $B_{Td} = B_{Id} = 0$ for all groups, then $\bar{Y}$ is model-unbiased; however, this is not likely to be the case. In almost all surveys, B_{Td} and B_{Id} cannot be estimated because the true values μ_j are unknown and no benchmarks are available. Also, the modes may be presented sequentially (e.g., telephone first and in-person second). Thus, the proportion that can be estimated from a sample is not Prp_{Id}, but is really the proportion who responded in-person given that they were presented with telephone and did not respond. If these persons had only been given the opportunity to respond in-person, their reporting error, i.e., the value of B_{Id}, might have been different. Similarly, Prp_{Td} may not be estimable. In other words, as shown in (1.4), the mode effect is confounded with the mode choice.

1.4. Statistical Error Models and Mixed-Mode Survey Inference

The recent mixed-mode surveys have adopted the ignorable mode effects assumption of the early mixed-mode surveys. These mixed-mode surveys mainly aim to decrease non-observational errors under a given survey budget. The assumption of ignorable mode effects has not been challenged by the empirical work to date. Past empirical work usually focused parts of the theoretical frameworks and full population quantities. But recent statistical methods have challenged this assumption and have specifically aimed to unconfound the mode choice and the mode effects in order to evaluate mode effects. However, these methods do not incorporate an explicit statistical error model from a Total Survey Error (TSE) view. Following Biemer and Stokes (1991) taxonomy, they adopt the psychometric view on measurement error.

Following the Total Survey Error (TSE) view, Chapter 2 shows the extension of the Lessler and Kalsbeek (1992) statistical error model for the survey mean estimator to the mixed-mode survey context. The extended survey mean estimator is restricted to single-frame mixed-mode surveys for the current discussion. Additionally the statistical error model is restricted to either one phase with multiple modes or multiple phases that considers one mode for a given phase. The statistical error model can be easily extended to the multiple mode x multiple phase case by incorporating separate terms for phase and mode. In this dissertation the terms phase and mode can be used interchangeably. The extended statistical error model is instrumental in comparing the existing methods to evaluate the mode effects in the mixed-mode surveys and motivate the proposed methods.

1.5. Proposed Mixed-Mode Survey Inference Method

As the previous research has shown, different modes of survey administration may produce data that do not all have the same accuracy. One way to explore whether there are mode effects is consider the set of respondents to the different modes separately. Then, by imputing the nonobserved responses for a particular mode, a complete sample data set for each mode can be generated as if all units had responded by that mode. These "completed" data sets can be combined to generate population estimates. Figure 1.4 is a schematic that illustrates the proposed method in the simplest form.

In Chapter 3, this estimation method is described for a two-mode survey design. The current evaluations ignore the final nonresponse and further calibration adjustments. As shown in Section 3.3.4, in combining completed data sets, different combination rules have been applied and these combination methods have been referred as plural (i.e. proposed methods).

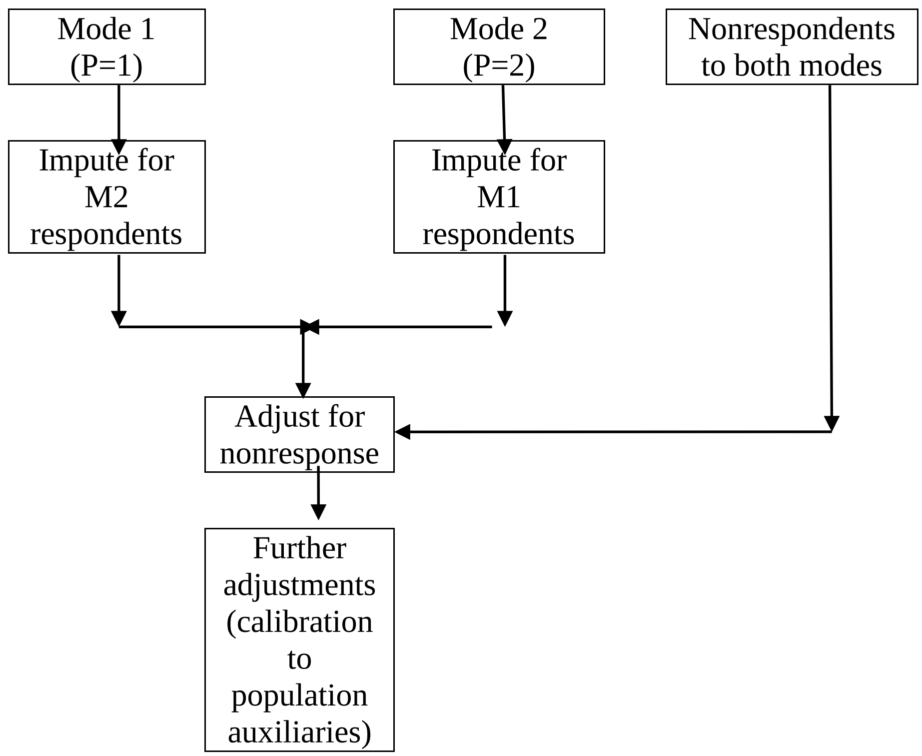


Figure 1.4 – Schematic Chart for the Proposed Mixed-Mode Survey Inference Method

1.6. Empirical and Simulation Evaluations

The proposed methods are evaluated in a series of empirical and simulation studies using CPS data. The dissertation work includes three studies. Two studies use a special data set that includes both survey responses and hypothetical true values for annual family income. The data, 1973 CPS Match Data, include both survey and Internal Revenue Service (IRS) Form 1040 data. When the IRS records data are assumed to be true values, the mode effects for a CPS-like mixed-mode survey can be evaluated. Chapter 4 details descriptive and regression analyses that investigate the mode effects for measures of total family income in this particular dataset. Chapter 4 also describes the covariates used in each of the imputation models.

Chapter 5 includes the first empirical study of the estimation methods developed in Chapter 3. The empirical study focuses on a variable of interest, wage and salary income, for which measurement complexities are minimal. A subsequent simulation study, discussed in Chapter 6, augments the empirical study data to generate samples of cases that represent more complicated measurement properties. Varying degrees of mode effects were simulated based on the observed data to evaluate the proposed methods under more complicated situations. Since both studies included benchmark values, which is not the usual case in survey practice, a third study conducts an empirical comparison analysis for both personal income and health insurance coverage for which no benchmark values are available. This empirical study is also discussed in Chapter 6.

The dissertation concludes with a discussion of conclusions and future research directions.

Chapter 2
Statistical Error Models for Mixed-Mode Survey Estimators

2.1. Mode Effects and Statistical Error Models

Generally, the definitions of *mode effects* found in the literature can be tied to one or multiple research purposes: (1) testing data comparability with respect to a benchmark or an alternative mode (Hochstim, 1967), (2) exploring differences in response patterns across modes of survey administration (Heerwegh, 2009), and (3) testing social and cognitive theories for possible differences in response behavior (Tourangeau & Smith, 1996).

While the second and third research purposes are informative, for survey estimation the interest is overall data comparability across modes. But from a statistical point of view, reporting one-time differences in estimates (Brambilla & McKinlay, 1987; Fowler, Gallagher, & Nederend, 1999), which is the usual path that the first research purpose follows, has limited generalizability in comparing the properties of an estimator under different data collection methods. Alternatively, the properties of survey estimators under different data collection methods can be studied using statistical error models under the Total Survey Error (TSE) framework (Biemer & Stokes, 2004; El Kasabi et al., (forthcoming).; Lessler & Kalsbeek, 1992; Peytchev, Ridenhour, & Krotki, 2010).

In particular, statistical measurement error models allow the bias and variance of estimators to be studied. Although mode effects are typically thought of contributing only to measurement error, the mode choice is the other mechanism which affects the survey error in mixed-mode surveys. Figure 2.1 illustrates an adaptation of the Groves et al. (2009) survey life cycle diagram to mixed-mode survey designs. The extended survey life cycle considers multiple response modes and conceptually illustrates the steps at which Total Survey Error (TSE) sources are introduced for a mixed-mode survey estimator of the finite population mean (Figure 2.1). According to Groves and Couper's framework for contact and survey participation (1998), each survey mode has features that can influence the contactability and survey participation decision. Accordingly, in cases where respondents are allowed to select a response mode, it is natural to consider a mode choice mechanism (conditioned on the modes available to the sample units) in addition to a conditional nonresponse mechanism. For a given phase, the nonresponse and the mode choice mechanism distinction is conditioned on the survey design. Allowing a mode choice mechanism implies that a unit does not report for one mode in a given phase but does report for another. For example, for the ACS-like sequential mixed-mode surveys, for the first phase, nonreporting units are generated by the mail phase. The mail nonrespondents may be telephone or in-person responses in the following phases. In this dissertation, for simplification purpose, in the illustrations the terms phase and mode are used interchangeably, p denotes mode. The mixed-mode survey designs either consider: (1) one mode per phase for multiple phases (e.g. ACS, see Table 1.1) or (2) one phase that uses multiple modes (e.g. CPS, see Table 1.2). The statistical error models discussed here can also be extended to multiple modes per phase with more than one phase being used in the full survey.

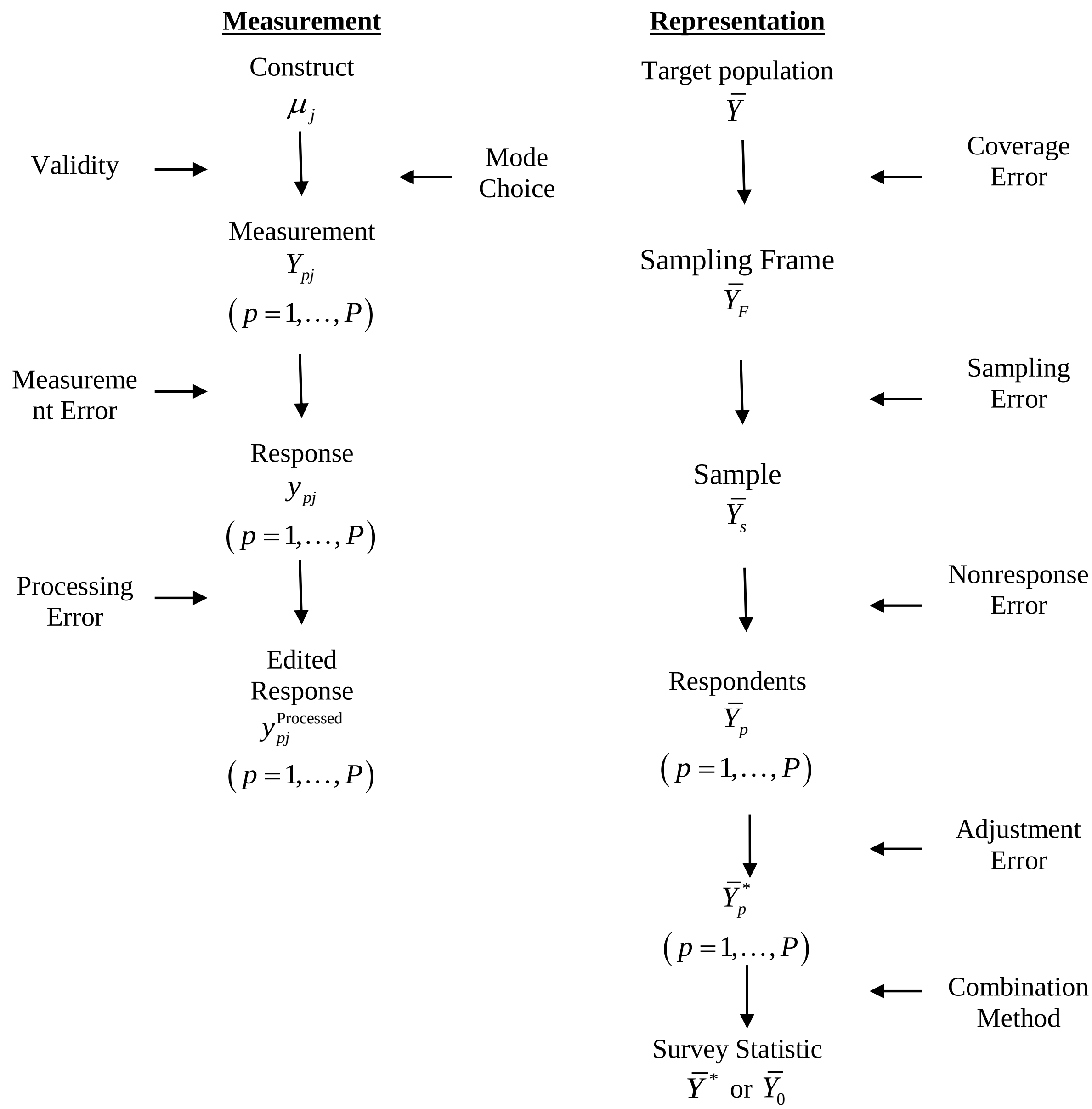


Figure 2.1 – Survey Life Cycle for Mixed-mode Surveys from a Quality Perspective
(adapted from Groves et al., 2009)

The statistical notations in the survey life cycle in Figure 2.1 are:

μ_j : The true value of a construct for the j th population element

Y_{pj} : The measured value of a construct for the j th population element collected by mode p

y_{pj} : The value of the response obtained for Y_{pj} collected by mode p

$y_{pj}^{\text{Processed}}$: The value of the response collected by mode p after editing and other processing steps

$\bar{Y}$: Population mean of the Y_j 's

$\bar{Y}_F$: Population mean of interest for the part of the population covered by the frame

$\bar{Y}_s$: Sample mean of interest

$\bar{Y}_p$: Mean of interest for respondents for a given mode p

$\bar{Y}_p^*$: Adjusted mean of interest for respondents for a given mode p

$\bar{Y}_0$: Mixed-mode survey mean that ignores mode effects

$\bar{Y}^*$: Mixed-mode survey mean

The estimator $\bar{Y}^*$ comes from combining the means from the different modes. How to do this "combining" is one of the main topics of this dissertation.

For the evaluation of the proposed method, complete coverage and response are assumed. Although these error sources are conceptually discussed, future research will include the extensions of the proposed methods to address coverage and final nonresponse error adjustments.

In the earlier years, mixed-mode survey estimators were mostly developed for special case survey designs which focused on improving representation of the survey population. These estimators assumed ignorable mode effects and complete response in the follow-up phase (Hansen & Hurwitz, 1946). In these special cases, mixed-mode surveys attempt to minimize the magnitude of the selected set of survey errors of representation, coverage and nonresponse-- implicitly assuming that differences in measurement errors are ignorable across survey modes. Although there are recent method developments which evaluate this assumption, this notion of ignorable mode effects is still dominant in mixed-mode survey inference. For example, the ACS-like sequential mixed-mode surveys aim to decrease nonresponse by offering alternative mode follow-ups, but at the same time they assume measurement differences across modes are ignorable. Doing so, the responses from different modes are combined as they are without any mode-specific adjustments.

Recently, some statistical methods have been proposed to quantify and isolate *mode effects* from inherently nonrandomized mode selection, *selection effects*, to validate mixed-mode survey estimation assumptions (Camillo & D'Attoma, 2011; De Leeuw, 2005; Jäckle et al., 2010; Lugtig et al., 2011; Vannieuwenhuyze et al., 2010, 2012; Voogt & Saris, 2005). These evaluation methods are not derived from general statistical error models that decompose survey error sources. To facilitate comparable evaluations of these methods and extensions to adjustment methods, the general single-mode survey descriptive statistical error model in Section 2.2.1 is extended to a mixed-mode survey error model in Section 2.2.2 under the framework presented in Figure 2.1. In addition, a general statistical error model allows derivation of the bias properties of a survey mean when the ignorable mode effects assumption is violated.

2.2. Statistical Error Models

The current mixed-mode survey statistical inference methods simply combine data from multiple modes without any mode adjustments under the assumption that mode effects are ignorable. The basis for this assumption may be found in early empirical studies of mode effects (Fowler et al., 1999; Groves & Kahn, 1979; Hochstim, 1967) that focused mainly on the comparability of data instead of the studies that focus on isolating mode effects and investigating the causes of the differences (Aquilino, 1994; Tourangeau & Smith, 1996). Although the need for incorporating social and cognitive theories of response behavior into more elaborate statistical error models has been emphasized in the earlier comprehensive reviews (Groves, 1999; Schwarz, Strack, Hippler, & Bishop, 1991), this school of thought has not been dominant in studying the mode effects in the later work. To date the barriers that Groves (1999) mentioned in his review continue to be significant in incorporating social and cognitive theories to statistical error models, particularly for mixed-mode survey inference. How to incorporate social and cognitive theories into mixed-mode survey inference is largely an unsolved problem. Importantly, this understudied link makes it difficult to extend the general statistical error models to elaborate on the sources of possible systematic mode effects, which is the interest of this dissertation.

Nevertheless, there are recent mathematical methods that study the implication of the violation of the ignorable mode effects assumption. Although these methods do not link the modeling of mode effects to social and cognitive theories, they are instrumental in understanding the implication of the confounded mode choice and mode effects on the accuracy and precision of survey inferences.

2.2.1. General model for a single-mode survey estimate

To construct a statistical error model for a single-frame mixed-mode survey estimator which aims to increase representativeness, it is possible to extend the descriptive statistical model for a single-mode survey mean estimator developed by Lessler and Kalsbeek (1992). This model formulates the effect of four survey errors, as shown in Figure 2.1, on the survey estimate of the population mean:

1. Measurement error
2. Nonresponse error
3. Sampling error
4. Coverage error

The true measure for population element j, assuming no mode effects, is Y_j $(j=1,2,\ldots,N)$ so that we wish to estimate:

$$\bar{Y} = \frac{\sum_{N} Y_j}{N} \tag{2.1}$$

Each error source can be distinguished as either stochastic or deterministic. With the exception of coverage error, Lessler and Kalsbeek (1992) consider each term to be a result of a stochastic process although for the household surveys, the coverage error can also be modeled as a stochastic error (Tourangeau, Shapiro, Kearney, & Ernst, 1997). Considering a case where Y_j's are imputed, explicitly or implicitly, the first two stochastic sources for survey error (measurement and nonresponse) can be incorporated in a model for the observed value for unit *j* as:

$y_j^* = R_j(y_{pj} + \varepsilon_j) + (1 - R_j)(y_{pj} + \varepsilon_{0j})$, where:

$$R_j = \begin{cases} 1 \text{ if population element } j \text{ responds, when selected} \\ 0 \qquad\qquad\qquad \text{if otherwise} \end{cases} \tag{2.2}$$

ε_j: is the error in measuring y_j if the population element j responds *(elementary response error),*

ε_{0j}: is the error in imputing a value for y_{pj} if the population element, j fails to respond *(elementary imputation error for nonresponse or measurement).*

In Figure 2.1, $\bar{Y}_P = \frac{\sum_{j \in U} R_j y_{pj}}{\sum_{j \in U} R_j}$, where U : is the population set (2.3)

represents the mean estimate if only respondents' data are used to estimate mean. For mixed-mode survey inference this estimator is referenced as $\bar{Y}_0$.

Lessler and Kalsbeek consider only random errors related to the measurement and imputation steps. In standard mixed-mode survey inference, this statistical error model is implicitly assumed and systematic differences are ignored. When the systematic differences are incorporated, the model described below in Section 2.2.2 is more appropriate.

Lessler and Kalsbeek include probability sampling as a third source of stochastic survey error and distinguish it as synthetic randomization. They impose two conditions on this stochastic error source: (1) marginal and joint probabilities of selection are known and (2) there exists a non-zero selection probability for all the members. The sample S is obtained from a frame *F*. U_F is the set of frame elements and *U* is the set of units in the target population. λ_k is the number of times frame element k is selected in the sample and $\eta_k = E_S(\lambda_k)$.

The fourth source of survey error depends on the linkage between the F frame and N population elements. This is a deterministic source of error since the sampling process is dependent on an existing frame.

$$\theta_{jk} = \begin{cases} 1 \text{ if population element } j \text{ is linked to one of the frame elements } k \\ 0 \qquad\qquad\qquad\qquad \text{if otherwise} \end{cases}$$

Three frame problems can be defined by as follows:

Undercoverage: θ_{jk} =0; for $k \in U_F$ (2.4)

Multiplicity: $\theta_j = \sum_{k \in U_F} \theta_{jk} > 1$ (2.5)

Overcoverage: $\theta_k = \sum_{j \in U} \theta_{jk} = 0$ (2.6)

If we assume that neither undercoverage or overcoverage exist in the frame, the population estimator will follow:

$$\bar{Y}^* = \sum_{j \in U} \sum_{k \in U_F} \frac{y_j^* \lambda_k \theta_{jk}}{\eta_k}, \text{ where:} \tag{2.7}$$

$\eta_k = E_S(\lambda_k)$ is the expected number of times the kth element is selected in the sample.

2.2.2. General model for a single-frame mixed-mode survey estimate

In single-frame mixed-mode surveys, measurement takes place via multiple modes. When the R_j nonresponse indicator is replaced with mode specific indicators R_{pj} in (2.2), the same logic can be used to derive mode-specific mean estimators. For specificity, we consider a two-mode survey with *T* = telephone and *I* = personal interview. The mixed-mode design is conditioned on one phase with multiple modes. Here we assume that overall response is complete and that nonresponse to one mode option (e.g. $R_{Tj} = 0$) implies response in an alternative mode (e.g. $R_{Ij} = 1$).

The formulation below can be extended to more than two modes. A model that describes the process of selecting a mode and then responding via that mode is:

$$y_j^* = R_{Tj} y_{Tj} + R_{Ij} y_{Ij} \text{, where:}$$

$$j = 1, 2, 3, \ldots, N, \tag{2.8}$$

$$R_{Tj} = \begin{cases} 1 & \text{if population unit } j \text{ responds in telephone mode} \\ 0 & \text{if otherwise} \end{cases},$$

$$R_{Ij} = \begin{cases} 1 & \text{if population unit } j \text{ responds in in-person mode} \\ 0 & \text{if otherwise} \end{cases}.$$

To reflect the possibility that the value reported when a person uses mode p may be incorrect, suppose that

$$y_{pj} = \mu_j + B_{pj} + \varepsilon_{pj} \quad (p = T \text{ or } I) \tag{2.9}$$

where ε_{pj} is a random error with mean 0. We consider the possible differences between μ_j and y_{pj} could be due to validity violations or/and measurement conditions that may vary by mode.

$B_{Tj} =$ reporting error for person j who responds by telephone,

$B_{Ij} =$ reporting error for person j who responds by in-person.

Letting E_M denote the expectation with respect to model (2.9), the average value reported by person j is $E_M\left(y_{pj}\right) = \mu_j + B_{pj}$, i.e., the report is biased compared to the desired value μ_j. Also, define U_p = set of persons that respond using mode p ($p = T$ or I). If we assume there is no undercoverage or overcoverage in the single-frame mixed-mode survey, then $U_F = U$. Given a particular split of the universe into U_T and U_I (i.e., conditioning on an outcome of the R_{pj} 's), the population estimator of the mean that does not correct for any differences in mode will be:

$$\bar{Y}^* = \sum_{p=(T,I)} \sum_{j \in U_p} \frac{y_j^* \lambda_k}{\eta_k} \tag{2.10}$$

where previous descriptions apply. Assume that the sample design is such that $\eta_k = E_S(\lambda_k)$. If $\bar{Y}^*$ was unbiased, its expectation (over all random quantities) would be $\bar{\mu} = \sum_{j \in U} \mu_j \big/ N$. However, if $B_{Tj}, B_{Ij} \neq 0$, i.e. in the presence of nonignorable mode effects, $\bar{Y}^*$ will be biased.

Considering mode selection to be a random process, we can also model R_{pj} as a random variable with mean $\mathrm{E}_R(R_{pj}) = g(\mathbf{x}_j; \psi_p) \equiv g_{pj}$ where $g(\bullet)$ is a function like the logit or probit and $\mathbf{x}_j$ is a vector of covariates for person *j*. In the two-mode case with no nonresponse, assume that a single parameter vector ψ applies and that $\mathrm{E}_R\left(R_{Tj}\right) = g(\mathbf{x}_j; \psi) = 1 - \mathrm{E}_R\left(R_{Ij}\right)$.

The order in which modes are made available to sample units can vary from one survey to another. If mode choice is modeled as random, this affects how one interprets the R_{pj} 's in (2.8). For example, in a mixed-mode survey, response modes could be made available in a sequence or concurrently (Cobben, Schouten, & Bethlehem, 2006). In a sequential mixed-mode survey, the response modes that are made available to the sampled units vary depending on the phase. In an ACS-like design, in the first phase only the mail response option is available. Sample units either respond by mail or do not respond at all. In the second phase, when the telephone mode is available, nonrespondents to the telephone mode can choose between mail and telephone modes to respond or do not respond at all. In contrast, in a concurrent Address Based Sample (ABS) Web-Mail survey, sample units can choose to select web or mail to respond or do not respond.

In a sequential mixed-mode survey, R_{pj} indicates whether unit *j* chose mode *p* in a given phase of the survey. The unit will have been offered other modes in previous phases. In that case, R_{pj} =1 in the current phase means that unit j did not choose any of the previously offered modes. In a concurrent mixed-mode survey, all modes are presented at once and R_{pj} could be modeled as the result of randomly choosing among all the modes. The sequential mixed-mode design corresponds to multiple phases with one mode, and concurrent mixed-mode survey design corresponds to a design with one phase with multiple modes in this dissertation.

As noted above, model (2.9) does not distinguish between construct and measurement. In our view, the validity of the measurement may demand different question formatting across different survey modes, although careful design steps need to be taken (Couper, 2008; Gray, Blake, & Campanelli, 2011; Martin et al., 2007; Nicolaas, Campanelli, Hope, Jäckle, & Lynn, 2011). In this dissertation, there is no attempt to distinguish the validity of the question format/wording from measurement errors in the statistical error models. In addition, processing and measurement errors are combined.

When subsampling is considered in the follow-up phases, phases (as in the ACS in-person phase), an additional stochastic source of survey error should be accounted for as synthetic randomization (probability sampling) conditioned on the response status.

The statistical error model described in this section is used for three purposes in the following sections: (1) to compare the existing mixed-mode survey mode effect evaluation methods (Section 2.3), (2) to understand the bias properties of $\bar{Y}_0$ (Section 3.1), and (3) to motivate and evaluate the proposed mixed-mode survey estimator (Section 3.3).

2.3. Existing Mode Effect Evaluation Methods

Current methods for evaluating mode effects focus on unconfounding the nonrandom selection of modes and mode effects in mixed-mode surveys. Regression model methods control for the nonrandom selection analytically (Jäckle et al., 2010). The mixture distribution method, introduced by Vannieuwenhuyze et al. (2010, 2012), computes the selection and mode effects for distribution parameter estimates, such as mean, of a variable of interest, Y, by defining Y as having a mixture distribution given the mixed-mode survey design. Using a parallel single mode survey and assuming the same population and measurement properties, mathematically it is possible to show that parameters related to the selection and the mode effects can be quantified for a two-mode survey. To do these comparisons, the mixture mode distribution method relies on two key assumptions which Vannieuwenhuyze et al. (2010) term *completeness* and *representativity*. These assumptions imply that parameter estimates obtained from the single-mode survey and the mixed-mode survey are unbiased estimates for the same survey population with respect to the non-observational survey error. In practice, this could be a strong assumption as the mixed-mode surveys are usually conducted to minimize the nonresponse bias under a certain budget constraint. On the other hand, the method conceptualizes the confounding nature of mode choice and mode effects. Additionally, the method enables the computation of required sample sizes to detect mode effects. Alternatively, propensity score matching methods (Lee & Valliant, 2007; Rosenbaum & Rubin, 1983, 1984) unconfound the mode choice and the mode effects based on the propensity score matching strata that have been formed using the available covariates. This method defines the mode effects as the mean differences between the matched groups. All these methods focus on the comparability of the survey data as opposed to determining the “best” performing mode.

2.3.1. Regression Model Methods

Jäckle et al. (2010) evaluate methods that define the mode effect as the differences in the mean or predicted response distributions between modes after controlling for some selected social-demographic variables. Although the data are obtained from randomized experiments, due to nonresponse the differences in the social-demographic distributions of respondents are controlled analytically. Jäckle et al. (2010) apply partial proportional odds models to test the linearity assumption for ordinal variables in addition to proportional odds and linear regression models. The methods derive underlying response distributions for alternative modes conditioned on some selected social-demographic variables. Although these models are not linked to the social and the cognitive theories, they illustrate how different modeling assumptions may yield different results in mode effect evaluations. Jäckle et al. (2010) discuss two regression model structures in particular that we summarize next.

Regression Model for Continuous Variables

$$\hat{\mu}_p = X\hat{\beta}^{OLS} + R_p\hat{\beta}_p^{OLS} \text{, where:} \tag{2.11}$$

$\hat{\mu}_p$: The predicted mean value for a given mode p, where $p = 1, 2, ..., P$,

X : Selected social-demographic covariates for respondents,

$\hat{\beta}^{OLS}$: Linear regression model parameter estimates associated with the social-demographics using ordinary least squares (OLS),

R_p: $N \; x \; P$ matrix that includes dummy indicators for modes $p = 1, 2, ..., P$,

$\hat{\beta}_p^{OLS}$: Mode effect estimate under linear regression model.

Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Variables

$$\Pr(y_p > c) = \frac{\exp(X\hat{\beta}_c^{Odds} + R_P\hat{\beta}_{pc}^{Odds})}{1 + [\exp(X\hat{\beta}_c^{Odds} + R_P\hat{\beta}_{pc}^{Odds})]} \text{, where:} \tag{2.12}$$

(2.12)

$\Pr(y_p > c)$: Predicted proportion of $y_p > c$ in which $c = 1, 2, ..., C$ index the response categories,

$\hat{\beta}_c^{Odds}$: Generalized ordered logit or partial proportional odds model parameter estimates for selected social-demographic covariates for respondents for a given category,

$\hat{\beta}_{pc}^{Odds}$: Mode effect estimate under generalized ordered logit or partial proportional odds model for a given category.

Generalized ordered logit models allow a relaxation of the assumption of the parallel line regression or proportional odds assumption that is generally implicit in modeling ordinal variables (Williams, 2006). The parallel line regression or proportional odds model assumption imposes the constraint that regression parameters except for the ordered intercepts to be same for all the response categories.

Jäckle et al. (2010) showed that in assessing the mode effects this assumption may yield different conclusions. This finding is particularly important since it reiterates the importance of tying the social and the cognitive theories to the statistical models. In the follow-up research, Lynn, Hope, Jäckle, Campanelli, and Nicolaas (2011) use these models to test specific mode effect hypotheses based on the social and the cognitive theories. This kind of work will help to improve both the mixed-mode design principles and set the modeling assumptions in the adjustment models.

The regression methods compare the response distributions for each mode under these models. Student t-tests for $\hat{\beta}_{p}^{OLS}$ and $\hat{\beta}_{Pc}^{Odds}$ are the statistical tests to determine the significance of mode effects. In addition, Wald tests can be used to test the parallel line regression assumption.

This dissertation examines only the regression models for continuous and binary variables to adjust the mode effects in the mixed-mode surveys in which nonrandomized selection of modes occur. Future research needs to test the methods for ordinal variables when the parallel line regression assumption is violated as described in (Jäckle et al., 2010).

2.3.2. Mixture Distribution Method

The method described in Vannieuwenhuyze et al. (2010, 2012) uses a single-mode survey as the reference distribution and estimates selection and mode effects for a two mode survey with respect to this reference distribution.

This method considers mode-specific distributions of Y, one as measured by in-person (I) survey mode Y_I, and another as measured by telephone (T) survey mode Y_T when two survey modes are considered. While the method is not specific to one parameter of a distribution, the mean is considered for this discussion. $\bar{\mu}_U = \frac{1}{N}\sum_{j \in U} \mu_j$ is the parameter of interest given the underlying Y distribution.

Assuming that one of the mode-specific distributions Y_I or Y_T can be measured, the density function of the mode-specific distribution can be written as a combination of conditioned distributions:

$f_T(Y) = \bar{g} f_T(Y \mid U_T) + (1-\bar{g}) f_T(Y \mid U_I)$, where:

$\bar{g} = \frac{N_T}{N}$ and, as in Section 2.2.2,

U : is the population set,

U_T : set of telephone respondents,

U_I : set of in-person respondents,

$$U = U_T \cup U_I . \tag{2.13}$$

As illustrated in Table 2.1, given a mixed-mode survey only conditional means can be observed for a given mode. The columns represent the conditional distribution of the variable of interest, Y. Column (1) represents a single-mode survey data that presumably collects survey data that represents all the members of the population using a telephone survey mode. Columns (2) and (3) together represent data collected by a mixed-mode survey. Presumably, this mixed-mode survey data also represent all the members of the population but in addition to telephone interview data, Column (2), some data are collected by in-person interviews, Column (3). Rows (2) and (3) illustrate conditional means given a two mode survey. Shaded cells in Rows (2) and (3) cannot be observed in a mixed-mode survey. On the other hand, a single mode survey can produce the corresponding conditional means using one mode for these unobserved cells.

Table 2.1 – Selection and Mode Effects as defined by Vannieuwenhuyze et al. (2010, 2012)

$Y_T \mid U$	$Y_T \mid U_T$	$Y_I \mid U_I$
(1)	(2)	(3)
$\mu_T \mid U_T$	$\mu_T \mid U_T$	
$\mu_T \mid U_I$		$\mu_I \mid U_I$

The Vannieuwenhuyze et al. (2010, 2012) formulations do not distinguish the "true value" and measurement error. But for comparison purposes, we will distinguish between the true value and measurement error considering the simple measurement error model in (2.9).

Suppose the mode choice mechanism is described by the model introduced in Section 2.2.2:

$$\Pr(R_{Tj}=1)=g_j \text{ and } \Pr(R_{Ij}=1)=1-g_j \tag{2.14}$$

Instead of defining a function for the selection mechanism $g(\bullet)$ explicitly, Vannieuwenhuyze et al. (2010, 2012) use $\overline{g}=\frac{N_T}{N}$ as shown in (2.13).

Next, we extend the Vannieuwenhuyze et al. (2010, 2012) definition of a *selection effects* using model (2.8). Both terms *mode choice* and *selection effects* refer to the nonrandom assignment of modes to respondents. As was pointed out earlier *mode choice* is the preferred term in this dissertation. One subtle difference between these terms is *selection effects* refer to the differences in quantities as a result of *mode choice,* where *mode* choice refers to the mechanism of nonrandom assignment of modes. To avoid confusion, in this section the term selection effects is used as Vannieuwenhuyze et al. (2010, 2012) defines it in a specific way. Otherwise, the term *mode choice* is used in the rest of the dissertation.

To define a selection effect, one mode is specified as the "reference" mode. We then imagine that two sets of units (U_T and U_I here) are enumerated using the same mode. In the present context, suppose that T is the reference mode. To specify this clearly, let $\left(\bar{Y}_T \mid p = T\right)$ denote the mean for the persons in U_T, assuming that they use mode T to respond. Let $\left(\bar{Y}_I \mid p = T\right)$ be the mean for persons in U_I, assuming that they also use mode T to respond. In a particular mixed-mode survey, $\left(\bar{Y}_I \mid p = T\right)$ is unobservable because persons in U_I responded only via mode I.

With these definitions, the T selection effect is the difference in the mean for units that are actually enumerated by T and the mean for units actually enumerated by I but assuming that both sets of units used mode T:

$$S_T(Y) = (\bar{Y}_T \mid p = T) - (\bar{Y}_I \mid p = T) \tag{2.15}$$

The model-expectation under the model specified by (2.8) and (2.9) of the T selection effect is:

$$\begin{aligned} E_M\left[S_T(\bar{Y})\right] &= E_M\left[(\bar{Y}_T \mid p = T) - (\bar{Y}_I \mid p = T)\right] \\ &= \frac{1}{N_T}\sum_{j \in U_T}(\mu_j + B_{Tj}) - \frac{1}{N_I}\sum_{j \in U_I}(\mu_j + B_{Tj}) \\ &= \underbrace{\frac{1}{N_T}\sum_{j \in U_T}\mu_j}_{\bar{\mu}_T} + \underbrace{\frac{1}{N_T}\sum_{j \in U_T}B_{Tj}}_{\bar{B}_T(T)} - \underbrace{\frac{1}{N_I}\sum_{j \in U_I}\mu_j}_{\bar{\mu}_I} - \underbrace{\frac{1}{N_I}\sum_{j \in U_I}B_{Tj}}_{\bar{B}_I(T)} \qquad (2.16) \\ &= \bar{\mu}_T + \bar{B}_T(T) - \bar{\mu}_I - \bar{B}_I(T) \\ &= \left[\bar{\mu}_T - \bar{\mu}_I\right] + \left[\bar{B}_T(T) - \bar{B}_I(T)\right] \end{aligned}$$

where $\bar{B}_T(T)$ is the mean telephone bias among persons who responded by telephone and $\bar{B}_I(T)$ is the mean telephone bias among persons who actually responded in-person. Thus, the telephone selection effect depends on both the difference in true means for the sets of units that actually responded via T and I and the difference in their average reporting biases.

Using similar notation, the in-person selection effect can be defined as:

$$S_I(\bar{Y}) = (\bar{Y}_I \mid p = I) - (\bar{Y}_T \mid p = I) \tag{2.17}$$

The model-expectation of the in-person selection effect can be defined as the difference in the U_I and U_T sets, assuming that both were enumerated with mode I :

$$\begin{aligned} E_M[S_I(\bar{Y})] &= E_M\left[(\bar{Y}_I \mid p=I)-(\bar{Y}_T \mid p=I)\right] \\ &= \frac{1}{N_I}\sum_{j\in U_I}(\mu_j+B_{Ij})-\frac{1}{N_T}\sum_{j\in U_T}(\mu_j+B_{Ij}) \\ &= \bar{\mu}_I+\bar{B}_I(I)-\bar{\mu}_T-\bar{B}_T(I) \\ &= \left[\bar{\mu}_I-\bar{\mu}_T\right]+\left[\bar{B}_I(I)-\bar{B}_T(I)\right] \end{aligned} \tag{2.18}$$

In this case, $\bar{B}_I(I)$ is the mean in-person bias among persons who responded in-person and $\bar{B}_T(I)$ is the mean in-person bias among persons who actually responded by telephone. The I selection effect depends on the difference in the true means and the difference in average bias, assuming that both sets of units used the in-person mode. Expressions (2.16) and (2.18) are composed of two parts: (1) the difference in conditional means and (2) the difference in mode specific average biases. Therefore, selection effects by definition capture the differences in the respondent composition between two modes.

In addition, the mixture distribution method defines mode effects with respect to a reference survey as the difference in the mean for a mode b and the mean for another mode a for the same set of units. The definition of the telephone vs. in-person measurement effect is then:

$$M_T(\bar{Y})=(\bar{Y}_T \mid p=T)-(\bar{Y}_T \mid p=I) \tag{2.19}$$

The model-expectation of telephone measurement effect is:

$$\begin{aligned} E_M[M_T(\bar{Y})] &= E_M\left[(\bar{Y}_T \mid p=T)-(\bar{Y}_T \mid p=I)\right] \\ &= \frac{1}{N_T}\sum_{j\in U_T}(\mu_j+B_{Tj})-\frac{1}{N_T}\sum_{j\in U_T}(\mu_j+B_{Ij}) \\ &= \frac{1}{N_T}\sum_{j\in U_T}(B_{Tj}-B_{Ij}) \\ &= \bar{B}_T(T)-\bar{B}_T(I) \end{aligned} \tag{2.20}$$

The mode effect could also be defined using the set of persons who responded in-person, i.e. U_I. Using similar notation to (2.19), we have

$$M_I(\bar{Y}) = (\bar{Y}_I \mid p = I) - (\bar{Y}_I \mid p = T) \tag{2.21}$$

The model-expectation of in-person measurement effect is:

$$\begin{aligned} E_M[M_I(\bar{Y})] &= E_M\left[(\bar{Y}_I \mid p = I) - (\bar{Y}_I \mid p = T)\right] \\ &= \frac{1}{N_I}\sum_{j \in U_I}(\mu_j + B_{Ij}) - \frac{1}{N_I}\sum_{j \in U_I}(\mu_j + B_{Tj}) \\ &= \frac{1}{N_I}\sum_{j \in U_I}(B_{Ij} - B_{Tj}) \\ &= \bar{B}_I(I) - \bar{B}_I(T) \end{aligned} \tag{2.22}$$

Expression (2.22) is the difference in the average biases for the U_I respondents assuming that they respond in the two different modes.

Next, consider two hypothetical means—one in which the population mean is based entirely on responses via I and one in which the entire population responds by T. To that end, define

$\bar{\bar{Y}}_T = N^{-1}\sum_U \left(y_j^* \middle| p = T\right)$ is the mean assuming that all persons respond by T

$\bar{\bar{Y}}_I = N^{-1}\sum_U \left(y_j^* \middle| p = I\right)$ is the mean assuming that all persons respond by I

Calculating the model expectations of $\bar{\bar{Y}}_T$ and $\bar{\bar{Y}}_I$ gives

$$\begin{aligned} \mathrm{E}_M\left(\bar{\bar{Y}}_T\right) &= N^{-1}\left[\sum\nolimits_{U_T}\left(\mu_j + B_{Tj}\right) + \sum\nolimits_{U_I}\left(\mu_j + B_{Tj}\right)\right] \\ &= \mathrm{g}\left[\bar{\mu}_T + \bar{B}_T\left(T\right)\right] + (1 - \mathrm{g})\left[\bar{\mu}_I + \bar{B}_I\left(T\right)\right] \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}_M\left(\bar{\bar{Y}}_I\right) &= N^{-1}\left[\sum\nolimits_{U_T}\left(\mu_j + B_{Ij}\right) + \sum\nolimits_{U_I}\left(\mu_j + B_{Ij}\right)\right] \\ &= \mathrm{g}\left[\bar{\mu}_T + \bar{B}_T\left(I\right)\right] + (1 - \mathrm{g})\left[\bar{\mu}_I + \bar{B}_I\left(I\right)\right] \end{aligned}$$

where, as defined earlier, $g = \frac{N_T}{N}$. Vannieuwenhuyze et al. (2012) define the difference $\bar{\bar{Y}}_T - \bar{\bar{Y}}_I$ as the marginal measurement effect, $M\left(\bar{Y}\right)$. The expectation of the marginal measurement effect is equal to

$$
\begin{aligned}
\mathrm{E}_M(M(\bar{Y})) &= \mathrm{E}_M\left(\bar{\bar{Y}}_T - \bar{\bar{Y}}_I\right) \\
&= \bar{g}\left[B_T(T) - B_T(I)\right] + (1-\bar{g})\left[\bar{B}_I(T) - \bar{B}_I(I)\right] \\
&= \bar{g}M_T(\bar{Y}) - (1-\bar{g})M_I(\bar{Y})
\end{aligned}
\tag{2.23}
$$

That is, the overall hypothetical marginal measurement effect for means can be written as a weighted difference of the two measurement effects described in (2.19) and (2.21).

Unlike the model-expectation of selection effects in (2.15) and (2.17), the model expectation of mode effects in (2.19) and (2.21) are defined conditional on the same group of respondents and include differences in average mode specific bias.

As a result of the mixture distribution properties, mathematically all the parameters required to define selection and mode effects can be derived. However, estimating counterfactual terms like $\bar{B}_I(T)$ and $\bar{B}_T(I)$ requires specialized data sets. If a comparison data set with true values is available, as is the case with the CPS-IRS file discussed in Chapter 3, then estimation is possible. Vannieuwenhuyze et al. (2012) give conditions under which a single-mode reference survey can be used for estimation in parallel to a mixed mode survey. The method assumes *completeness* and *representativity*, i.e. complete response and no change in the measurement mechanisms between single mode (reference survey) and mixed mode survey.

Although rewriting the selection and mode effects definitions in terms of $\bar{\mu}_I, \bar{\mu}_T, \bar{B}_I(\mathrm{T}), \bar{B}_I(\mathrm{I}), \bar{B}_T(\mathrm{T})$, and $\bar{B}_T(\mathrm{I})$ is not instrumental in the estimation of the selection and the mode effects in the application of mixture distribution method, it helps to motivate the comparisons to the proposed method and the extensions of the method. A possible extension is to use $\hat{M}(\bar{Y})$ and $\hat{\sigma}^2_{\hat{M}(\bar{Y})}$ to adjust for the mode effects.

2.3.3. Propensity Score Matching

Another method to assess mode effects in mixed-mode surveys is the propensity score matching method (Lee & Valliant, 2007; Rosenbaum & Rubin, 1983, 1984). In the context of mixed-mode survey data analysis, the propensity score matching method relies on detecting groups with similar mode choice probability scores, g_j, in alternative response modes and compare mode effects as defined by differences in means, for example $\bar{Y}_T - \bar{Y}_I$, for a given group which has a $\bar{g}$ score on the average (Camillo & D'Attoma, 2011; Lugtig et al., 2011). Although this method does not assume an underlying measurement error, the mode effects are implicitly defined as the difference in the average systematic reporting errors between modes for a given matching group.

Although this seems to be a straightforward adjustment for mode choice mechanism, the issue of unbalanced data presents itself as a problem as in the other propensity score matching method applications. Unbalanced data occurs when there are no matched cases given a g_j for a given set of covariates. When there is a different coverage by frame, this is inherited in survey data (see Figure 1 in Lugtig et al. (2011)). For example, Lugtig et al. (2011) study households who were invited by mail to take a web survey and households with registered landline telephone numbers who were contacted by telephone. Inherently, the telephone frame excludes households with only non-landline telephones. So by definition web data include responses from the non-landline telephone households and a portion of the web responses cannot be matched to telephone responses due to coverage differences. In addition to coverage differences, some responses may not be matchable because of differential nonresponse. Both Lugtig et al. (2011) and Camillo & D'Attoma (2011) studies exclude unmatched data to make the evaluations of the mode effects. In general, imbalance data lead to more restricted modeling assumptions in the propensity score matching (Iacus, King, & Giuseppe Porro, 2012). To test the imbalance in the data, Camillo & D'Attoma (2011) extended the propensity score matching method to a global imbalance test.

2.4. Existing Mode Effect Adjustment Methods

The methods described in Section 2.3 aim to assess mode effects and they can also be extended to estimation adjustment methods. On the other hand, to date only a particular adjustment method has been discussed by Buelens and Van den Brakel (2011) in the literature, that is the response mode calibration estimator. This response mode calibration estimator does not attempt to unconfound mode choice and mode effects, instead sets the total measurement error to a constant for the population total. Setting the total measurement error to a constant allows unbiased measurement of change in totals as shown in Section 2.4.1.

2.4.1. Response Mode in Calibration Estimator

Buelens and Van den Brakel (2011) extended the classical GREG estimator for the mean of Y. Assuming a linear association between the variable of interest and a subset of covariates:

$$y_j = \mu(X_j, \beta^{(\mu)}) + \varepsilon_j \text{ , where:} \tag{2.24}$$

$j = 1, 2, 3, ...N$ indexes individual population persons,

μ_j can depend on X_j, a vector of covariates for person j,

$\varepsilon_j \overset{iid}{\sim} (0, \sigma^2)$.

Following the usual calibration notation, the corresponding GREG estimator for the total of Y, T, is

$$\hat{T}_{yr} = \sum\nolimits_{j \in S} w_j y_j \text{ , where:} \tag{2.25}$$

$$w_j = \frac{1}{\pi_j}(1 + \frac{X_j^T}{\text{var}_M(y_j)}(\sum\nolimits_S (\frac{X_j X_j'}{\text{var}_M(y_j)\pi_j})^{-1}(T_X - \hat{T}_{X\pi})) \text{, where:}$$

π_j : the probability that unit j is included in sample S,

$T_X = \sum\nolimits_{j \in U} x_j$,

$\hat{T}_{X\pi}$: the π estimator of T_X .

When the population size, N , is known, the GREG estimator for the mean is:

$$\bar{Y}_r = \frac{\hat{T}_{yr}}{N} \tag{2.26}$$

When systematic reporting error by mode is introduced into (2.24):

$y_j = \mu(X_j, \beta^{(\mu)}) + R_{Pj}B_P + \varepsilon_j$, where:

$j = 1, 2, 3, ...N$ indexes individual population persons,

$p = 1, 2, 3, ..., P$ denotes survey response mode,

μ_j can depend on X_j , a vector of covariates for person j ,

$B_p =$ reporting error for person j who responds by mode p ,

$$R_{pj} = \begin{cases} 1 \text{ if population unit } j \text{ responds in } p \text{ mode} \\ 0 \qquad\qquad \text{if otherwise} \end{cases}$$

$$\varepsilon_j \overset{iid}{\sim} (0, \sigma^2)$$

The classical GREG estimator for a total can be extended to:

$$\hat{T}_{yr} = \sum_{j \in S} w_j (\mu(X_j, \beta^{(\mu)}) + R_{pj}B_p + \varepsilon_j) \tag{2.27}$$

Then the expectation of $\hat{T}_{yr}$ with respect to the sampling and Y-response model is:

$$E_S E_M (\hat{T}_{yr}) = T_y + E_S \left[\sum_S w_j (R_{pj} B_p) \right] \tag{2.28}$$

Buelens and Van den Brakel (2011) define the second part of (2.28) as total measurement error and rewrite it as:

$$\begin{aligned} E_S \left[\sum_S w_j (R_{pj} B_p) \right] &= \sum_P B_p \sum_S w_j R_{pj} \\ &= \sum_P B_p \hat{T}_p \end{aligned} \tag{2.29}$$

When the total measurement error (2.29) is plugged into (2.28):

$$E_S E_M (\hat{T}_{yr}) = T_y + \sum_P B_p E_S \left[\hat{T}_p \right] \tag{2.30}$$

As the expectation of $\hat{T}_{yr}$ with respect to the sampling and Y-response model in (2.30) shows $\hat{T}_{yr}$ is not an unbiased estimator unless $B_{pj} = 0$. Buelens and Van den Brakel (2011) focus on the implication of the total measurement error when the research interest is to estimate the difference in T_y over time. Buelens and Van den Brakel (2011) claim that although it is plausible to consider the B_p to be constants over time, due to nonrandom assignment of modes and possible design variations $E_S\left[\hat{T}_p\right]$ is not expected to be constant over time. In other words, differences in total measurement error are confounded with the real differences from time 1 to time 2 as in:

$$E_S E_M(\hat{T}^1_{yr} - \hat{T}^2_{yr}) = T^1_y - T^2_y + \sum\nolimits_P B_p\left(E_S\left[\hat{T}^1_p\right] - E_S\left[\hat{T}^2_p\right]\right) \qquad (2.31)$$

The superscripts (1) and (2) correspond to time (1) and (2) in (2.31). $E_S\left[\hat{T}^1_p\right] - E_S\left[\hat{T}^2_p\right]$ requires to be zero for $\hat{T}^1_{yr} - \hat{T}^2_{yr}$ to be unbiased. To meet that condition, Buelens and Van den Brakel (2011) replace $\hat{T}^1_p$ and $\hat{T}^2_p$ with constants, Υ_p, to offset the effect of mode effects in estimation of the difference between time 1 and time 2. When the time superscript is ignored this condition implies that:

$$\hat{T}_p = \sum\nolimits_S w_j R_{pj} = \Upsilon_p \qquad (2.32)$$

Υ_P are chosen arbitrarily and treated as population controls. For example, Buelens and Van den Brakel (2011) chose Υ_P to be equal to the reference survey response mode proportions conducted in their example application. Alternatively, response propensities can be used to estimate population mode response proportions. (2.32) is achieved by including the response mode indicator in the GREG weighting model and mode calibrated GREG estimator is:

$$\hat{T}^*_{yr} = \sum\nolimits_{j \in S} w^p_j y_j \qquad (2.33)$$

Unless the calibration corrects for the differential nonresponse fully, mode calibration would add bias to the total estimator. Buelens and Van den Brakel (2011) propose two alternative ways to check this assumption. The first approach depends on the availability of variables which are not subject to mode effects and have both survey and population values. GREG survey estimates and population estimates could be compared to detect whether GREG calibration completely corrects for the differential nonresponse. Secondly, they suggest applying different levels of calibration to conduct an empirical comparison analysis. They acknowledge a more appropriate method would be based on experimental designs. This method does not adjust for the bias in the total estimator, it calibrates the total measurement error to be equal constants to offset the difference in the difference estimator.

In summary, each of the methods described in this section has some limitations but each one offers novel pieces to the mixed-mode survey inference puzzle. Jäckle et al. (2010) show how different modeling assumptions could yield to different conclusions about the mode effects. The authors follow up their modeling work with research that links social and cognitive theories and modeling assumptions.

The Vannieuwenhuyze et al. (2010, 2012) method is limited in incorporating different measurement error model structures for the random error terms. For example, interviewers and regional offices impose different error structures for measurement error models. The method is restricted to two survey response modes. In addition, it assumes a constant (average) response propensity for a mode. The method also needs to be extended for mode effect adjustment.

Propensity score matching methods are also limited in incorporating different measurement error model structures for random error terms (Camillo & D'Attoma, 2011; Lugtig et al., 2011) The problem could be redefined by isolating all the sources of error, coverage and nonresponse and selection effects in the application of the method. This method is promising as an exploratory analysis to determine whether cell sizes are appropriate for methods such those proposed in Chapter 3 in which nonobserved responses to a specific mode will be explicitly imputed.

Although mixture distribution (Vannieuwenhuyze et al. 2010, 2012) and propensity adjustment methods (Camillo & D'Attoma, 2011; Lugtig et al., 2011) focus on only assessing the severity of mode effects, they can be extended as adjustment methods. On the other hand, the GREG mode calibration estimator offers an indirect assessment feature for mode effects and focuses on the calibration of mode effects to offset the confounding mode effect on the difference estimator (Buelens & Van den Brakel, 2011). Buelens & Van den Brakel (2011) chose to set the arbitrary constants to the proportions of response modes from a reference survey in their case study. Alternatively, an explicit mode choice function that defines mode response propensities can be incorporated into the estimator. In this method, one of the key assumptions is the complete effectiveness of the calibration method for the coverage and nonresponse error. To test this assumption requires validation data.

In addition to the described methods in this section, there is current research that extends fractional imputation to mixed-mode survey inference (Kim, 2011).

The proposed imputation method as further described in this section addresses some of the shortcomings of the existing methods and allows for mode effects adjustment. But it relies on the modeling assumptions, as do all of the existing methods, and should be tied to social and cognitive theories. Also, it should not be considered as a substitute for exploring the mode effects under the proper experimental designs as described in Tourangeau et al. (2000b).

Chapter 3

Proposed Mixed-Mode Survey Inference Methods

3.1. Bias Properties of Mixed-Mode Survey Mean Estimator

In the presence of mode effects, the bias of $\bar{Y}$ in (1.4) that simply combines data from different modes is not known. For example, the Tobacco Use Supplement to the Current Population Survey (TUS-CPS) data are used to produce estimates of the prevalence of current smoking in the U.S. (Soulakova et al., 2009). Both the regression analysis of Soulakova et al. (2009) and randomized experiments conducted by Beland and St- Pierre (2008) supported the finding that mode effects for self-reported smoking status were significant for some of the subgroups. In particular, Beland and St-Pierre (2008) reported that 18-29 years old whites and males were more likely to report being a smoker in the in-person mode. Soulakova et al. (2009) found 18-24, 24-44 and 45-64 year olds were more likely to report to be a smoker in the in-person mode compared to a 65+ age group. Soulakova et al. (2009) also found that differences between the two modes were higher for the males. In other words, $B_{Td} \neq B_{Id} \neq 0$ in (1.4). Furthermore, if we consider R_{Tj} and R_{Ij} to be random variables, members of these subgroups may randomly choose to respond either in telephone or in-person in each replication of the TUS-CPS. When mode choice is stochastic, the proportions of the telephone and in-person responses by groups, Prp_{Td} and Prp_{Id}, will vary in mixed-mode surveys. This means that possible underreporting of smoking status in the telephone mode makes a random contribution to the overall bias of the estimated prevalence of smoking for a random proportion of the sample in each replication of the survey. Thus, the estimates from different survey replications may be subject to unknown and varying levels of biases. In other words, if the ignorable mode effects assumption is violated, the bias properties of the estimators that combine survey responses from different modes without an adjustment are not known. To address this issue inferences based on data from mixed-mode surveys should incorporate mode effect evaluations and adjustments.

3.2. Evaluation of Mode Effects in Mixed-Mode Surveys

As illustrated in (1.2), $\bar{Y}$ combines conditional means of responses from telephone and in-person means based on the mode choice and that is what is available from a mixed-mode survey design. Therefore, mode effects cannot be evaluated simply by statistical tests for differences in parameter estimates based on the data collected using the different modes available in a mixed-mode survey design. For example, in the 2012 March CPS the differences in mean personal income and percent health insurance coverage by survey response mode may result from a combination of mode choice and potential mode effects. As a result, the ignorability of mode effects cannot be evaluated simply by comparing mean differences for the two response modes given the mixed-mode survey design without accounting for the mode choice.

In an attempt to account for mode choice in mode effects evaluations and mode effects adjustments in statistical inference methods, this dissertation proposes to use multiple imputation methods. Although methods can be also applied by implementing single imputations, multiple imputations are used to estimate the variance of the estimators. In mixed-mode surveys for a given person only one mode condition is observed and any inference related to mode effects includes a speculation about how the respondent would have responded in the other mode (Rosenbaum & Rubin, 1984). Following this counterfactual approach, the proposed methods analytically control for mode choice and impute mode-specific data for the complete sample under each alternative mode. These mode-specific data, which are a combination of observed and imputed data, are used to estimate mode-specific population means, $\bar{Y}_p^*$, where p denotes phase and mode, interchangeably.

In the presence of nonignorable mode effects, the mode-specific data estimates may then be adjusted to an internal standard, i.e. best mode, or to an external standard if one is available. The adjusted mode-specific data estimates are combined to produce an adjusted estimate and confidence interval in the next step. Alternative combination methods are further discussed in the following sections.

An exact definition of ignorable mode effects has intentionally not been used. This is because of the possible differences in the availability of data in real-life situations. For example, in this dissertation there are three settings which impose different restrictions on the mode effects evaluations. In the first setting, samples from a finite population with known true values have been drawn. The evaluations of the performance of the alternative estimators were based on the relative differences. In the second setting, hypothetical populations were created based on the real survey data from a study that used 1973 CPS data matched to tax return data from the U.S. Internal Revenue Service (IRS) (Greenlees et al., 1982). The hypothetical populations preserved the 1973 CPS match data mode choice and mode effects conditions but used an artificially generated analysis variable. In this setting, individual benchmark values were available. Mode effects were evaluated in regression models which were fit on the individual level differences. F-tests for the mode main effects and interactions with the group identifiers were used to evaluate mode effects. In the third setting, individual and population level benchmarks were not available. As an alternative to F-tests, repeated measurement ANOVA overall tests were used to detect substantial differences between estimates.

Section 3.3 further discusses the measurement model (1.1) and mean estimator (1.2) in the context of proposed methods. Furthermore, in controlling the mode choice, two imputation models as described in Section 3.3.3 are applied:

1) Mode choice is dependent on the available covariates but each person has a particular mode that is used for responding (ignorable

mode choice). In this case, the mode choice is ignorable since under the choice model the expected value of R_{Tj} and R_{Ij} in (1.1) depends on known covariates only and can be adjusted for. For this set of the computations U_T and U_I are considered to be fixed sets.

2) Mode choice is dependent on the available covariates and the distribution of the survey variable of interest (nonignorable mode choice). In this non-ignorable mode choice, the mean of R_{Tj} and R_{Ij} depends on known covariates and on the variable of interest.

3.3. Multiple Imputation Methods

3.3.1. Response and Choice Models in Mixed-Mode Surveys

Under multiple theoretical realizations of response, R_{Tj} and R_{Ij} can be considered as random variables in (2.8) in which case a more elaborate formulation can be considered.

Suppose $R_{Tj} = 1 - R_{Ij}$, R_{Ij} is a random variable with $\mathrm{E}_R(R_{Tj}) = g(X_j;\psi) \equiv g_j$ where $g(\bullet)$ is logistic, probit, or some other binary regression equation. Note that X_j is a vector of covariates for person j that can contain dummies for social-demographic group and ψ is a vector of regression model parameters.

Suppose $\mu_j = \mu(X_j, \beta^{(\mu)})$ where $\beta^{(\mu)}$ and ψ are different parameters. Rewriting (1.1) gives

$$y_j = \mu(X_j, \beta^{(\mu)}) + R_{Tj}B_{Tj} + R_{Ij}B_{Ij} + \varepsilon_j \tag{3.1}$$

where B_{Tj} and B_{Ij} are mode effects associated with person j. The mode effects can differ among persons with this formulation. In this section, the group level notation used in Section 1.3 is not used. (Alternatively, we could have written $B_{Tj} = Z_j^T B_T$ and $B_{Ij} = Z_j^T B_I$ where Z_j is a vector of covariates for case j, and B_T and B_T are mode effect parameters. This formulation also allows persons to have different mode effects.)

3.3.2. Expectations of Response and Choice Models in Mixed-Mode Surveys

In this section, E_M denotes expectation with respect to the response model (Y-model) and E_R denotes expectation with respect to the mode choice model (R-model).

3.3.2.1. Mean Estimator Ignoring Mode Effect

The combined expectation for unit *j* over the mode choice and response models is:

$$\begin{aligned}\mathrm{E}_M\mathrm{E}_R[y_j] &= \mathrm{E}_M[\mu_j + g_j B_{Tj} + (1-g_j)B_{Ij} + \varepsilon_j] \\ &= \mu_j + g_j B_{Tj} + (1-g_j)B_{Ij}\end{aligned} \tag{3.2}$$

The finite population mean is:

$$\bar{Y} = \frac{1}{N}\sum_{j\in U} y_j$$

where U is the finite population. (3.3)

Under (3.1) the combined expectation of $\bar{Y}$ is

$$\begin{aligned}\mathrm{E}_M\mathrm{E}_R[\bar{Y}] &= \frac{1}{N}\sum_{j\in U}\mathrm{E}_M\mathrm{E}_R[y_j] \\ &= \frac{1}{N}\sum_{j\in U}\left[\mu_j + g_j B_{Tj} + (1-g_j)B_{Ij}\right] \\ &= \underbrace{\frac{1}{N}\sum_{j\in U}\mu_j}_{\bar{\mu}_U} + \underbrace{\frac{1}{N}\sum_{j\in U}\left[g_j B_{Tj} + (1-g_j)B_{Ij}\right]}_{(\overline{gB})_U} \\ &= \bar{\mu} + (\overline{gB})_U\end{aligned} \tag{3.4}$$

where $\bar{\mu}$ is the average expected value in population and $(\overline{gB})_U$ is the average combined effect of random mode choice g_j and mode effect (B_{Tj}, B_{Ij}). Thus, the bias of $\bar{Y}$ is $\mathrm{E}_M \mathrm{E}_R[\bar{Y} - \bar{\mu}] = (\overline{gB})_U$. The estimation error of the population mean therefore is generally not estimable in mixed-mode surveys, because estimating B_{Tj} and B_{Ij} requires "truth" for each unit and the truth is not usually available. It is possible to estimate $g_j = g(X_j; \psi)$ as long as a functional form of g can be specified which reasonably models the probability that a person chooses one mode over another. We denote the estimator that ignores the mode effects by $\bar{Y}_0$ in the next section.

3.3.2.2. Proposed Mean Estimator

The proposed multiple imputation methods impute counterfactual data i.e. as if they had been reported by another mode. For example, when the modes are telephone and in-person (e.g. CPS), two steps are taken. First, a completed data set is produced by imputing in-person respondents as if they had responded by telephone. Second, the telephone respondents are imputed as if they had responded in-person. This approach produces two completed data sets that can be combined in different ways to give an estimate of the population mean. Figure 1.4 shows the schematic chart for the proposed mixed-mode survey inference method.

The next section analyzes the proposed procedure of imputing mode-specific counterfactual data. Similar to the previous sections, the analysis does not account for sampling. Although the analysis is specific to a two-mode design, the discussion can be extended to surveys with more than two modes.

For this situation, we define the following:

$$\bar{Y}_P = \frac{\sum_{j \in U} R_{Pj} y_j}{\sum_{j \in U} R_{Pj}} = \frac{1}{N_P} \sum_{j \in U_P} y_j \text{, where:} \qquad (3.5)$$

$\bar{Y}_P$ is response mean for persons who responded by telephone or in-person in a realization of the survey,

U_T : set of persons with R_{Tj} =1,

U_I : set of persons with R_{Ij} =1.

(3.5) also holds when R_{Tj} and R_{Ij} are considered to be stochastic on the interval (0,1).

Using the same notation as in Chapter 2, the expectation over response model of $\bar{Y}_T$ is

$$\begin{aligned} \mathrm{E}_M[\bar{Y}_T] &= \frac{1}{N_T}\sum_{j\in U_T}(\mu_j + B_{Tj}) \\ &= \underbrace{\frac{1}{N_T}\sum_{j\in U_T}\mu_j}_{\bar{\mu}_T} + \underbrace{\frac{1}{N_T}\sum_{j\in U_T}B_{Tj}}_{\bar{B}_T(T)} \qquad (3.6) \\ &= \bar{\mu}_T + \bar{B}_T(T) \end{aligned}$$

Similarly,

$$\begin{aligned} \mathrm{E}_M[\bar{Y}_I] &= \frac{1}{N_I}\sum_{j\in U_I}(\mu_j + B_{Ij}) \\ &= \underbrace{\frac{1}{N_I}\sum_{j\in U_I}\mu_j}_{\bar{\mu}_I} + \underbrace{\frac{1}{N_I}\sum_{j\in U_I}B_{Ij}}_{\bar{B}_I(I)} \qquad (3.7) \\ &= \bar{\mu}_I + \bar{B}_I(I) \end{aligned}$$

Note that $E_M\left(\bar{Y}_T - \bar{Y}_I\right) = (\bar{\mu}_T - \bar{\mu}_I) + \left[\bar{B}_T(T) - \bar{B}_I(I)\right]$ does not estimate the difference in actual means or the mode effect. Also, notice that this expression in similar to but not exactly the same as the expected selection effects in (2.16) and (2.18).

For persons who responded by I , we impute values as if they had responded by T and the reverse for persons who responded by *T*:

y^*_{Tj} $(j \in U_I)$ is the imputed telephone value for persons who responded by I

y^*_{Ij} $(j \in U_T)$ is the imputed in-person value for persons who responded by T

Assume $U = U_T \cup U_I$ is the full population. To do the imputations, we will use the telephone reports to create the y^*_{Tj} imputations for the cases that respond in-person. Similarly, the in-person reports will be used to create imputations for the cases that respond by telephone. In those circumstances, it is reasonable to suppose that, on average, the imputations for a set of cases (U_T or U_I) are contaminated by the reporting errors associated with the cases used to create the imputations. In particular, suppose that the expectations with respect to the imputation mechanism are

$$E_{IMP}\left[y^*_{Tj}\right] = \mu_j + B_{Tj} \text{ and}$$

$$E_{IMP}\left[y^*_{Ij}\right] = \mu_j + B_{Ij}. \tag{3.8}$$

Define the means that use imputed data as

$$\bar{Y}^*_T = \frac{1}{N}\left[\sum_{j \in U_T} y_j + \sum_{j \in U_I} y^*_{Tj}\right] \tag{3.9}$$

$$\bar{Y}^*_I = \frac{1}{N}\left[\sum_{j \in U_I} y_j + \sum_{j \in U_T} y^*_{Ij}\right] \tag{3.10}$$

In the next section, we discuss ways of combining $\bar{Y}^*_T$ and $\bar{Y}^*_I$ to estimate the population mean.

3.3.2.3. Estimation Errors for Alternative Estimators

As before E_M denotes expectation with respect to the response model (*Y*-model), E_R is the expectation over the mode choice model (R-model), and

E_{IMP} denotes expectation with respect to the imputation model. Suppose that $\tilde{y}_j$ is the true value for unit *j* which obeys the model

$$\tilde{y}_j = \mu_j + \delta_j$$

where $\delta_j \overset{iid}{\sim} \left(0, \sigma^2\right)$ are independent error terms.

The estimation error for the mean computed as if all cases responded by telephone is $\bar{Y}_T^* - \bar{Y}$, which can be written as

$$\begin{aligned} \bar{Y}_T^* - \bar{Y} &= \frac{1}{N}[\sum_{j \in U_T} y_j + \sum_{j \in U_I} y_{Tj}^*] - \frac{1}{N}[\sum_{j \in U_T} \tilde{y}_j + \sum_{j \in U_I} \tilde{y}_j] \\ &= \frac{1}{N}[\sum_{j \in U_T} (y_j - \tilde{y}_j) + \sum_{j \in U_I} (y_{Tj}^* - \tilde{y}_j)] \end{aligned} \tag{3.11}$$

The expectation with respect to the *Y*-model and the imputation model, conditional on the sets of units that responded by *T* or *I* is

$$\begin{aligned} E_M E_{IMP}[\bar{Y}_T^* - \bar{Y} \mid U_T, U_I] &= P_T \frac{1}{N_T}\left\{\sum_{j \in U_T} B_{Tj}\right\} + P_I \frac{1}{N_I}\left\{\sum_{j \in U_I} B_{Tj}\right\} \\ &= \frac{1}{N}\sum_{j \in U} B_{Tj} \equiv \bar{B}_{UT} \end{aligned} \tag{3.12}$$

where $P_T = N_T / N$, $P_I = N_I / N$. The expectation of the first term in (3.11) is

$$E_M[(y_j - \tilde{y}_j) \mid U_T, U_I] = \mu_j + B_{Tj} - \mu_j = B_{Tj}$$

Similarly, the expectation of the second term in (3.11) is

$$E_M E_{IMP}[(y_{Tj}^* - \tilde{y}_j) \mid U_T, U_I] = B_{Tj}.$$

Consequently, the expectation of (3.11) reduces to

$$\begin{aligned} E_M E_{IMP}[\bar{Y}_T^* - \bar{Y} \mid U_T, U_I] &= P_T \frac{1}{N_T}\left\{\sum_{j \in U_T} B_{Tj}\right\} + P_I \frac{1}{N_I}\left\{\sum_{j \in U_I} B_{Tj}\right\} \\ &= \frac{1}{N}\sum_{j \in U} B_{Tj} \equiv \bar{B}_{UT} \end{aligned} \tag{3.13}$$

Thus, the imputations for U_I are contaminated by the telephone mode effect, and conditional on the realized modes selected by respondents, the imputed estimate $\bar{Y}_T^*$ inherits the average reporting error associated with the telephone mode.

Similar, to (3.13), the expectation of the mean as if all cases had responded in person is

$$\begin{aligned} \mathrm{E}_M \mathrm{E}_{IMP}[\bar{Y}_I^* - \bar{Y} \mid U_T, U_I] &= P_I \frac{1}{N_I}\left\{\sum_{j \in U_I} B_{Ij}\right\} + P_T \frac{1}{N_T}\left\{\sum_{j \in U_T} B_{Ij}\right\} \\ &= \frac{1}{N}\sum_{j \in U} B_{Ij} \equiv \bar{B}_{UI} \end{aligned} \tag{3.14}$$

$\bar{B}_{UI}$ in (3.14) is the average mode effect in the population if all cases responding by I .

To remove the condition on U_T and U_I in (3.13), the expectation over the mode choice model (R-model) will be taken. First, B_{UT} can be written as

$$\begin{aligned} \bar{B}_{UT} &= \frac{1}{N}\sum_{j \in U} B_{Tj} \\ &= \frac{1}{N}\left[\sum_{j \in U} R_{Tj} B_{Tj} + \sum_{j \in U} (1 - R_{Tj}) B_{Tj}\right] \end{aligned} \tag{3.15}$$

Since $\mathrm{E}_R[R_{Tj}] = \Pr(j \in U_T) = g_j$, we have

$$\begin{aligned} \mathrm{E}_R[\bar{B}_{UT}] &= \frac{1}{N}\left[\sum_{j \in U} g_j B_{Tj} + \sum_{j \in U} \left(1 - g_j\right) B_{Tj}\right] \\ &= \bar{B}_{UT} \end{aligned} \tag{3.16}$$

Note that $\bar{B}_{UT}$ differs from the means, $\bar{B}_I(T), \bar{B}_I(I), \bar{B}_T(T),$ and $\bar{B}_T(I)$ defined in Chapter 2. Similarly,

$$\bar{B}_{UI} = \frac{1}{N} \sum_{j \in U} B_{Ij}$$

$$= \frac{1}{N} \left[\sum_{j \in U} (1 - R_{Ij}) B_{Ij} + \sum_{j \in U} R_{Ij} B_{Ij} \right] \tag{3.17}$$

and

$$\begin{aligned} \mathrm{E}_R[\bar{B}_{UI}] &= \frac{1}{N} \left[\sum_{j \in U} g_j B_{Ij} + \sum_{j \in U} \left(1 - g_j\right) B_{Ij} \right] \\ &= \bar{B}_{UI} \end{aligned} \tag{3.18}$$

On the other hand, the procedure that ignores mode is:

$$\bar{Y}_0 = \frac{1}{N} \left[\sum_{j \in U_T} y_j + \sum_{j \in U_I} y_j \right] \tag{3.19}$$

Its estimation error is

$$\bar{Y}_0 - \bar{Y} = \frac{1}{N} \left[\sum_{j \in U_T} (y_j - \tilde{y}_j) + \sum_{j \in U_I} (y_j - \tilde{y}_j) \right] \tag{3.20}$$

with expectation

$$\mathrm{E}_M[\bar{Y}_0 - \bar{Y} \mid U_T, U_I] = \frac{1}{N} \left[\sum_{j \in U_T} B_{Tj} + \sum_{j \in U_I} B_{Ij} \right] \tag{3.21}$$

If we remove the condition on U_T and U_I and rewrite (3.21):

$$\begin{aligned} \mathrm{E}_R \mathrm{E}_M[\bar{Y}_0 - \bar{Y}] &= \frac{1}{N} E_R \left\{ \sum_{j \in U} \left[R_{Tj} B_{Tj} + \left(1 - R_{Tj}\right) B_{Ij} \right] \right\} \\ &= \frac{1}{N} \left\{ \sum_{j \in U} \left[g_j B_{Tj} + \left(1 - g_j\right) B_{Ij} \right] \right\} \\ &= \overline{(gB)}_U \end{aligned}$$

as shown earlier in (3.4).

Thus, $\mathrm{E}_R\mathrm{E}_M[\bar{Y}_0 - \bar{Y}]$ is a weighted average of telephone and in-person mode effects. The sizes of the mode effects is usually unknown, i.e. it may not be known whether $\bar{B}_{UT} < \overline{(gB)}_U$ or $\bar{B}_{UI} < \overline{(gB)}_U$. We will have $\bar{B}_{UI} \le \overline{(gB)}_U \le \bar{B}_{UT}$ or the reverse depending on which of $\bar{B}_{UI}$ or $\bar{B}_{UT}$ is smaller since $g_j B_{Tj} + (1-g_j)B_{Ij}$ is a convex combination.

As an alternative estimator, we propose

$$\bar{Y}^* = \alpha\bar{Y}_T^* + (1-\alpha)\bar{Y}_I^*,\ 0 \le \alpha \le 1 \tag{3.22}$$

The bias of $\bar{Y}^*$ is $\mathrm{E}_R\mathrm{E}_M\mathrm{E}_{IMP}\left[\bar{Y}^* - \bar{Y}\right] = \alpha\bar{B}_{UT} + (1-\alpha)\bar{B}_{UI}$, implying that the bias of $\bar{Y}^*$ depends on α and sizes of average mode effects.

The smallest bias would be obtained by choosing the mode with the smaller bias and using only $\bar{Y}_T^*$ or $\bar{Y}_I^*$ but this would waste the data collected via the other mode.

Note that $\mathrm{E}_R\mathrm{E}_M\mathrm{E}_{IMP}\left[\bar{Y}_T^* - \bar{Y}_I^*\right] = \bar{B}_{UT} - \bar{B}_{UI}$. So $\bar{Y}_T^* - \bar{Y}_I^*$ can be used to estimate the difference in the average difference in mode effects in the population. This difference incorporates the possibility that the effect of mode can differ among persons.

Next, we examine different choices for the combining weight α. Define $v_T = \mathrm{var}_M(\bar{Y}_T^*)$, $v_I = \mathrm{var}_M(\bar{Y}_I^*)$. To find the minimum mean square error (MSE) combination for $\bar{Y}^*$, first note that the MSE is equal to

$$MSE(\bar{Y}^*) = \alpha^2 v_T + (1-\alpha)^2 v_I + 2\alpha(1-\alpha)\underbrace{\mathrm{cov}(Y_T^*, Y_T^*)}_{C_{TI}} + [\alpha\bar{B}_{UT} + (1-\alpha)\bar{B}_{UI}]^2$$

The first derivative with respect to α is

$$\frac{\partial MSE(\bar{Y}^*)}{\partial \alpha} = 2\alpha v_T - 2(1-\alpha)v_I + 2(1-2\alpha)C_{TI} + 2[\alpha \bar{B}_{UT} + (1-\alpha)\bar{B}_{UI}][\bar{B}_{UT} - \bar{B}_{UI}]$$
$$\propto \alpha v_T - (1-\alpha)v_I + (1-2\alpha)C_{TI} + [\bar{B}_{UI} + \alpha(\bar{B}_{UT} - \bar{B}_{UI})][\bar{B}_{UT} - \bar{B}_{UI}]$$
$$= \alpha v_T - v_I + \alpha v_I + C_{TI} - 2\alpha C_{TI} + \bar{B}_{UI}(\bar{B}_{UT} - \bar{B}_{UI}) + \alpha(\bar{B}_{UT} - \bar{B}_{UI})^2$$
$$= \alpha \left[\underbrace{(v_T + v_I - 2C_{TI})}_{Var_M(Y_T^* - Y_I^*)} + (\bar{B}_{UT} - \bar{B}_{UI})^2 \right] - v_I + C_{TI} + \bar{B}_{UI}(\bar{B}_{UT} - \bar{B}_{UI})$$

To find the optimal α, set the derivative equal to 0 and solve, giving

$$\alpha \left[Var_M(Y_T^* - Y_I^*) + (\bar{B}_{UT} - \bar{B}_{UI})^2 \right] = v_I - C_{TI} - \bar{B}_{UI}(\bar{B}_{UT} - \bar{B}_{UI}) \text{ and}$$

$$\alpha_{opt} = \frac{v_I - C_{TI} - \bar{B}_{UI}(\bar{B}_{UT} - \bar{B}_{UI})}{\left[Var_M(Y_T^* - Y_I^*) + (\bar{B}_{UT} - \bar{B}_{UI})^2 \right]}$$
$$= \frac{v_I - \rho_{TI}\sqrt{v_T v_I} - \bar{B}_{UI}(\bar{B}_{UT} - \bar{B}_{UI})}{\left[v_I + v_T + 2\rho_{TI}\sqrt{v_T v_I} + (\bar{B}_{UT} - \bar{B}_{UI})^2 \right]} \tag{3.23}$$

where $\rho_{TI} = corr(\bar{Y}_T^*, \bar{Y}_I^*)$. In general, α_{opt} decreases as $\bar{B}_{UI}$ decreases and increases as v_I increases.

For each component of α_{opt} we can further simplify the form in some special cases:

Using $C_{TI} = \rho_{TI}\sqrt{v_T v_I}$ and $v_T = v_I$

$$v_I - C_{TI} = v_I - \rho_{TI} v_I = v_I(1 - \rho_{TI}) \tag{3.24}$$

$$Var_M(Y_T^* - Y_I^*) = 2v_I(1 - \rho_{TI}) \tag{3.25}$$

When we plug (3.24) and (3.25) into (3.23), the optimal weight becomes

$$\alpha_{opt} = \frac{v_I(1-\rho_{TI}) - \bar{B}_{UI}(\bar{B}_{UT} - \bar{B}_{UI})}{\left[2v_I(1-\rho_{TI}) + (\bar{B}_{UT} - \bar{B}_{UI})^2 \right]}$$

If there is no mode effect, $\bar{B}_{UT} = \bar{B}_{UI} = 0$. This is also true if individual mode effects average out to be zero. Having no mode effect implies that $\alpha_{opt} = \frac{1}{2}$, which makes sense when $v_T = v_I$. If $v_T \neq v_I$ but $\bar{B}_{UT} = \bar{B}_{UI}$, then

$$\alpha_{opt} = \frac{v_I - \rho_{TI}\sqrt{v_T v_I}}{v_I + v_T + 2\rho_{TI}\sqrt{v_T v_I}}.$$

If $\rho_{TI} = 0$ then $\alpha_{opt} = \frac{v_I}{v_I + v_T}$. It is the prescription to "weight inversely according to the variance" described in Section 3.3.4.

3.3.3. Imputation models

There are two model applications in controlling the mode choice: (1) ignorable mode choice in which mode choice is dependent on the available covariates, and (2) nonignorable mode choice in which the selection of mode is dependent on the available covariates and the distribution of the survey variable of interest.

3.3.3.1. Ignorable Mode Choice Imputation Models

In the ignorable mode choice imputation the following modeling and imputations steps were followed for the continuous and binary variables. In the applications the continuous variable Y is total family income or personal income and the binary variable Y is health insurance coverage. X is the matrix of available household and householder characteristics.

Ignorable Mode Choice Imputation Model for Continuous Variables: The usual noninformative prior distribution normal linear regression model motivates the ignorable mode choice imputation model. A special case of ignorable mode choice imputation model is applied in which U_T and U_I are fixed sets. In the model parameterizations, subscript p is ignored for simplication purposes. The normal linear regression model is $Y_j \sim N(X_j\beta, \sigma^2)$. Assuming the standard noninformative prior distribution $\Pr(\beta, \log\sigma) \propto \text{constant}$ or, equivalently $\Pr(\beta, \sigma \mid X) \propto \frac{1}{\sigma^2}$, the conditional posterior distribution for β is $(\beta \mid \sigma^2, y) \sim MVN(\hat{\beta}, (X^T X)^{-1}\sigma^2)$ where $\hat{\beta} = (X^T X)^{-1} X^T Y$. The marginal posterior distribution of σ^2 is $\Pr(\sigma^2 \mid y) \sim Inv-\chi^2(n-k, s^2)$ where $s^2 = \frac{1}{n-k}(y - X\hat{\beta})^T (y - X\hat{\beta})$, n is the sample size and k is the number of parameters. Given these posterior distributions of the regression parameters, y's for the alternative mode respondents can be drawn from $y_j^* = N(x_j\hat{\beta}, \hat{\sigma}^2 \nu)$, where ν is a random variable.

In the computations M=5 completed data sets were saved. While M=5 is often used, more recent evidence shows that a greater number of imputations is required when the missing fraction is high (Graham, Olchowski, & Gilreath, 2007). In this dissertation, M=5 is used, but the value of M (number of imputations) will be determined empirically in the future extensions of this research. The multiple imputation method for imputing y's for the respondents to the alternative mode follows these steps:

1. Compute $y_j^* = X_j^{(Y)} \hat{\beta}^{(Y)} + \hat{\sigma}\varepsilon_j$, where $\varepsilon_j \overset{iid}{\sim} N(0,1)$
2. Save the y_j^* as the imputed value for observation j.

Due to computational demands the current implementation of the multiple imputation does not use multiple draws of $(\hat{\beta}^{(Y)}, \hat{\sigma}^2)$, but multiple draws of $(\hat{\beta}^{(Y)}, \hat{\sigma}^2)$ will be incorporated in the future implementations.

Ignorable Model Choice Imputation Model for Binary Variables: The model that will be used for a binary variable is

$$\text{logit}[\Pr(Y_j = 1 \mid X_j)] = X_j \beta^{(Y)} \text{ where } \text{logit}(a) = \log\left[\frac{a}{1-a}\right] \quad (3.26)$$

Using the large-sample normal approximation and noninformative prior, the posterior distribution of $\beta^{(Y)} \sim MVN(\hat{\beta}^{(Y)}, V(\hat{\beta}^{(Y)}))$ where $\hat{\beta}^{(Y)}$ are the maximum likelihood estimates for (3.26) and $V(\hat{\beta}^{(Y)})$ is the inverse of the information matrix evaluated at $\hat{\beta}^{Y}$. The multiple imputation method for imputing *y*'s for the alternative mode respondents follows the steps:

1- Compute Cholesky decomposition of $\hat{V}(\hat{\beta}^{(Y)})$, denoted by TT' (Instead of (Draw β^{*} from $MVN(\hat{\beta}^{(Y)}, V(\hat{\beta}^{(Y)}))$). Generate *p* normal deviates z where *p* is the length of $\hat{\beta}^{(Y)}$ and construct $\beta_{*}^{(Y)} = \hat{\beta}^{(Y)} + Tz$.

2- Compute $p_j^{*} = \dfrac{\exp(X_j \beta_{*}^{Y})}{1 + \exp(X_j \beta_{*}^{Y})}$ for the alternative mode respondents.

3- Draw u_j independently from a uniform(0,1) distribution, and if $u_j > p_j^{*}$ impute y_j^{*}=0, otherwise y_j^{*}=1.

The steps 1-3 are repeated M times, and M completed data sets are saved.

3.3.3.2. Nonignorable Mode Choice Imputation Models

In this case, nonignorable nonresponse models are extended to impute data for alternative mode data (Glynn, Laird, & Rubin, 1993; Greenlees et al., 1982; Little & Rubin, 2002). There are three nonignorable nonresponse models that the Bayesian framework distinguishes: (1) selection models (Heckman, 1979), (2) pattern-mixture models (Glynn, Laird, & Rubin, 1986; Glynn et al., 1993;Little, 1993) and (3) pattern-set mixture models (Little, 1993; Little & Rubin, 2002). Greenlees, Reece, and Zieschang (1982) (from now on denoted by GRZ) used a Bayesian selection model to impute nonrespondent data. Glynn, Laird and Rubin (1986) extended the selection bias model to include the follow-up nonrespondent data.

Most of the literature imposes the normality assumption on Y in the selection bias modeling (Greene, 2011; Rubin, 1987; Little & Rubin, 2002) although many important variables collected in surveys are non-normal.

Selection bias models can be estimated by two-step Heckman or maximum likelihood methods. While the two-step Heckman method is the most common estimation method in the literature, there are problems cited for this method (Greene, 2011; Stolzenberg & Relles, 1997). The method in this thesis applies the maximum likelihood estimation approach as described in the next section.

Nonignorable Model Choice Imputation Model for Continuous Variables: For the nonignorable mode choice imputation model, the full likelihood (shown in (3.29) below) is built up by multiplying the likelihoods for respondents to alternative modes (for example, $p = T$ and I in CPS design). The likelihood functions are conditioned on the selection mechanism and the distributional assumption for the response variable (Greenlees et al., 1982) as shown in (3.27) and (3.28). Suppose that $X_j^{(R)}$ and $X_j^{(Y)}$ are the covariates on which we condition the mode choice mechanism and response variable respectively. Again in these parameterizations, subscript p is ignored for simplification purposes. $\beta^{(R)}$ and $\beta^{(Y)}$ are the model parameters for the selection (mode choice model) and regression (response model) equations, respectively. Assuming a normal distribution for the response variable (3.27) and a logistic function for the mode choice mechanism (3.28), the full likelihood function for the telephone mode is as follows:

$$(Y_j \mid X_j^{(Y)};\theta) \sim N(X_j^{(Y)}\beta^{(Y)}, \sigma^2) \text{ where } \theta = (\beta^{(Y)}, \sigma^2) \tag{3.27}$$

$$\Pr(R_{Tj} = 1 \mid X_j^{(R)}, Y_j; \psi) = \left[1 + \exp(-X_j^{(R)}\beta^{(R)} - \gamma Y_j)\right]^{-1} \text{ where}$$
$$\psi = (\beta^{(R)}, \gamma) \tag{3.28}$$

$$L_{full}(\theta, \psi \mid Y, R_{Tj} = 1) = \prod_{j \in U_T} \frac{1}{\left[1 + \exp(-X_j^{(R)}\beta^{(R)} - \gamma Y_j)\right]} \frac{1}{\sigma} \Phi\left(\frac{Y_j - X_j^{(Y)}\beta^{(Y)}}{\sigma}\right) \times$$
$$\prod_{j \in U_I} \int_{-\infty}^{\infty} \left(1 - \frac{1}{\left[1 + \exp(-X_j^{(R)}\beta^{(R)} - \gamma Y_j)\right]}\right) \frac{1}{\sigma} \Phi\left(\frac{Y_j - X_j^{(Y)}\beta^{(Y)}}{\sigma}\right) dy_j \tag{3.29}$$

The full likelihood function is the multiplication of the conditional density function of the mode respondents $j \in U_T$ and alternative mode respondents $j \in U_I$. Since Y is not observed for the alternative mode respondents $j \in U_I$, integration over the density function is used. Following this mechanism, a full likelihood function can also be fit for the in-person mode in which Y is observed for the in-person mode respondents $j \in U_I$ and Y is not observed for the alternative mode respondents $j \in U_T$.

Given the full likelihood function (3.29), maximum likelihood estimation was performed using the quasi-Newton method with the BFGS algorithm (Broyden, Dennis Jr., & More, 1973; Broyden, 1969; Fletcher, 1970; Goldfarb, 1970; Shanno & Kettler, 1970) as employed by R-programming software. Since the integral cannot be exactly evaluated, it is approximated by ten-point Gauss-Hermite quadrature.

The GRZ approach allows imputation of the expected values of y_j conditional on X and mode choice. While in their imputation model, GRZ used the conditional expectation for y_j's which was sufficient for their statistical analysis, they suggested drawing values from the Y distribution conditioned on $X_j^{(R)}$ and $X_j^{(Y)}$ values to do the actual imputations. This will avoid underestimation of the Y variance. The maximum likelihood estimates $(\hat{\beta}^{(Y)}, \hat{\sigma}^2, \hat{\beta}^{(R)}, \hat{\gamma})$ given (3.29) were plugged into an imputation model (Greenlees et al., 1982) as follows:

1. Draw $(Y_j \mid X_j^{(Y)}; \theta) \sim N(X_j^{(Y)}\hat{\beta}^{(Y)}, \hat{\sigma}^2)$
2. Compute $\hat{y}_j = X_j^{(Y)}\hat{\beta}^{(Y)} + \hat{\sigma}\varepsilon_j$, where $\varepsilon_j \overset{iid}{\sim} N(0,1)$
3. Compute
$$\Pr(R_{pj} = 0 \mid \hat{y}_j, X_j^{(R)}, \hat{\beta}^{(R)}, \hat{\gamma}) = 1 - \frac{1}{[1 + \exp(-X_j^{(R)}\hat{\beta}^{(R)} - \hat{\gamma}\hat{y}_j]}.$$
4. Draw a random number η from a uniform distribution $[0,1]$.

5. Save the $\hat{y}_j$ as the imputed value for observation j if $\Pr(R_{pj} = 0 \mid \hat{y}_i, X_j^{(R)}, \hat{\beta}^{(R)}, \hat{\gamma}) \geq \eta$; otherwise repeat the imputation steps 1-5.

The current implementation does not use multiple draws of $(\hat{\beta}^{(Y)}, \hat{\sigma}^2, \hat{\beta}^{(R)}, \hat{\gamma})$, but multiple draws of $(\hat{\beta}^{(Y)}, \hat{\sigma}^2, \hat{\beta}^{(R)}, \hat{\gamma})$ will be incorporated in the future implementations.

Nonignorable Mode Choice - Multiple Imputation Selection Models for Binary Variables:

A proposed imputation method for a binary variable, following a bivariate probit model, is described in (3.30) and (3.31) (Greene, 2011).

$$\text{Selection equation: } R_{pj}^{*} = X_j^{(R)}\beta^{(R)} + \varepsilon_j^{(R)},\ R_{pj} = \begin{cases} 1, \text{ if } R_{pj}^{*} \leq 0, \\ 0, \text{ if } \mathrm{R}_{pj}^{*} > 0 \end{cases} \tag{3.30}$$

$$\text{Regression model: } Y_j^{*} = X_j^{(Y)}\beta^{(Y)} + \varepsilon_j^{(Y)}, \quad Y_j = \begin{cases} 1, \text{ if } \mathrm{Y}_j^{*} \leq 0, \\ 0, \text{ if } \mathrm{Y}_j^{*} > 0 \end{cases} \text{ and}$$

$$\begin{pmatrix} \varepsilon_j^{(R)} \\ \varepsilon_j^{(Y)} \end{pmatrix} \sim N\left(\underset{\sim}{0}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \tag{3.31}$$

The estimation of the parameters $(\beta^{(R)}, \beta^{(Y)}, \rho)$ is done through the computation of multivariate normal probabilities with arbitrary correlation matrices as implemented in the R function `pmvnorm` (Genz, 1992, 1993). Using the large-sample normal approximation and noninformative prior, the posterior distribution of $\beta^{(R)} \sim MVN(\hat{\beta}^{(R)}, V(\hat{\beta}_{(R)}))$ and the posterior distribution of $\beta^{(Y)} \sim MVN(\hat{\beta}^{(Y)}, V(\hat{\beta}_{(Y)}))$. $\hat{\beta}^{(R)}$ and $\hat{\beta}^{(Y)}$ are the maximum likelihood estimates for (3.30) and (3.31); $V(\hat{\beta}_{(R)})$ and $V(\hat{\beta}_{(Y)})$ are the inverse of the information matrix evaluated at $\hat{\beta}_{(R)}$ and $\hat{\beta}_{(Y)}$. Given these posterior distributions of the regression parameters, the multiple imputation method for imputing y's for the alternative mode respondents follows the steps:

1. Draw randomly $\varepsilon_j^{*(R)}$ and $\varepsilon_j^{*(Y)}$ from $N\left(\underset{\sim}{0}, \begin{bmatrix} 1 & \hat{\rho} \\ \hat{\rho} & 1 \end{bmatrix}\right)$.
2. Compute Cholesky decompositions of $\hat{V}(\hat{\beta}^{(R)}))$, $\hat{V}(\hat{\beta}^{(Y)}))$, denoted by $TT'^{(R)}$ and $TT'^{(Y)}$ (Instead of drawing randomly $\beta^{*(R)}$ and $\beta^{*(Y)}$ from $MVN(\hat{\beta}^{(M)}, V(\hat{\beta}_{(M)}))$ and $MVN(\hat{\beta}^{(Y)}, V(\hat{\beta}_{(Y)}))$ respectively). Generate $p^{(R)}$ and $p^{(Y)}$ $z^{(R)}$ and $z^{(Y)}$ normal deviates where $p^{(R)}$ and $p^{(Y)}$ are the length of $\hat{\beta}^{(R)}$ and $\hat{\beta}^{(Y)}$. Construct $\beta_*^{(R)} = \hat{\beta}^{(R)} + T^{(R)} z^{(R)}$ a $\beta_*^{(Y)} = \hat{\beta}^{(Y)} + T^{(Y)} z^{(Y)}$
3. Compute $\hat{y}_j = X_j^{(Y)} \hat{\beta}^{*(Y)} + \varepsilon_j^{*(Y)}$.
4. Compute $\hat{R}_{pj} = X_j^{(R)} \hat{\beta}^{*(R)} + \varepsilon_j^{*(R)}$.
5. Draw $u_j^{(R)}$ and $u_j^{(Y)}$ independently from uniform(0,1) distributions, and if $u_j^{(R)} > \hat{R}_{pj}$ and $u_j^{(Y)} > \hat{y}_j$ where p denotes a mode.
6. For nonreporting units for a given p, i.e. $R_{pj} = 0$:
   a. If $u_j^{(R)} > \hat{R}_{pj}$ and $u_j^{(Y)} > \hat{y}_j$ then save $y_j^* = 0$;
      If $u_j^{(R)} > \hat{R}_{pj}$ and $u_j^{(Y)} <= \hat{y}_j$ then save $y_j^* = 1$;
      otherwise repeat the imputation steps 1-6.

The steps 1-3 are repeated M times, and M completed data sets are saved. The current implementation does not use multiple draws of $(\hat{\beta}^{(Y)}, \hat{\rho}, \hat{\beta}^{(R)})$, but multiple draws of $(\hat{\beta}^{(Y)}, \hat{\rho}, \hat{\beta}^{(R)})$ will be incorporated in the future implementations (Chib & Greenberg, 1998).

3.3.4. Special Cases of α : Alternative Empirical Combination Methods

Mode-specific mean estimates, $\bar{Y}_T^*$ and $\bar{Y}_I^*$, are computed based on both observed and imputed counterfactual data using the multiple imputation technique. As defined in (3.22), a proposed estimator combines mode-specific mean estimates:

$$\bar{Y}^* = \alpha \bar{Y}_T^* + (1-\alpha)\bar{Y}_I^*, \ 0 \le \alpha \le 1$$

$$\alpha_{opt} = \frac{v_I(1-\rho_{TI}) - \bar{B}_{UI}(\bar{B}_{UT} - \bar{B}_{UI})}{\left[2v_I(1-\rho_{TI}) + (\bar{B}_{UT} - \bar{B}_{UI})^2\right]}$$ which minimizes the MSE($\bar{Y}^*$)

Other special cases are as follows:

Method 1 (CM_1) – Simple average estimator

$$\bar{Y}^* = \sum_P w_{p,CM_1}^* \bar{Y}_p^*, \text{ where } w_{p,CM_1}^* = \frac{1}{2}$$

Method 2 (CM_2) – Weighted inversely according to the variances of the estimated means

$$\bar{Y}^* = \sum_P w_{p,CM_2}^* \bar{Y}_p^*, \text{ where } w_{p,CM_2}^* = \frac{\frac{1}{Var(\bar{Y}_p^*)}}{\sum_P \frac{1}{Var(\bar{Y}_p^*)}}$$

Method 3 (CM_3) – Weighted inversely according to the mean square errors of the estimated means

$$\bar{Y}^* = \sum_P w_{p,CM_3}^* \bar{Y}_p^*, \text{ where } w_{p,CM_3}^* = \frac{\frac{1}{MSE(\bar{Y}_p^*)}}{\sum_P \frac{1}{MSE(\bar{Y}_p^*)}}$$

where $\bar{Y}_p^*$ are mode specific mean estimates and $p = 1, 2$ corresponds to telephone and in-person modes in the CPS computations.

The bias properties of the described imputation method and the standard method that ignores mode effects are evaluated in the empirical and simulation studies as described in Chapter 5 and Chapter 6. The following chapter describes the data, mode choice and response regression models used in the empirical and the simulation studies.

Chapter 4

Modeling of Mode Choice and Mode Effects in the 1973 Current Population Survey (CPS) Match and 2012 CPS March Data

4.1. Introduction

This chapter has four aims: (1) provide a full description of the datasets that are used in Chapters 5 and 6, (2) investigate the mode choice mechanism in the 1973 CPS Match and 2012 CPS March data using logistic regression models, (3) investigate the mode effects in the 1973 CPS Match data in which unit level benchmarks are available, and (4) describe and present the variable selection for the imputation models. As described in Chapter 3, two kinds of imputation models are considered: (1) ignorable mode choice, and (2) nonignorable mode choice. Ignorable mode choice imputation models include only the response models (Y-model). A linear regression model and a logistic regression model are used for income and health insurance coverage, respectively. Nonignorable mode choice imputation models include both the mode choice model (R-model) and the response model (Y-model). For the mode choice, a logistic regression model and a probit regression model are used for income and health insurance coverage, respectively.

4.2. Data Description

4.2.1. 1973 CPS Match Data[1]

The Current Population Survey (CPS) is a rotating panel survey that produces data on the U.S. labor force. The panel rotation scheme follows a 4-8-4 pattern for a selected household. A sample household is interviewed for two four consecutive months which are eight months apart. CPS is a mixed-mode survey which includes telephone and in-person modes. Except for the first and fifth wave interviews, interviews are mostly conducted by telephone, but for the first and fifth waves the dominant mode is in-person.

In a joint project, the U.S. Census Bureau and Social Security Administration matched the 1973 CPS March data with Social Security benefit and earnings records and released the data to the public. Additionally, a limited set of tax items provided by the Internal Revenue Service (IRS) from the 1972 Federal Income Tax are also available for a subset of respondents in the same dataset.

In addition to the survey mode, there are some other measurement error sources in the CPS data collection, such as proxy reporting, and dependent interviewing, that may contribute into the varying biases. For this investigation, not all the measurement error sources are taken into account. In the Chapter 5 application a subset of data is selected to eliminate other possible measurement errors to a degree. In Chapter 6, an augmented subset data is used as described in this section. Also, since the CPS telephone interviewing was not centralized in 1973, there may be a possibility of greater levels of interviewer-related survey error on the survey estimates. However, the data for the interviewers are not available in this dataset to perform this evaluation.

---

[1] [ICPSR 7616]. ICPSR version. Washington, DC: U.S. Dept. of Commerce, Bureau of the Census and Social Security Administration, Long-Range Research Branch [producer], 197?. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2001. doi:10.3886/ICPSR07616

In the Chapter 6 study, a subset of the 1973 public-use Current Population Survey (CPS) and Social Security Records Exact Match data (which is referred as 1973 CPS match data) are used to create hypothetical populations. This dataset contains a limited set of tax filing items for a subset of respondents, including Adjusted Gross Income (AGI) from the 1972 Federal Income Tax year provided by the Internal Revenue Service (IRS). The IRS data are assumed to be an external standard to use in adjusting survey responses.

The 1973 CPS match data are restricted to the following: records for which both CPS and IRS records are available, specifically primary families, single taxpayers or married taxpayers whose spouse was present and who filed jointly, and those who reported total family income and adjusted gross income (AGI) less than 50,000 USD[2]. Wage and salary income, and total family income are the variables of study in the Chapter 5 and Chapter 6 investigation, respectively. The data exclude records with item nonresponse on any of the original CPS income type and total family income measures. For the simulation in Chapter 6, a *modified total family income* variable is constructed by summing up the eight income types, listed below, that were reported in the CPS March Supplement for the householder and spouse. The original CPS family income was constructed by summing up the income for all the family members who are 15 and older. This construct is referred to as CPS *constructed total family income* in the remainder of the text. Since the modified total family income is the sum of reported income for the householder and spouse (where present), this calculation of income is more comparable to IRS AGI (Form 1040) than the CPS constructed total family income measure, which includes income from all family members age 15 and over.

The variable of study, modified total family income is computed by summing eight income types from CPS: (1) Wages and salaries, (2) Non-farm

[2] Two cases were excluded who reported 13 for Adjusted Gross Income for their yearly income. These cases had large values of Cook's D in a regression model for AGI.

self-employed (SE) income, (3) Farm self-employed (SE) income, (4) Social security and railroad retirement benefits, (5) Property income, (6) Public assistance income, (7) Other government transfer income, and (8) Other income. The income categories that are used for the modified and the CPS-constructed total family income are same.

IRS AGI includes (1) Wages, salaries, tips, and other employee compensation, (2) Dividends, (3) Interest income, (4) Business income or (loss), (5) Net gain or (loss) from sale or exchange of capital assets, (6) Net gain or (loss) from supplemental schedule of gains and losses, (7) Pensions, annuities, rents, royalties, partnerships, estates or trusts, etc., (8) Farm income or (loss), (9) Fully taxable pensions and annuities, (10) 50% of capital gain distributions, (11) State Income tax refunds standard deduction, (12) Alimony received, and (13) Other income.

Because the categories used for CPS income components are not exactly the same as those used by the IRS, it is not possible to construct a value from CPS that is conceptually identical to IRS AGI. For this analysis, the amount of welfare payments received—an example of public assistance payments— is not excluded from the modified total family income as it is considered to be a part of the CPS income construct even though it is not a part of the IRS AGI (Form 1040). A variable for whether the family received welfare payments is included in the models used to adjust the comparison of CPS derived from family income and IRS AGI amounts. The final data file includes 15,999 records.

Figure 4.1 shows the scatterplots of the CPS constructed total family income, modified total family income, and AGI from IRS records along with the product moment correlation coefficients. Although the correlation between the constructed total family income and modified total family income is strong, 0.96, the residual discrepancies between these two measures could not be resolved. One computation difference between the two measures is that to be comparable to AGI modified CPS total family income excludes the income earned by the children (family members 15 and younger). But differences between these two measures with regard to presence of children were not systematic enough to explain the discrepancies. The mode choice and regression analyses of mode response differences use modified CPS total family income which had a higher correlation with IRS AGI (Form 1040).

Figure 4.2 also shows the relationship between the modified CPS income computed for this analysis and IRS AGI. There is a strong relationship between the two since $\hat{\rho}$ =0.86. However, as the scatterplot shows, there are many discrepancies between the two measures of income for individual families. Assuming that IRS AGI is the truth, the bulk of the differences must be due to reporting errors by CPS respondents on either total income or the components of income. Whether this is willful misreporting, recall error, failure to report all types of income in the CPS, or some other reason cannot be known based on the 1973 data.

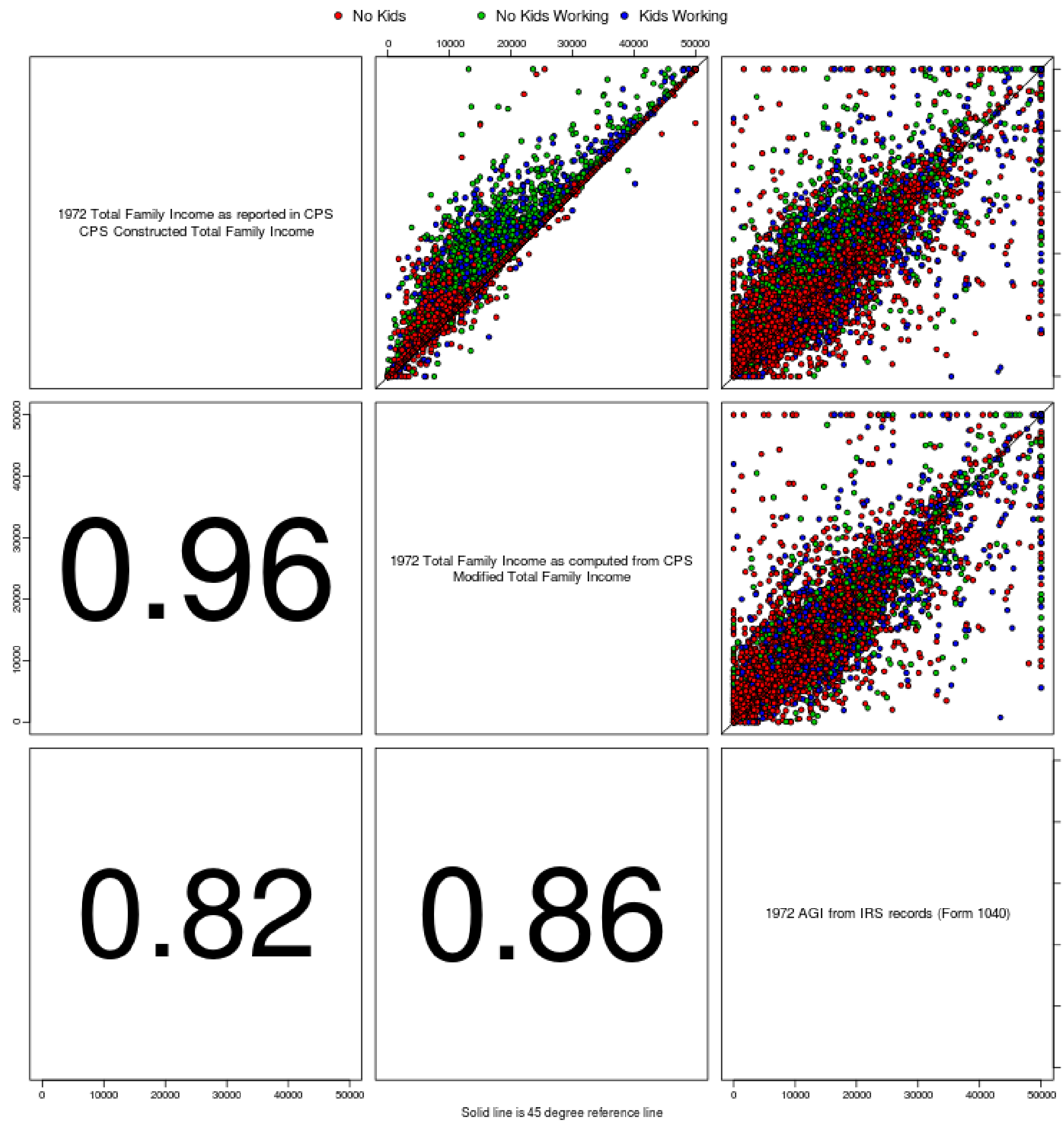


Figure 4.1 – Family Income Construct Correlations. CPS Imputations for missing items are excluded. Incomes are top-coded at 50,000 USD. The CPS constructed total family income measure includes income from all family members age 15 and over.

Figure 4.2 shows the Figure 4.1 data as color coded by response mode. The scatterplot shows the modified total family income versus AGI and each data point represents a family. The red and the blue dots represent data points by telephone and in-person modes, respectively. Ideally, all the data points should be clustered around the solid 45 degree line.

Figure 4.3 shows the distribution of differences, $(y_j^{CPS} - y_j^{AGI})$, of telephone and in-person modified total family income responses. There were essentially no differences in the distribution of differences of telephone and in-person total family income. Differences by mode were further analyzed in regression analysis to address whether mode effects were ignorable in the 1973 CPS match data.

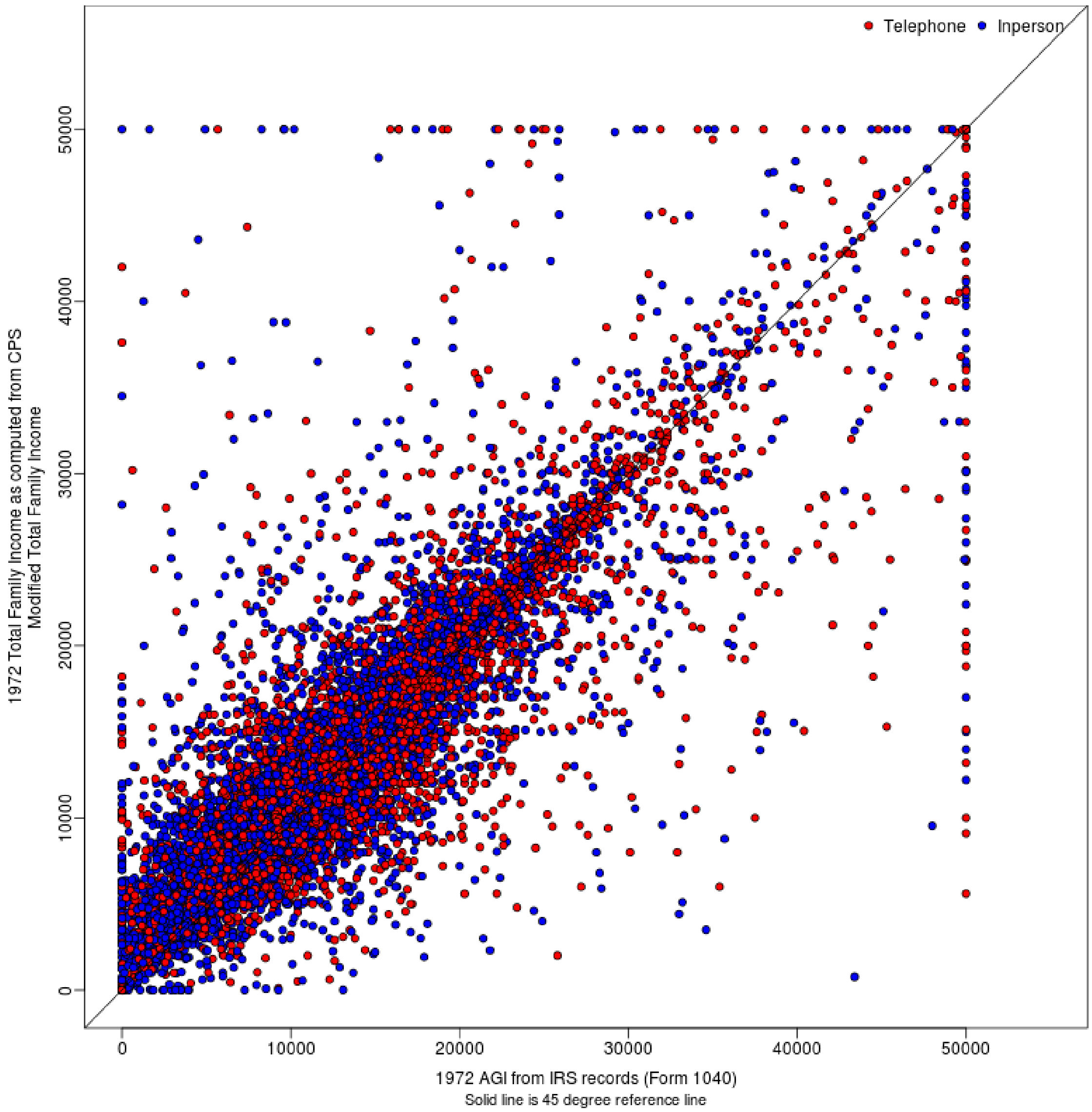


Figure 4.2 – Scatterplot of Modified Total Family Income versus IRS AGI (Form 1040) in 1973 CPS Match Data

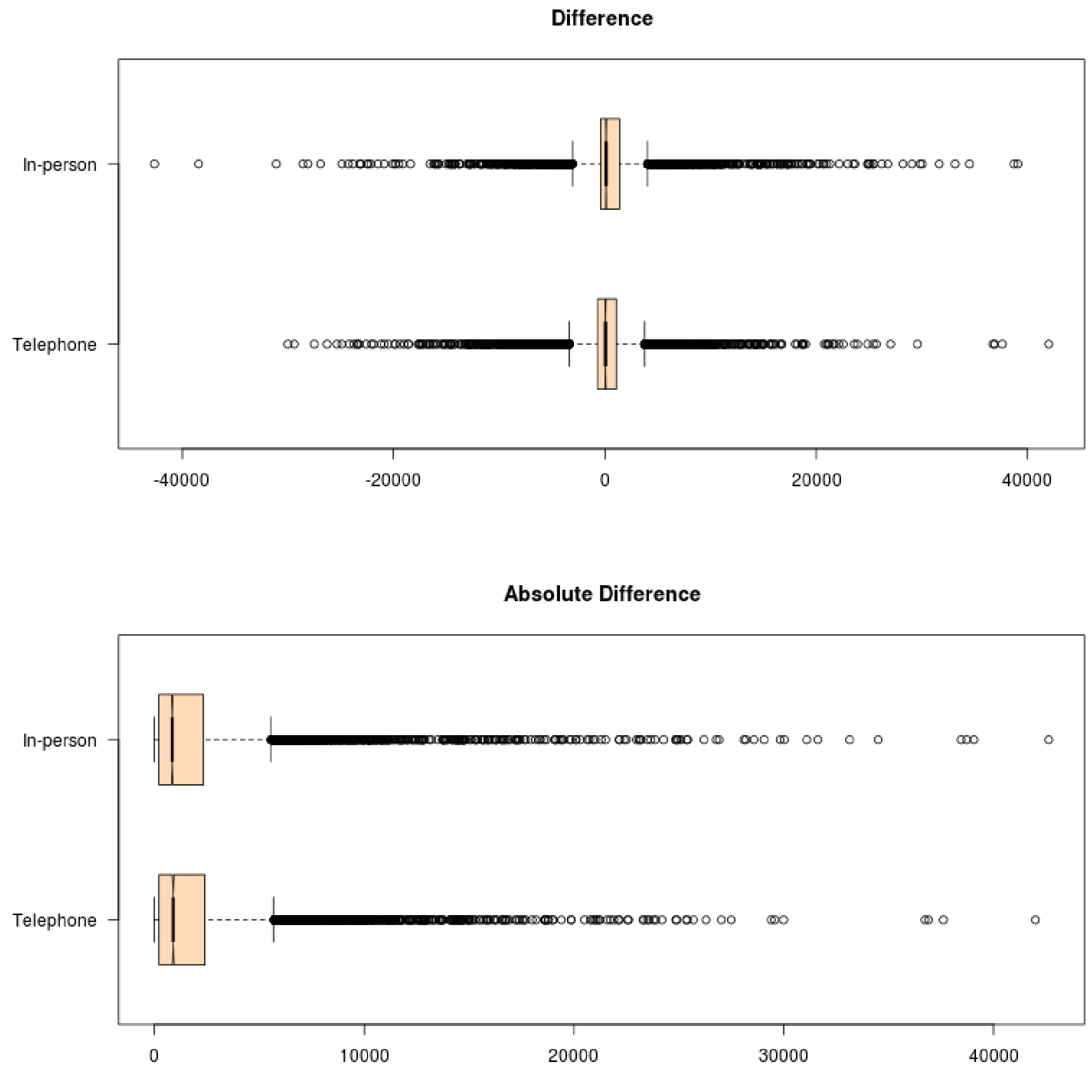


Figure 4.3 – Boxplots of Differences in Modified Total Family Income and IRS AGI by Response Mode in 1973 CPS Match Data

In addition to householder characteristics (Table 4.1), household characteristics such as presence of children, Standard Metropolitan Statistical Area (SMSA) residence, occupied unit tenure, living quarters, region, and welfare receipt status were also controlled in the regression models (Table 4.2). Section 4.2.3 summarizes the differences between the householder and household characteristics in 1973 CPS Match Data and 2012 CPS March data as reported in Table 4.1 and 4.2.

Table 4.1 – Householder Covariate Percentages in the Special 1973 CPS Match File and 2012 CPS Data (*1973 and 2012 categories are not comparable.)

Variable	**Category**	**1973 Match Data %**	**2012 Weighted %**
Marital Status	Married	98.29	49.51
	Single	1.71	50.49
Sex	Male	99.07	50.67
	Female	0.93	49.33
Age	15-24	7.59	5.26
	25-29	12.42	7.77
	30-34	12.33	8.99
	35-39	10.38	8.7
	40-44	11.82	9.4
	45-49	11.76	9.8
	50-54	11.64	10.1
	55-59	9.61	9.8
	60-64	7.30	8.7
	65-69	3.26	6.7
	70-74	1.13	4.9
	75+	0.75	10.0
Education Attainment	None	0.28	9.5
	Elementary School	16.81	12.0
	High School	54.85	28.5
	College	28.06	50.0
Race-Ethnicity-White	0	5.92	31.8
	1	94.08	68.2
CPS Income	None	0.15	-
	Wages only	35.56	61.3
	Self-employment only	3.10	4.7
	Other only	0.31	-
	Wages and Self-employment	3.69	-
	Other and Self-employment	5.11	-
	Other and Wages	45.68	-
	Wages, Self-employment and Other	6.39	-
	Nonworker	-	34.0
Part-time/Full-time Status in 1972	Full-time	95.68	54.9
	Part-time	4.32	11.1
	Nonworker	-	34.0
Work Class 1973*	Other	9.29	-
	Professional	27.36	-
	Sales	12.64	-
	Craft	45.84	-
	Laborer	4.88	-
Work Class 2012	Management	-	7.75
	Business and financial operations	-	3.34
	Computer and mathematical sciences	-	1.93
	Architecture and engineering	-	1.50
	Life, physical, and social sciences	-	0.66

	Community and social service	-	1.15
	Legal	-	0.85
	Education, training, and library	-	3.93
	Arts, design, entertainment, sports	-	1.29
	Healthcare practitioner and technician	-	3.64
	Healthcare support	-	1.70
	Protective service	-	1.62
	Food preparation and serving related	-	2.95
	Building and grounds cleaning and maintenance	-	2.50
	Personal care and service	-	2.16
	Sales and related	-	6.72
	Office and administrative support	-	7.89
	Farming, fishing, and forestry	-	0.44
	Construction and extraction	-	3.44
	Installation, maintenance, and repair	-	2.17
	Production	-	3.99
	Transportation and material moving	-	4.04
	Armed Forces	-	0.34
	Nonworker	-	34.00
Industry 1973*	Other	8.56	-
	Agriculture	6.04	-
	Construction	12.40	-
	Manufacturing	35.64	-
	Transportation	9.21	-
	Trade	19.44	-
	Service	8.71	-
Industry 2012	Agriculture, forestry, fishing and hunting	-	1.0
	Mining	-	0.5
	Construction	-	4.3
	Manufacturing	-	7.1
	Wholesale and retail trade	-	8.3
	Transportation and utilities	-	3.4
	Information	-	1.5
	Financial activities	-	4.7
	Professional and business	-	8.1
	Educational and health services	-	14.9
	Leisure and hospitality	-	5.0
	Other services	-	3.2
	Public administration	-	3.7
	Armed Forces	-	0.3
	Nonworker	-	34.0
Employment Status of Spouse	Not working	45.79	15.6
	Full-time	35.92	26.5
	Part-time	16.57	5.7
	Single	1.71	52.1

Table 4.2 – Household Covariate Frequencies and Percentages in the 1973 CPS Match Data

Variable	Category	1973 Match Data %	2012 Weighted %
Kids in the HH			
	No Kids older than 14	63.72	-
	Kids older than 14- no income	16.50	-
	Kids older than 14- with income	19.78	-
Presence of children			
	No kids under 14	-	71.74
	Kids under 14	-	28.26
SMSA Residence			
	Not in SMSA	33.21	-
	in SMSA: Central City	25.37	-
	in SMSA:Ring	41.42	-
Principal city/Balance status			
	Principal city	-	28.34
	Balance of CBSA	-	41.69
	Non CBSA	-	15.42
	Not identified	-	14.55
Occupied Unit Tenure			
	Unknown	1.98	
	Owned or being bought	74.67	64.41
	Rented for cash	21.57	34.15
	Occupied without payment or cash rent	1.78	1.45
Living Quarters			
	Other	96.53	95.45
	Trailer-Permanent	3.47	4.55
Census Region and Division of Residence			
	Northeast	23.21	17.65
	North Central (Midwest)	30.73	22.55
	South	29.26	38.24
	West	16.80	21.56
Welfare Receipt Status			
	0	98.77	98.89
	1	1.23	1.11

Further information related to the dataset can be found in previous studies that used subsets of the 1973 public-use Current Population Survey (CPS) and Social Security Records Exact Match data to evaluate the properties the imputation methods for income item nonresponse in the CPS (David, Little, Samuhel, & Triest, 1986; Glynn et al., 1993; Greenlees et al., 1982).

4.2.2. 2012 CPS March Data

CPS March 2012 respondent data are used to perform empirical comparison analyses of the proposed inference methods in a condition where no benchmark values are available for the variables of study: (1) personal income as reported in CPS March Supplement, and (2) health insurance coverage. The details of the empirical comparison analyses are discussed in Section 6.3. CPS March Supplement measures of income and health insurance coverage are merged with the CPS March data to determine the response mode. The nonrespondents to the CPS March 2012 are excluded from the analysis. Future research will include the evaluations of the nonresponse adjustments for the proposed inference methods.

The variables of study are not modified as in the 1973 CPS match data application as the comparison of the measures to a benchmark is not relevant. Instead March 2012 CPS reported measure is used. The imputed values for personal income and health insurance coverage are excluded from the analyses. The unit of analysis is householders as in the 1973 CPS match data application.

4.2.3. Descriptive Statistics and Comparisons of 1973 and 2012

Table 4.1 -Table 4.2 and Figure 4.4 show how the distributions of householder and household characteristics differ between 1973 and 2012 data. The householder and the household characteristics were skewed in terms of marital status, sex, and race in 1973 compared to the March 2012 CPS data. These shifts in the distributions were both a result of the selection criteria for the analysis subsets of 1973 and 2012 CPS data and the changing social-demographics of the U.S. population. In the comparisons, weighted percentages of 2012 data were used to reflect the population distributions. Since it is a specific subset of 1973 CPS data, 1973 match percentages are unweighted. The shift in the 2012 householder data was towards being older, a college graduate, and a single. The householders were equally males and

females in 2012. The percentage of non-working spouses drops in 2012. In 2012, about one third of the respondents were non-whites compared to 6% in 1973 CPS match data.

One of the characteristics whose distribution was much different for 1973 and 2012 was industry and occupation of longest job in the last year. While about 46% of the householders worked in craft occupations in 1973, about one fourth worked in professional occupations. Thirty five percent of householders worked in manufacturing. Trade was the second most common industry that householders worked in. While the categories were not exactly same, the difference in the agriculture percentage between 1973 and 2012 years is worth noting. 1973 match data indicated that 6% of householders worked in agriculture. 2012 CPS data shows that only 1% of householders worked in this industry.

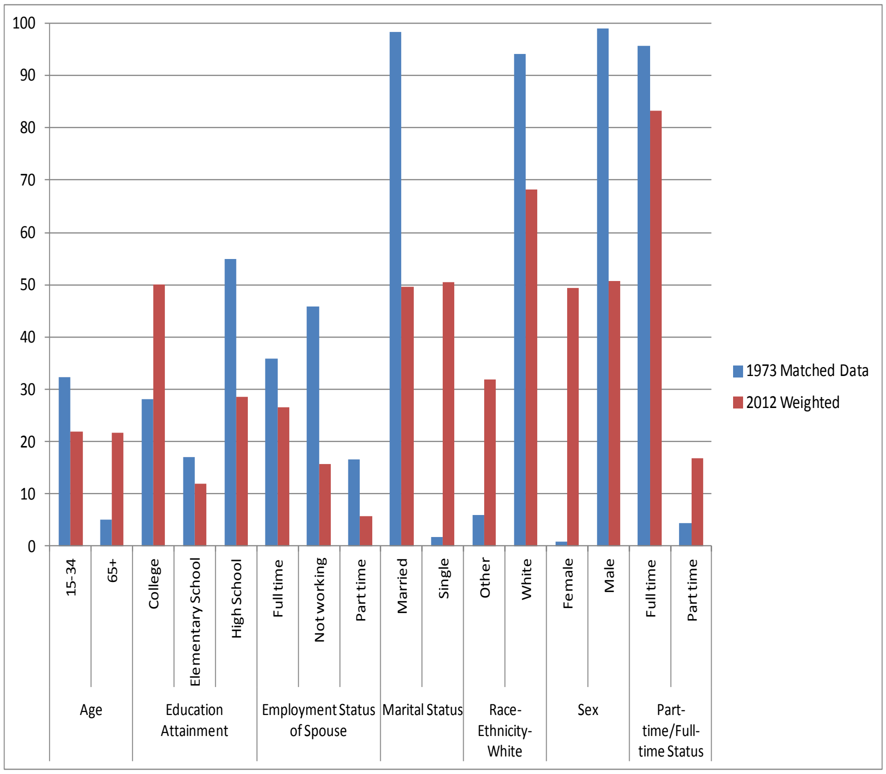


Figure 4.4 – Distribution Differences in Householder Characteristics between 1973 CPS Match Data and 2012 Weighted CPS Data

These differences are important for three reasons. First, they suggest possible changes in the survey population (although 1973 CPS match data is only a subset of the 1973 CPS data). The differences are confounded by the selection criteria and does not purely reflect the differences between 1973 and 2012 survey populations. Second, characteristics such as age, education, and race have been studied and hypothesized to be indicators of cognitive abilities, and social conditions that may yield nonignorable mode effects (Aquilino, 1994; Holbrook et al., 2003; Schwarz et al., 1991). Due to skewed distributions or small cell sizes, the power of the regression analyses of differences was limited in detecting effects in the 1973 match data. Lastly, more complicated family and income structures may make the measurement of family income difficult and cause an increase in the survey measurement errors of income in the CPS 2012 data (Körmendi, 1988; Moore et al., 2000).

Up to this point, householder and household characteristics have been discussed to illustrate the 1973 CPS match data characteristics. These householder and household characteristics were used as covariates in the regression and imputation models as will be further discussed. In addition to householder and household characteristics, survey response and matching characteristics also influence the survey measurements of income in the CPS 1973 data. As a household survey, CPS allows proxy reporting. Results of studies of the accuracy of proxy reporting in surveys are mixed (Tourangeau et al., 2000b). For example, a proxy respondent may decrease the accuracy of survey reports if he or she attempts to recall factual responses from memory without relying on the actual records or a proxy respondent may be more motivated to use records to report income since he or she is more aware of the lack of knowledge. In parallel, only 28.6% of householders were indicated as the survey respondents in 1973 CPS match data. Thus, in addition to householder and household characteristics, regression and imputation models include the householder March respondent indicator as a control variable for proxy reporting.

4.3. Regression Models for Mode Choice and Measurement Discrepancy in the 1973 CPS Match Data

4.3.1. Model of Mode Choice

Regression models were used to explore the mode choice mechanism and the total family income response characteristics by mode. For the mode choice mechanism, a logistic regression model for the probability of responding using the in-person mode was fit. Table 4.3 shows the proportions of response modes by month in sample.

Table 4.4 summarizes Significant Type III tests of main effects ($p<0.01$) in the logistic regression of responding by in-person mode. In terms of the mode choice, some covariates were found to be related to responding via in-person versus telephone modes. As one of the design factors, month in sample (MIS) was considered to be the main factor underlying the mode choice. In parallel to the logistic regression model analyses, we also conducted regression tree analyses which determined that some MIS interactions, including interactions with region, education, industry, children's and spouse's working status, SMSA residence and tenure of occupied residence, were significant predictors of mode choices. Among these interactions only MIS x Region and MIS x SMSA residence interactions were detected as significant by Type III tests. To increase the prediction accuracy all the MIS interactions were included in the final model choice imputation model (Table 4.5). Also, March Respondent and White variables were dropped from the final mode choice imputation model to avoid small cell sizes.

Table 4.3 – Response Mode Distribution by Month in Sample

Response Mode	Month in Sample							
	W1	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**
Telephone	2%	34%	60%	66%	6%	53%	63%	65%
In-person	98%	66%	40%	34%	94%	47%	37%	35%

Table 4.4 – Mode Choice Logistic Model Type III Tests, 1973 CPS Match Data

Factors	LR Chisq	Df	Pr(> Chisq)
Month in sample (MIS)	4915.44	7	0.00
Education in years	68.37	1	0.00
Industry householder worked in	30.16	5	0.00
Children's and spouse's working status	25.15	6	0.00
SMSA residence	723.34	2	0.00
Tenure of occupied residence	18.25	1	0.00
Job class	15.62	4	0.00
March respondent	13.87	1	0.00
White	15.53	1	0.00

Table 4.5 – Mode Choice Logistic Model including Interactions Type III Tests, 1973 CPS Match Data

Factors	**LR Chisq**	**Df**	**Pr(> Chisq)**
Month in sample (MIS)	44.09	7	0.00
Education in years	2.58	1	0.11
Industry householder worked in	2.11	5	0.83
Children's and spouse's working status	9.69	6	0.14
SMSA residence	4.15	2	0.13
Tenure of occupied residence	0.03	1	0.87
Region	13.52	3	0.00
March respondent	11.77	1	0.00
White	12.84	1	0.00
MIS x Region	57.41	21	0.00
MIS x Education	5.66	7	0.58
MIS x Industry	39.03	35	0.29
MIS x Children's and spouse's working status	46.79	42	0.28
MIS x SMSA residence	36.33	14	0.00
MIS x Tenure	9.73	7	0.20

4.3.2. Model of Response Differences

In exploring the total family income response characteristics by mode, the absolute difference of income, $(y_j^{CPS} - y_j^{AGI})$ was used where y_j^{CPS} was the modified total family income for family j . Since IRS AGI data were available as the standard to compare survey data against, both telephone and in-person differences could be evaluated in regression models in which other factors were held constant.

Including or excluding the householders whose records show zero AGI (n=80) and who reported Income Type as none (n=15) did not change the results of the absolute difference regression analyses. The final model structure of differences was determined based on the ANOVA Type III tests. Alternative model structures were tested starting with the main effects only for household and householder level covariates. Mode interactions for significant predictors were tested in the following model.

Table 4.6 shows the Type III test statistics for each of the covariates included in the final model. For the purpose of this regression analysis, mode effects were considered to be the beta coefficients for the in-person and the in-person interactions. Although the overall mode effect was not significant ($p=0.60$), mode effects at some subgroup levels were significant. The interactions of welfare receipt status (Figure 4.5), age (Figure 4.6) and income type (Figure 4.7) with the response mode were statistically significant ($p<0.05$).

Table 4.6 – Type III tests for the Factors in the Linear Regression Model of Absolute Difference, 1973 CPS Match Data

Covariates	**Sum Sq**	**Df**	**F value**	**Pr(> F)**
(Intercept)	4431987.82	1	0.49	0.48
Occupied Unit Tenure	88751646.86	3	3.29	0.02
Month in sample (MIS)	136198143.2	7	2.16	0.03
In-person	2507067.67	1	0.28	0.60
Welfare payment recipient	114821382.8	1	12.75	0.00
Householder March respondent	170429.97	1	0.02	0.89
Education attainment	743157133.3	1	82.55	0.00
Age (categorical)	983615851.6	11	9.93	0.00
Race-ethnicity(White)	63097456.24	1	7.01	0.01
Part-time/Full-time status	501614545.6	4	13.93	0.00
Industry for the longest job in 1972	709559144.6	5	15.76	0.00
Income type	2412227252	7	38.28	0.00
Children’s and spouse’s working status	478212268.8	6	8.85	0.00
In-person mode x Welfare payment recipient	40415660.11	1	4.49	0.03
In-person mode x Age (categorical)	192508061.1	11	1.94	0.03
In-person mode x Income type	179072791	7	2.84	0.01

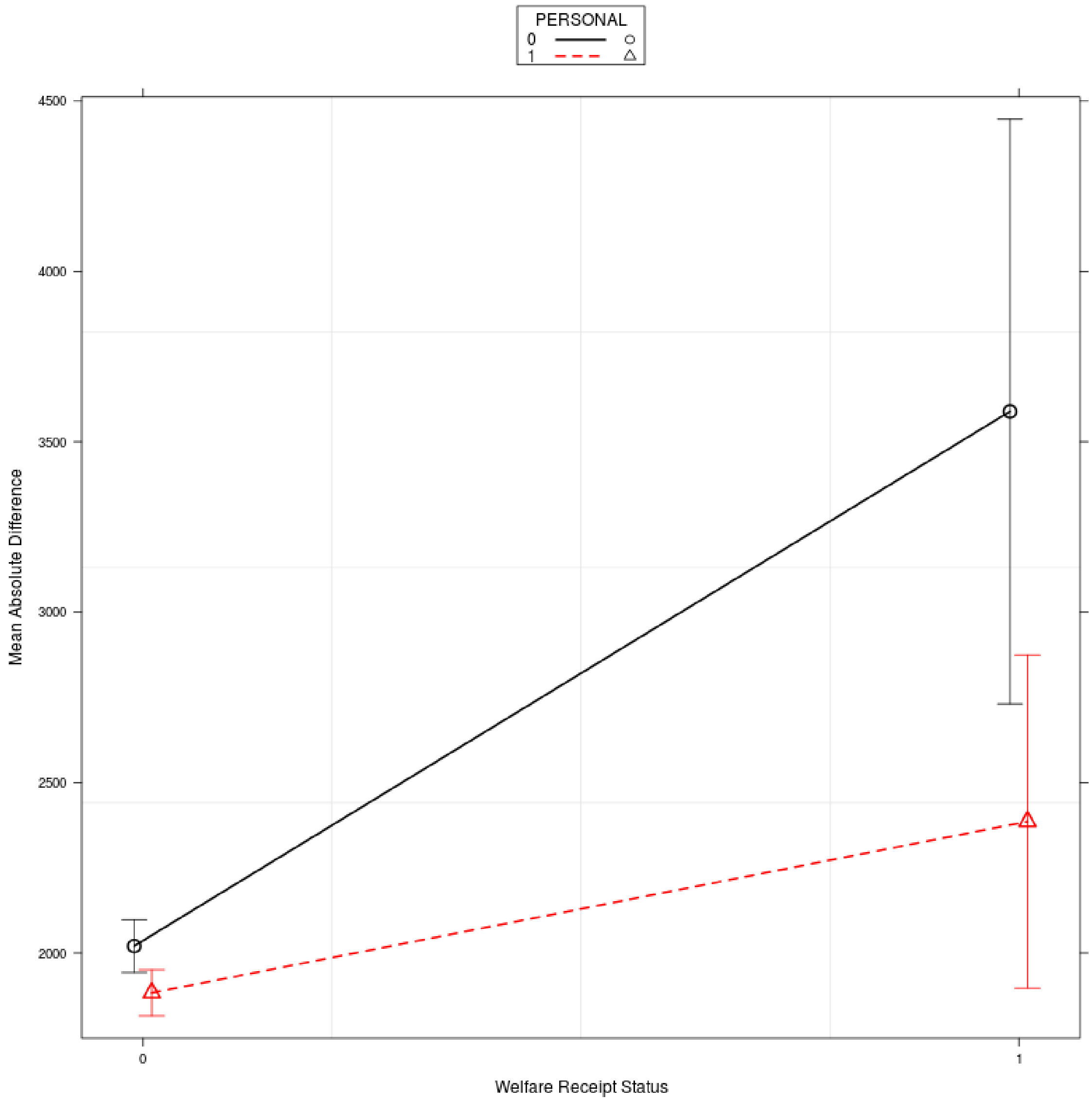


Figure 4.5 – Marginal Mean Absolute Difference for Welfare Receipt Status Controlling for all the Covariates in the Final Model (Table 4.6) by Response Mode (Personal=1: In-person, Personal=0: Telephone). The corresponding bars represent 95% confidence intervals.

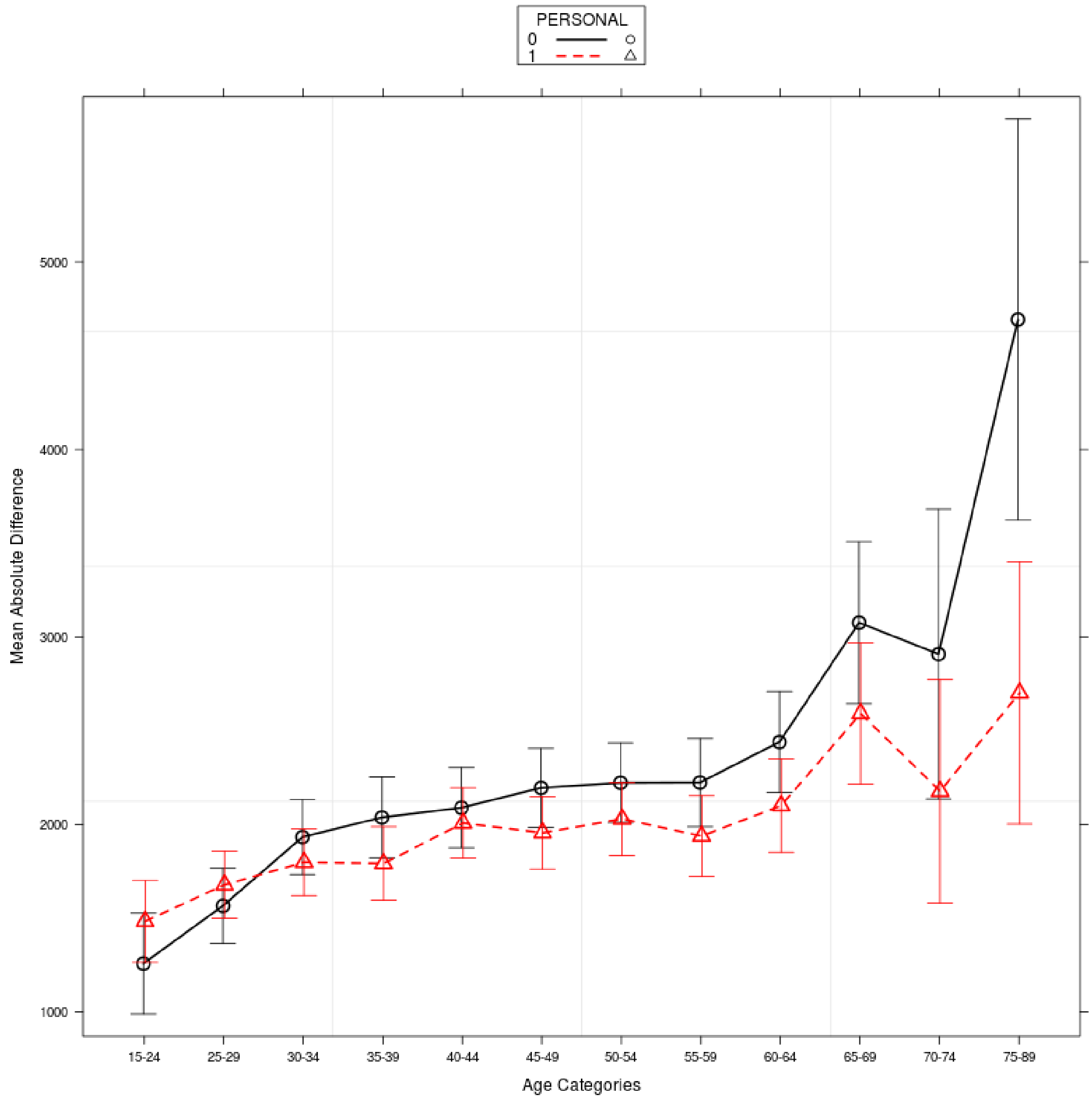


Figure 4.6 – Marginal Mean Absolute Difference for Age Controlling for all the Covariates in the Final Model (Table 4.6) by Response Mode (Personal=1: In-person, Personal=0: Telephone). The corresponding bars represent 95% confidence intervals.

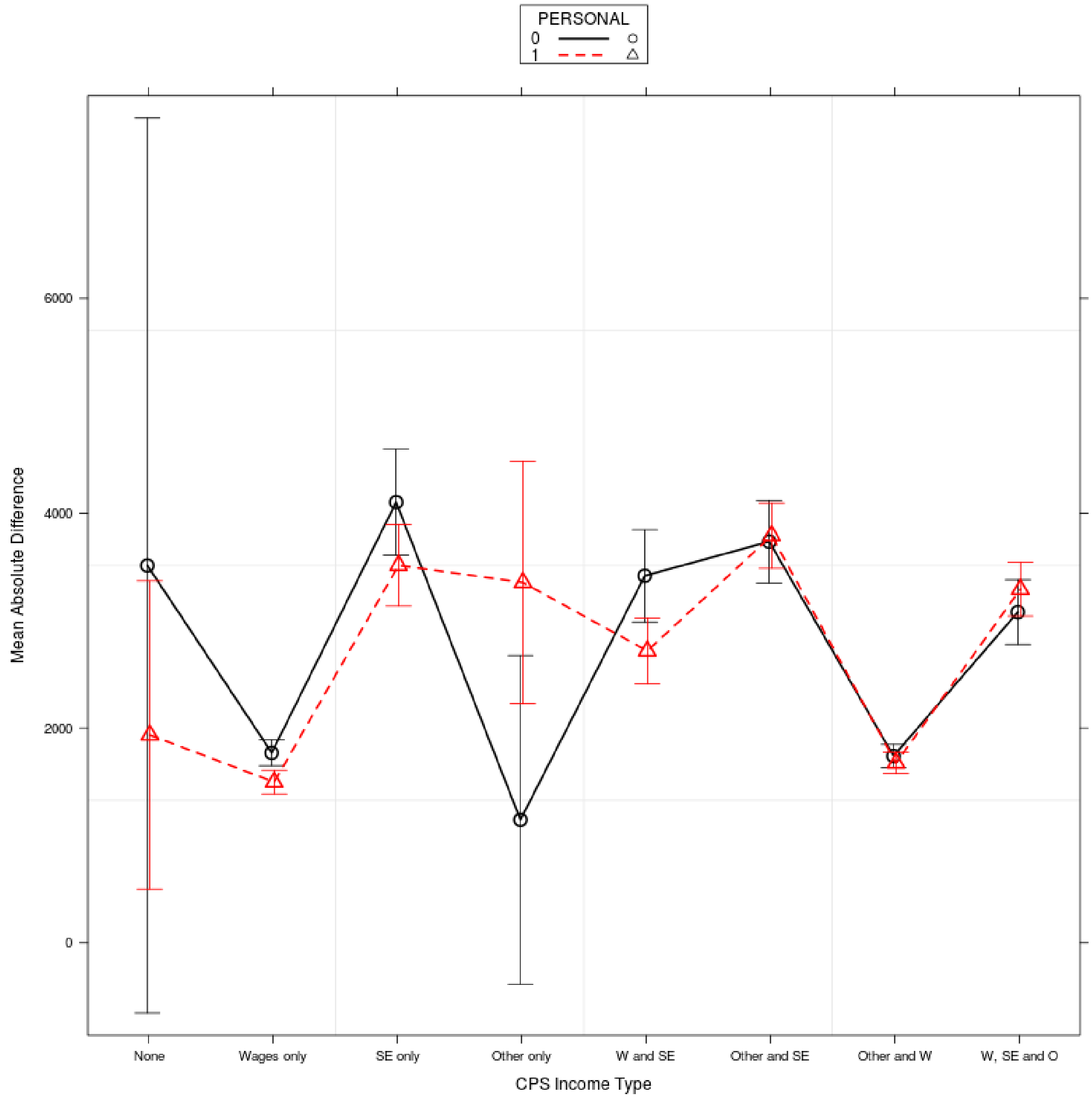


Figure 4.7 – Marginal Mean Absolute Difference for Income Type Controlling for all the Covariates in the Final Model (Table 4.6) by Response Mode (Personal=1: In-person, Personal=0: Telephone). The corresponding bars represent 95% confidence intervals.

Marginal mean absolute differences were computed by using the effects package in R (Fox, 2003) for the covariates with significant mode interactions (Figures 4.5-4.7). Marginal mean absolute differences are estimated marginal means for each cell given the best linear unbiased estimator (BLUE) of the corresponding cell means (Searle, Speed, & Milliken, 1980). Wide confidence intervals in Figures 4.5-4.7 indicate small cell sizes. The marginal mean absolute difference was higher for householders who have not received a welfare receipt in in-person mode (Figure 4.5). Although the differences in marginal absolute difference were significantly higher for the 75-80 age group in telephone mode, there was an increasing pattern in mode effects starting in 60-64 age group (Figure 4.6). Although it is expected to see larger mode effects for the more complicated income structures, only in the wage only group, telephone mode yielded the larger difference (Figure 4.7).

One of the possible reasons for no differences at the aggregate may be the skewed distributions for the subgroups including race, gender, spouse's employment status, and education attainment. For example, blacks were suspected to be more prone to mode effects, in particular due to social desirability bias compared to the whites (Aquilino, 1994) and the vast majority of 1973 data were collected from whites (96%). The distributions of these characteristics in the 1973 data were quite different from those in 2012. Therefore, the conclusion of ignorable overall mode effects may not apply to 2012 data.

4.3.3. Model for Total Family Income based on 1973 CPS Match Data

The regression models were fit separately for the subsets of the CPS 1973 telephone and in-person respondents. For the modified total income, the regression models included the covariates from the GRZ model and covariates that are considered to be related to measurement error. The visual inspection of the residual density plots did not reveal severe departure from normality assumption. Thus, regression models were fit on the original income scale without a log transformation. For both telephone and in-person regression models, all the factors were statistically significant ($p<0.05$), except gender and month in sample in the telephone respondent regression (not reported in detail in this dissertation). The R squares are 0.36 and 0.38 for telephone and in-person regression models, respectively.

4.4. Regression Models for CPS March 2012 Data

4.4.1. Models for Mode Choice based on CPS 2012 data

The exploratory analysis of predicted probabilities for health insurance coverage showed that it is more informative to stratify the sample in four groups by age and work status of householders: 65+ vs. <65, and worker vs. nonworker (Table 4.7). The mode choice logistic regression models were fit separately for these four groups.

Table 4.8 shows the ANOVA Type III tests in the final model in which all the predictors were significant. A stepwise approach was followed to determine the final model structure. Initially significance for householder and household level beta coefficients were tested by ANOVA Type III tests, and month in sample interactions of these significant predictors were tested in a following model. The regression models suggested that the state and month in sample interactions produced small cell sizes and no variation in response variable Health Insurance Coverage. Therefore, for the final Health Insurance Coverage imputation models month in sample interactions were dropped.

Table 4.7 – Response Mode % by Age x Work Status

Age x Work status	n	In-person %	Telephone %
65+, Worker	2,040	35.74	64.26
65+, Nonworker	7,531	42.70	57.30
<65, Worker	25,991	40.44	59.56
<65, Nonworker	6,761	48.41	51.59

Table 4.8 – Mode Choice Logistic Model Type III Tests, CPS March 2012 Respondents

Covariates	Df	65+, Worker (n=2,040)		65+, Nonworker (n=7,531)		<65, Worker (n=25,991)		<65, Nonworker (n=6,761)	
		LR Chisq	Pr(> Chisq)	LR Chisq	Pr(> Chisq)	LR Chisq	Pr(> Chisq)	LR Chisq	Pr(> Chisq)
Month in sample (MIS)	7	473.67	0.00	1750.46	0.00	4935.96	0.00	1069.24	0.00
State	50	74.55	0.01	205.30	0.00	285.53	0.00	126.16	0.00
Living quarters	1	-	-	11.96	0.00	-	-		
Tenure	2	6.26	0.04	30.12	0.00	77.24	0.00	33.93	0.00
Telephone in household	1	-	-	-	-	43.52	0.00	29.74	0.00
Telephone available (Universe=No telephone in household)	2	-	-	29.24	0.00	-	-	-	-
Telephone interview acceptable (Universe=Telephone available)	1	112.58	0.00	455.84	0.00	904.01	0.00	391.58	0.00
Principal city/Balance status	3	-	-	-	-	10.72	0.01	-	-
Metropolitan area (CBSA) size	6	-	-	-	-	18.42	0.01	-	-
Sex	1	-	-	12.12	0.00	-	-	7.44	0.01
Age (Categorical)	8	-	-	-	-	34.16	0.00	15.68	0.05
Level of school completed/degree received	3	-	-	18.32	0.00	53.24	0.00	19.97	0.00
Race-ethnicity	3	-	-	27.79	0.00	98.06	0.00	12.85	0.01
Occupation of longest job	4	-	-	-	-	22.75	0.00	-	-
Employment status	3	-	-	-	-	9.58	0.02	-	-
Spouse's employment status and presence of children	9	-	-	-	-	43.98	0.00	34.36	0.00
Householder March respondent	1	4.90	0.03	2.70	0.10	126.15	0.00	0.83	0.36
MIS x State	350	-	-	413.55	0.01	405.33	0.02	-	-

MIS x Telephone in household	7	-	-	-	-	17.42	0.01	-	-
MIS x Telephone interview acceptable	7	22.53	0.00	15.71	0.03	30.92	0.00	27.13	0.00
MIS x Metropolitan (CBSA) size	42	-	-	-	-	69.46	0.00	-	-
MIS x Householder March Respondent	7	-	-	-	-	-	-	20.03	0.01

4.4.2. Models for Personal Income based on CPS 2012 data

The stepwise regression method described in Section 4.4.1 was used to select the covariates for the total personal income regression models. All the factors shown in Table 4.8 are included in the R-model which was a part of the nonignorable mode choice model imputations. Table 4.9 presents the covariates included in the final Y-model. Table 4.1 and 4.2 present the distributions for these variables. For some of the covariates combined variables were computed, for example spouse's employment status and presence of children were combined. The log transformation was used in the imputation models. In the prediction computations the bias correction for the log transformation was applied (Newman, 1993). The models included the state and month in sample covariates to incorporate the sampling design in the regression models. The sampling weights were recomputed at the state and MIS level to reflect the unequal probabilities of selection. As an illustration of the fitted models, beta coefficient estimates and their standard errors for the response imputation models for log personal income are presented in Table 4.10. For brevity, beta coefficients for the other models have not been included in this thesis.

Table 4.9 – Covariates included in the Response Imputation Models (Y-Model)

Covariate
Month in sample (MIS)
State
Living quarters
Tenure
Principal city/Balance status
Metropolitan area (CBSA) size
Householder March respondent
Spanish speaking households
Sex
Age (Categorical)
Level of school completed/degree received
Race-ethnicity
Employment status
Occupation of longest job
Industry of longest job
Part-time/Full-time Status
Sources of earnings
Spouse's employment status and presence of children (Family Type)

Table 4.10 – Beta Coefficient Estimates and Standard Errors for the Response Imputation Model for the Natural Logarithm of Personal Income

	In-person	**Telephone**
Parameters	**Estimate (SE)**	**Estimate (SE)**
Intercept	9.87 (0.04)	9.86 (0.05)
MIS=2 vs. MIS=1	-0.02 (0.01)	-0.01 (0.02)
MIS=3	-0.03 (0.01)	-0.04 (0.02)
MIS=4	-0.02 (0.01)	-0.02 (0.02)
MIS=5	-0.02 (0.01)	-0.01 (0.02)
MIS=6	-0.02 (0.01)	-0.01 (0.02)
MIS=7	-0.04 (0.01)	-0.02 (0.02)
MIS=8	-0.03 (0.01)	-0.01 (0.02)
State=AK vs. AL	0.01 (0.05)	0.13 (0.05)
State=AZ	-0.05 (0.05)	-0.10 (0.05)
State=AR	-0.12 (0.04)	-0.04 (0.05)
State=CA	0.01 (0.03)	0.10 (0.04)
State=CO	-0.06 (0.04)	0.02 (0.04)
State=CT	0.08 (0.04)	0.15 (0.04)
State=DE	-0.04 (0.04)	-0.02 (0.05)
State=DC	0.04 (0.04)	0.10(0.05)
State=FL	-0.06 (0.03)	-0.03 (0.04)
State=GA	-0.12 (0.04)	0.02 (0.04)
State=HI	0.04 (0.05)	0.13 (0.05)

State=ID	-0.1 (0.04)	-0.04 (0.05)
State=IL	-0.03 (0.04)	-0.02 (0.04)
State=IN	-0.04 (0.04)	-0.06 (0.04)
State=IA	-0.07 (0.04)	-0.02 (0.04)
State=KS	-0.05 (0.04)	-0.04 (0.04)
State=KY	-0.08 (0.04)	0.01 (0.04)
State=LA	-0.07 (0.04)	-0.09 (0.05)
State=ME	-0.05 (0.04)	-0.04 (0.04)
State=MD	0.00 (0.04)	0.06 (0.04)
State=MA	-0.03 (0.05)	0.01 (0.04)
State=MI	-0.08 (0.04)	-0.03 (0.04)
State=MN	-0.05 (0.04)	0.01 (0.04)
State=MS	-0.06 (0.04)	-0.05 (0.05)
State=MO	-0.06 (0.04)	-0.04 (0.04)
State=MT	-0.12 (0.05)	-0.05 (0.04)
State=NE	-0.04 (0.05)	-0.01 (0.04)
State=NV	-0.08 (0.04)	0.01 (0.04)
State=NH	-0.01 (0.04)	0.02 (0.04)
State=NJ	0.04 (0.04)	0.12 (0.04)
State=NM	-0.02 (0.05)	0.07 (0.05)
State=NY	-0.02 (0.04)	0.06 (0.04)
State=NC	-0.06 (0.04)	-0.02 (0.04)
State=ND	-0.02 (0.05)	0.01 (0.04)
State=OH	-0.03 (0.04)	-0.01 (0.04)
State=OK	-0.05 (0.05)	-0.08 (0.04)
State=OR	-0.04 (0.05)	0.01 (0.04)
State=PA	-0.04 (0.04)	0.00 (0.04)
State=RI	0.02 (0.04)	0.08 (0.04)
State=SC	-0.10 (0.04)	0.00 (0.05)
State=SD	-0.11 (0.04)	-0.06 (0.04)
State=TN	-0.12 (0.04)	-0.08 (0.04)
State=TX	-0.04 (0.03)	-0.01 (0.04)
State=UT	-0.08 (0.04)	-0.05 (0.05)
State=VT	-0.02 (0.04)	-0.03 (0.04)
State=VA	0.01 (0.04)	0.03 (0.04)
State=WA	0.00 (0.04)	0.09 (0.04)
State=WV	-0.02 (0.04)	-0.06 (0.05)
State=WI	-0.04 (0.04)	0.02 (0.04)
State=WY	-0.01 (0.04)	0.07 (0.04)
Living quarter= Other vs. House, apt.,flat	-0.09 (0.02)	-0.07 (0.02)
Tenure=Rent vs. Owned or being bought	-0.12 (0.01)	-0.13 (0.01)
Tenure=No cash rent	-0.20 (0.03)	-0.22 (0.03)

Principal city/Balance status=Balance of CBSA vs. Principal city	0.02 (0.01)	0.00 (0.01)
Principal city/Balance status=Non CBSA	-0.04 (0.02)	-0.04 (0.02)
Principal city/Balance status=Not identified	0.04 (0.02)	-0.01 (0.02)
Metropolitan area (CBSA) size=250,000 - 499,999 vs. 100,000-249,999	0.03 (0.02)	0.01 (0.02)
Metropolitan area (CBSA) size=500,000 - 999,999	0.04 (0.02)	0.01 (0.02)
Metropolitan area (CBSA) size=1,000,000 - 2,499,999	0.06 (0.02)	0.04 (0.02)
Metropolitan area (CBSA) size=2,500,000 - 4,999,999	0.09 (0.02)	0.05 (0.02)
Metropolitan area (CBSA) size=5,000,000+	0.07 (0.02)	0.05 (0.02)
Metropolitan area (CBSA) size=Not identified	0.05 (0.02)	-0.01 (0.02)
Householder March respondent vs. Not	0.05 (0.01)	0.04 (0.01)
Spanish speaking households vs. Other	-0.11 (0.02)	-0.15 (0.03)
Female vs. Male	-0.20 (0.01)	-0.21 (0.01)
Age=25-29 vs. 15-24	0.10 (0.02)	0.09 (0.02)
Age=30-34	0.16 (0.02)	0.15 (0.02)
Age=35-39	0.17 (0.02)	0.19 (0.02)
Age=40-44	0.19 (0.02)	0.23 (0.02)
Age=45-49	0.18 (0.02)	0.21 (0.02)
Age=50-54	0.22 (0.02)	0.24 (0.02)
Age=55-59	0.21 (0.02)	0.24 (0.02)
Age=60-64	0.26 (0.02)	0.31 (0.02)
Age=65-69	0.45 (0.02)	0.48 (0.02)
Age=70-74	0.48 (0.02)	0.52 (0.03)
Age=75+	0.53 (0.02)	0.52 (0.02)
Education=Highschool vs. Less than 12 grade	0.05 (0.01)	0.06 (0.01)
Education=College	0.16 (0.01)	0.17 (0.01)
Education=Graduate	0.35 (0.02)	0.36 (0.02)
Race/Ethnicity=Black only vs. White	-0.07 (0.01)	-0.09 (0.01)
Race/Ethnicity=Other	-0.11 (0.01)	-0.09 (0.01)
Race/Ethnicity=Hispanic	-0.11 (0.01)	-0.09 (0.01)
Worker vs. Nonworker	1.14 (0.03)	1.17 (0.03)
Occupation of longest job =Service occupations vs. Management, professional occupations	-0.29 (0.01)	-0.31 (0.01)
Occupation of longest job =Sales and office occupations	-0.22 (0.01)	-0.25 (0.01)

Occupation of longest job =Natural resources, construction, and maintenance occupations	-0.21 (0.02)	-0.24 (0.02)
Occupation of longest job=Production,transportation and material moving occupations	-0.32 (0.02)	-0.33 (0.02)
Industry of longest job=Construction vs. Agriculture	-0.07 (0.03)	-0.06 (0.03)
Industry of longest job==Manufacturing	-0.01 (0.03)	0.00 (0.03)
Industry of longest job==Wholesale and retail trade	-0.12 (0.03)	-0.11 (0.03)
Industry of longest job==Transportation and utilities	0.04 (0.04)	0.00 (0.03)
Industry of longest job==Information and financial activities	-0.02 (0.03)	0.02 (0.03)
Industry of longest job==Professional and other services	-0.13 (0.03)	-0.13 (0.03)
Industry of longest job==Public administration	0.00 (0.03)	0.02 (0.03)
Working status=Full year-Part time vs. Full year-Full time	-0.34 (0.02)	-0.39 (0.01)
Working status=Part year-Full time	-0.32 (0.01)	-0.31 (0.01)
Working status=Part year-Part time	-0.55 (0.02)	-0.62 (0.02)
Sources of earnings=Self employment vs. other	-0.20 (0.02)	-0.22 (0.02)
Family type=Married-fulltimewrksps-wthkids vs. Married-fulltimewrksps-nokids	-0.04 (0.02)	-0.02 (0.01)
Family type=Married-prttimewrksps-nokids	0.08 (0.02)	0.10 (0.02)
Family type=Married-prttimewrksps-wthkids	0.08 (0.02)	0.13 (0.02)
Family type=Married-ntwrksps-nokids	0.03 (0.02)	0.10 (0.01)
Family type=Married-ntwrksps-wthkids	0.10 (0.02)	0.12 (0.02)
Family type=Ntmarried-nokids	0.07 (0.01)	0.10 (0.01)
Family type=Ntmarried-wthkids	0.09 (0.02)	0.13 (0.02)
Family type=Single-nokids	-0.01 (0.01)	0.03 (0.01)
Family type=Single-wthkids	0.06 (0.02)	0.10 (0.02)

4.4.3. Models for Health Insurance Coverage based on CPS 2012 data

The same stepwise regression method was used to select the covariates for the health insurance coverage logistic regression models. Although the state covariate was intended to be used in the regression models to incorporate the sampling design, this was problematic as there was no variation in the response variable in some cells. Therefore the state covariate was dropped from the imputation models. Table 4.11 presents the covariates in the final Y-models. The covariates presented in Table 4.8, except MIS interactions, were used in R-models of the nonignorable mode choice models. Due to intensive computational requirements, MIS interactions were dropped in R-models.

Table 4.11 – Covariates included in the Response Imputation Models (Y-Model)

Covariate	**65+, Worker (n=2,040)**	**65+, Nonworker (n=7,531)**	**<65, Worker (n=25,991)**	**<65, Nonworker (n=6,761)**
Month in sample (MIS)	x	x	x	x
State				
Living quarters			x	
Tenure	x	x	x	x
Telephone available (Universe=No telephone in household)				x
Telephone interview acceptable (Universe=Telephone available)	x			x
Principal city/Balance status		x	x	
Metropolitan area (CBSA) size			x	
Age (Categorical)				x
Sex			x	x
Level of school completed/degree received			x	x
Race-ethnicity		x	x	x
Spanish speaking households		x	x	
Occupation of longest job			x	
Industry of longest job	x		x	
Part-time/Full-time Status	x		x	
Sources of earnings			x	
Spouse's employment status and presence of children	x	x	x	x
Householder March Respondent	x	x	x	x

Table 4.12 reports unadjusted mode-specific means for the variables of interest in the three studies. The first study uses a subset of 1973 CPS Match data to estimate mean wage and salary income. The difference in adjusted means between telephone and in-person respondents is \$1,369. The second study creates hypothetical populations using a subset of 1973 CPS Match data to investigate the bias properties of total family income. The empirical comparison study investigates the differences in mean estimates of personal income and health insurance coverage using alternative estimation methods of inference. The direction of the differences is all consistently lower for the in-person respondents for any of these variables of interest. As discussed in Chapter 2 and 3, these difference in the distribution of reported values is not due to measurement error only, it is likely due to both mode choice and mode effects. This difference is $(\bar{Y}_T \mid P = T) - (\bar{Y}_I \mid P = I)$ according to the notation that is used in Chapter 2. Put more simply, the differences in the in-person and telephone means may be due to a different mix of demographics for the persons responding to each mode rather than to a difference in the modes themselves.

Table 4.12– Unadjusted Means for Variables of Interest in three studies

Variable of Interest	Data Source	In-person	Telephone
Wage and salary income	1973 CPS Match data	12,021	13,390
Total family income	1973 CPS Match data	12,245	13,870
Personal income	2012 CPS March data	33,162	41,704
Health insurance coverage	2012 CPS March data	0.83	0.89

Chapter 5
Empirical Evaluations of Mixed-Mode Survey Inference Methods

5.1. Introduction

Following the Chapter 3 discussion of the theoretical statistical properties of the proposed methods, this chapter presents results from the empirical evaluations of the proposed methods using a subset of the data described in Section 4.2.1. The chapter starts by outlining the specific research questions. Then the related descriptives of the dataset used in the evaluations are discussed. The following section includes the description of the simulation and lists the covariates that are used in the imputation models. Later, the results based on the relative differences and absolute relative differences for each of the simulation variation are discussed.

For the empirical evaluation of the proposed method for mixed-mode survey inference, this chapter uses a subset of public-use Current Population Survey, 1973, and Social Security Records Exact Match data set. As described in Chapter 4, CPS is a mixed-mode survey and the 1973 CPS Match Data includes the 1972 person level Internal Revenue Service (IRS) income data. IRS income match data provide benchmarks to evaluate the proposed methods. While the analytical methods discussed in this paper are also applicable to the other survey items, wage and salary income is chosen for testing the proposed methods. To evaluate the proposed method empirically, random samples are drawn. Given a drawn sample, the standard method and the proposed methods are applied for wage and salary income. There are three parameters that are varied in the computations: 1) sample size (400 and 800), 2) item missing inclusion (included and excluded), and 3) imputation model (ignorable mode choice and nonignorable mode choice). The relative differences of the estimates are computed with respect to the population mean of wage and salary income and are compared across methods by 95% confidence intervals.

The proposed methods approach the mixed-mode survey response patterns as a special case of a missing data problem and use a series of multiple imputation models to create completed mode-specific data vectors conditioned on the observed data for response mode and sample unit covariates. These mode-specific completed data vectors are used to address two research questions in particular: (1) Are the measurement error differences between modes ignorable? and, (2) What are the properties of statistical inference methods that incorporate nonignorable measurement error differences under a mixed-mode survey design?

To explore the first research question, multiple imputation inference techniques are applied to the completed mode-specific data vectors to compute sample means and standard errors(Rubin, 1987). These means and standard errors are used to compare the differences in the mean estimates of the population distribution of the variable of interest by mode. To explore the second research question, the empirical properties of alternative methods in combining separate mode-specific mean estimates are investigated.

5.2. Current Population Survey, 1973, and Social Security Records: Exact Match Data[3]

In contrast to the 1973 CPS subset that is described in Chapter 4 and used as the basis to generate simulated populations in Chapter 6, the empirical evaluations are constrained to a subset of data that eliminates some of the possible measurement error sources that are not directly related to the mode of interview. The analysis dataset includes household heads, who:

- are married,
- reported a non-farm residence,

[3] [ICPSR 7616]. ICPSR version. Washington, DC: U.S. Dept. of Commerce, Bureau of the Census and Social Security Administration, Long-Range Research Branch [producer], 197?. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2001. doi:10.3886/ICPSR07616

- worked in a non-agricultural industry full-time in 1972,
- were married taxpayers filing jointly,
- whose spouses are present,
- whose spouses did not work in 1972,
- whose source of income is wages and salaries only,
- and had IRS matched records that were identified as a good-match.

In addition, among this subset persons who reported wage and salary income less than $600 were also excluded. Since there is no variation in the income value for the CPS and the IRS top-coded records, respondents who reported income greater than $50,000, and this proposed method is expected to be implemented on the raw data, these top-coded records were excluded from further analysis as well. The final sample size for this subset is n=5,422. In the simulations, this subset was considered as the population and random replicates were sampled to perform the empirical evaluations.

Table 5.1 reports the response mode distribution by wave, i.e., month in sample, for the 1973 sample. The distribution of response mode follows a similar pattern in this subset of data to the larger data set described previously in Chapter 4 (Table 4.3). While in-person mode is the dominant mode in the first and the fifth waves, telephone mode is preferred by about two-thirds of the sample in the other months, except the second month.

In this investigation, the variable of interest is the wage and salary income as reported in the CPS and the mean wage and salary income is the estimate of interest. Table 5.2 reports the unweighted quintiles of the wage and salary income by mode. Without controlling for individual-level and household covariates, the comparison suggests that the distribution of reported wage and salary income differs by mode in this subset of data. On average, in-person respondents earn $1,369 less per year than telephone respondents. After controlling for personal characteristics, education, work experience, race (white vs. other), occupation type (professional, sales, craft, laborer), and industry (construction, manufacturing, transportation, trade, service) and residential (household) characteristics, central city, suburb, region, the mean difference in average wage and salary earning for the two modes shrinks by about two thirds, but it is still significant.

As discussed in Chapter 2 and 3, this difference in the distribution of reported values for wage and salary income is not due to measurement error only, it is likely due to both mode choice and mode effects. By the notation that is used in Chapter 2, this difference is $(\bar{Y}_T \mid P=T)-(\bar{Y}_I \mid P=I)$. Given the nonrandom assignment nature of the CPS, the significant difference could not be attributed only to mode choice, i.e. people with higher income are likely to choose telephone, or to mode effects.

Since the 1972 person level IRS wage and salary income data are available, they can be compared against CPS-reported wage and salary income for this same year. The average relative differences by mode,

$$(\mathrm{RelDif}_{U_T} \mid P=T)=\frac{\sum_{j\in U_T}(y_j-y_{IRS})}{\sum_{j\in U} R_{Tj}} \text{ and } (\mathrm{RelDif}_{U_I} \mid P=I)=\frac{\sum_{j\in U_I}(y_j-y_{IRS})}{\sum_{j\in U} R_{Ij}}$$

are not significantly different between the in-person and the telephone modes (p=0.06).

The mode choice model covariates, $X^{(R)}$, and the outcome model covariates, $X^{(Y)}$ used in this chapter are the same as in the GRZ selection and outcome models with two exceptions. In this dissertation, the response mode is the dependent variable in the mode choice model, and month in sample (MIS) is included as one of the mode choice model covariates.

GRZ and the extensions of their work have studied the properties of the imputation models for the item missing data in reported wage and salary income (Greenlees et al., 1982; Glynn et al., 1986, 1993). This dissertation includes an indicator for whether an item is missing or not as a simulation parameter. While the overall item missing percent for this subset is 10%, the telephone mode yielded a higher item nonresponse rate (Table 5.3). Here, item missing data includes both the refusals and the other types of missing data.

Table 5.1 – Response Mode Distribution by Month in Sample

Response Mode	Month in Sample							
	M1	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**
Telephone	3%	35%	64%	72%	8%	59%	72%	69%
In-person	97%	65%	36%	28%	92%	41%	28%	31%

Table 5.2 – Sample Quintiles of Reported Wage and Salary Income

Response Mode	Quintiles						
	0%	10%	25%	50%	75%	90%	100%
Telephone	1,000	7,300	9,400	12,000	15,800	21,000	48,000
In-person	900	6,000	8,284	11,000	14,500	19,300	45,000

Table 5.3 – Sample Percentage of Item Missing in Reported Wage and Salary Income

Response Mode	% of Item Missing
Telephone	12%
In-person	8%
Overall	10%

5.3. Evaluation of Proposed Mixed-Mode Methods of Survey Inference

To assess the proposed mixed-mode methods of survey inference, the 1973 CPS Match data set on wage and salary income was used as a "population" to derive samples and simulate the performance of the proposed methods and the standard method. A total of eight simulations were performed varying three parameters: (1) Replicate sample size (400 and 800), (2) Whether to include households with missing incomes, i.e., item missing, in the imputations or not, and (3) Imputation model specification: ignorable mode choice regression model versus nonignorable mode choice regression model (see Section 3.3.3). As shown in Tables 5.4 and 5.5, in this set of imputations, Greenlees et al. (1982) mode choice and response model structures were used. In addition, month in sample was included in the models.

Table 5.4 – Covariates included in the Mode Choice Imputation Models (R-Model)

Covariate	**Definition**
V1040	Household head Age in years
EDUCATION	Number of years of education completed by the household head
WHITE	Unity if the race of the house- hold head is white; zero otherwise
NORTH	Unity if the household resides in the North Central region; zero otherwise
SOUTH	Unity if the household resides in the South region; zero otherwise
WEST	Unity if the household resides in the West region; zero otherwise
V1001	Month in sample

Table 5.5 – Covariates included in the Response Imputation Models (Y-Model)

Covariate	Definition
EDUCATION	Number of years of education completed by the household head
EDUCATION2	EDUCATION squared
EXPERIENCE	AGE - EDUCATION
EXPERIENCE	EXPERIENCE squared
WHITE	Unity if the race of the house- hold head is white; zero otherwise
CENTRALCITY	Unity if the household resides in the central city of an SMSA; zero otherwise
SUBURB	Unity if the household resides in the ring of the SMSA; zero otherwise
NORTH	Unity if the household resides in the North Central region; zero otherwise
SOUTH	Unity if the household resides in the South region; zero otherwise
WEST	Unity if the household resides in the West region; zero otherwise
PROFESSIONAL	Unity if the household head's occupation is professional or managerial; zero otherwise
SALES	Unity if the household head's occupation is sales or clerical; zero otherwise
CRAFT	Unity if the household head's occupation is craft or operative; zero otherwise
LABORER	Unity if the household head's occupation is laborer; zero otherwise
CONSTRUCTION	Unity if the household head is employed in the construction or mining industries; zero otherwise
MANUFACTURING	Unity if the household head is employed in the manufacturing industry; zero otherwise
TRANSPORTATION	Unity if the household head is employed in the transportation, communication, or utilities in- dustries; zero otherwise
TRADE	Unity if the household head is employed in the wholesale or retail trade industries; zero otherwise
SERVICE	Unity if the household head is employed in the personal service, entertainment, or recreation service industries; zero otherwise
V1001	Month in sample

An equal number of respondents was drawn from each wave in each of the replicates under fixed sample sizes of 400 and 800 from the subset of the

CPS data as defined. The MIS variable is used as the stratum variable in selection to preserve the mode choice mechanism characteristics in each simulated sample. The MIS variable is also included in the mode choice model. The completed data sets were then created for telephone and in-person modes using both imputation models.

The completed data sets were used to compute the mode-specific mean wage and salary income. Under the proposed method, multiple imputations were combined using the usual multiple imputation combination rules to produce mode-specific means (Rubin, 1987). The distributions of the simulation sample estimates of the mode-specific means for wage and salary income were compared in terms of three evaluation criteria: (1) Number of significant differences, (2) Mean absolute relative difference, and (3) Mean relative difference. The mode-specific estimates were combined under four methods: (1) Simple average (Combination Method 1 – CM1), (2) Minimum variance (Combination Method 2 – CM2), (3) Minimum mean square error (Combination Method 3 – CM3), and (4) Ignoring the measurement differences (the standard combination method – CM4). The methods are described in Section 3.3.4. Each simulation included 50 replicates and 5 imputations per replicate.

5.4. Results

Table 5.6 summarizes the first part of the results from the simulation exercise. A larger number of significant differences between the mode-specific means were observed for the sample size=800 simulations at 95% confidence level for any of the simulations (see Table 5.6). But the largest proportion of significant differences was only 22%. In particular, 11 out of 50 samples generated significant differences under the nonignorable mode choice imputation model (excluding cases with item missing data on the variable of interest).

Figures 5.1-5.4 report the second and third parts of the results. Figures 5.1 and 5.2 correspond to the relative and absolute relative differences under the ignorable mode choice imputation model. In parallel, Figures 5.3 and 5.4 report relative and absolute relative differences under the nonignorable mode choice imputation model.

First, the discussion focuses on the absolute relative differences and then moves to the relative differences. In addition to the differences between the combination method estimates, differences between the mode-specific estimates are of interest. As discussed in Section 3.3.2.3, $\mathrm{E}_R\mathrm{E}_M\mathrm{E}_{IMP}\left[\bar{Y}_T^* - \bar{Y}_I^*\right] = \bar{B}_{UT} - \bar{B}_{UI}$. So $(\bar{Y}_T^* - \bar{Y}_I^*)$, i.e. mode-specific estimates, can be used to estimate the difference in the average difference in mode effects in the population. This difference incorporates the possibility that the effect of mode can differ among persons. In this set of comparisons, we compare the absolute relative differences for $\bar{Y}_T^*$ and $\bar{Y}_I^*$ to investigate the ignorability of the mode effects. The absolute relative difference was larger for the in-person mode-specific means on the average than the ones for the telephone mode-specific means in the sample size=800 under the ignorable mode choice imputation model simulation (Figure 5.2). The corresponding sample size=400 simulations could not capture the significant difference.

Based on the ignorable mode choice mode, Figure 5.5 and Figure 5.6 show the patterns for relative differences and absolute relative differences for in-person and telephone mode-specific estimates of means compared to those for the standard method (CM4) estimates. Similarly, Figure 5.7 and Figure 5.8 show relative differences and absolute relative differences patterns under the nonignorable mode choice imputation model.

Figures 5.5-5.8 further decomposes the relative and absolute differences for in-person and telephone means into observed and imputed means. For example, in-person means are decomposed into in-person observed data means (inperson.O) and in-person imputed data means (inperson.I). The relative and the absolute relative differences for the standard method lies between the relative differences and the absolute differences for the in-person observed and the telephone observed.

To isolate the mode choice and to understand the direction of the imputation error, differences between the observed and the imputed data conditioned on the subset of in-person and telephone respondent subsets can be explored. For example, differences in relative bias between the in-person observed and the telephone imputed can be used to understand the direction of the imputation error by isolating the mode choice. In Figure 5.5 and Figure 5.7, the imputed data have greater relative difference and absolute relative difference on the average compared to the observed data. In Figure 5.5 and Figure 5.7, one clear pattern is the direction of the relative differences for the telephone imputed data. In contrast to the other data, the mean of relative differences is positive. The effect of this opposite direction can be seen in the absolute relative differences for the telephone mode-specific estimates as shown in Figure 5.6 and Figure 5.8. The mean absolute relative difference does not lie between the means for telephone observed and imputed data. As the Figure 5.6 and Figure 5.8 show the mean of absolute relative difference for the standard method (CM4) lies between the means of absolute relative differences for in-person and telephone observed data as the direction of the relative differences are same for the means.

The difference was eliminated under the nonignorable mode choice imputation model (Figure 5.4). The same pattern was observed for both simulation variations of including or excluding the item missing cases. Since the differences could be explained by the nonignorable mode choice imputation model, the mode effects are considered to be ignorable for this particular dataset and wage and salary income. For the interest of the completeness of the exercise, combination methods were applied on mode-specific estimates.

In terms of the absolute relative difference, the Combination Methods 1, 2 and 3 yielded the same difference levels on the average across all the variations of the simulations. These three methods outperformed the standard method (CM4) in which telephone and in-person responses were combined without any adjustments. The standard method (CM4) yielded substantial negative relative difference across all the simulation variations. These differences for the standard method (CM4) were greater in the simulations where item missing data on wage and salary income was imputed. This is some evidence that a separate mechanism should be considered for the item missing imputations. Future research will include the extension to include a separate imputation for the item missing values.

For a population mean estimation in addition to absolute relative differences, relative differences are also a research interest. Figures 5.1 and 5.3 show relative differences for two imputation models, ignorable mode choice and nonignorable mode choice, respectively, for the corresponding estimation methods. Under the ignorable mode choice imputation model (Figure 5.1), the only case where the confidence interval includes zero is CM3 (for sample size=800 simulations), which is the minimum mean squared error combination. Under the nonignorable mode choice imputation model, the confidence intervals include zero for CM1, CM2, and CM3. But, for sample size=800 and item missing excluded condition, none of the confidence intervals include zero. This is further evidence to that imputation models may need to incorporate factors for whether items are missing or not.

However, note that CM1, CM2, and CM3 are all closer to being unbiased in Figures 5.1 – 5.4 than the standard method CM4, which ignores the possibility of mode biases. This is true for both sample sizes and regardless of whether items with missing values are included in the imputations.

Table 5.6 – Number of Significant Differences at 95% confidence level between Telephone and In-person Mode-specific Estimates

	400x50x5		**800x50x5**	
	Ignorable Mode Choice Regression Model	**Nonignorable Mode Choice Model**	**Ignorable Mode Choice Regression Model**	**Nonignorable Mode Choice Model**
Item Missing Excluded	3	5	6	11
Item Missing Included	0	3	3	9

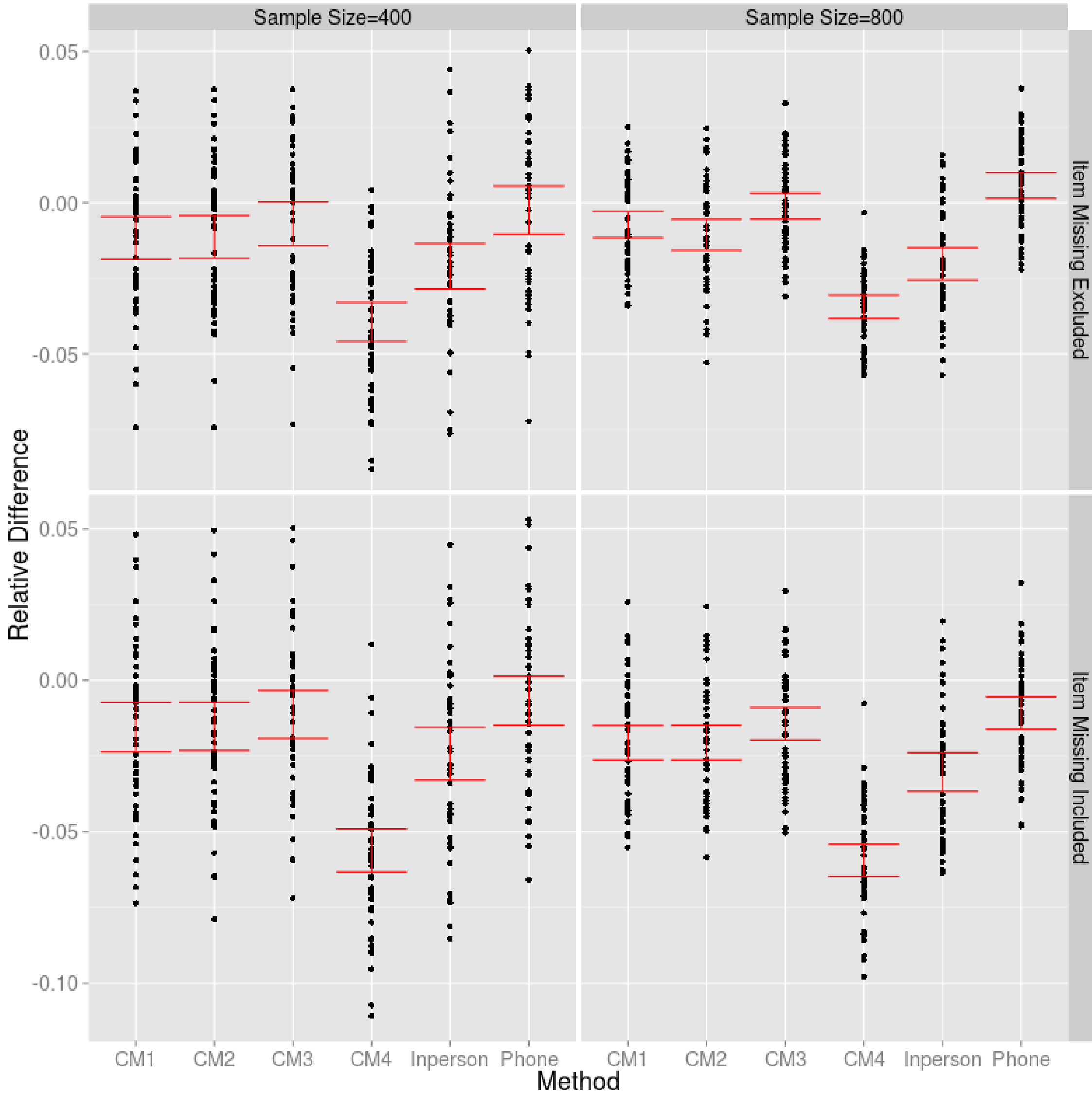


Figure 5.1 – Under the ignorable mode choice imputation model relative differences ($\text{RelDiff}_{CM} = \left(\bar{y}_{CM_i} - \bar{y}_{IRS} / \bar{y}_{IRS}\right)$, where CM_i for $i = 1,2,3,4$ is the combination method)[4] in 50 samples[5] of estimates of wage and salary income mean with the four alternative methods of estimation (CM1, CM2, CM3, and CM4=Standard Combination Method) and mode-specific imputed data (Telephone and In-person) by item missing treatment procedure. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean relative difference.

[4] Same formula is used for the mode specific mean estimates in which CM_i is replaced by the telephone and in-person estimates.
[5] The model parameters are not estimated in one replicate in sample size=400 simulations due to zero sample size cells

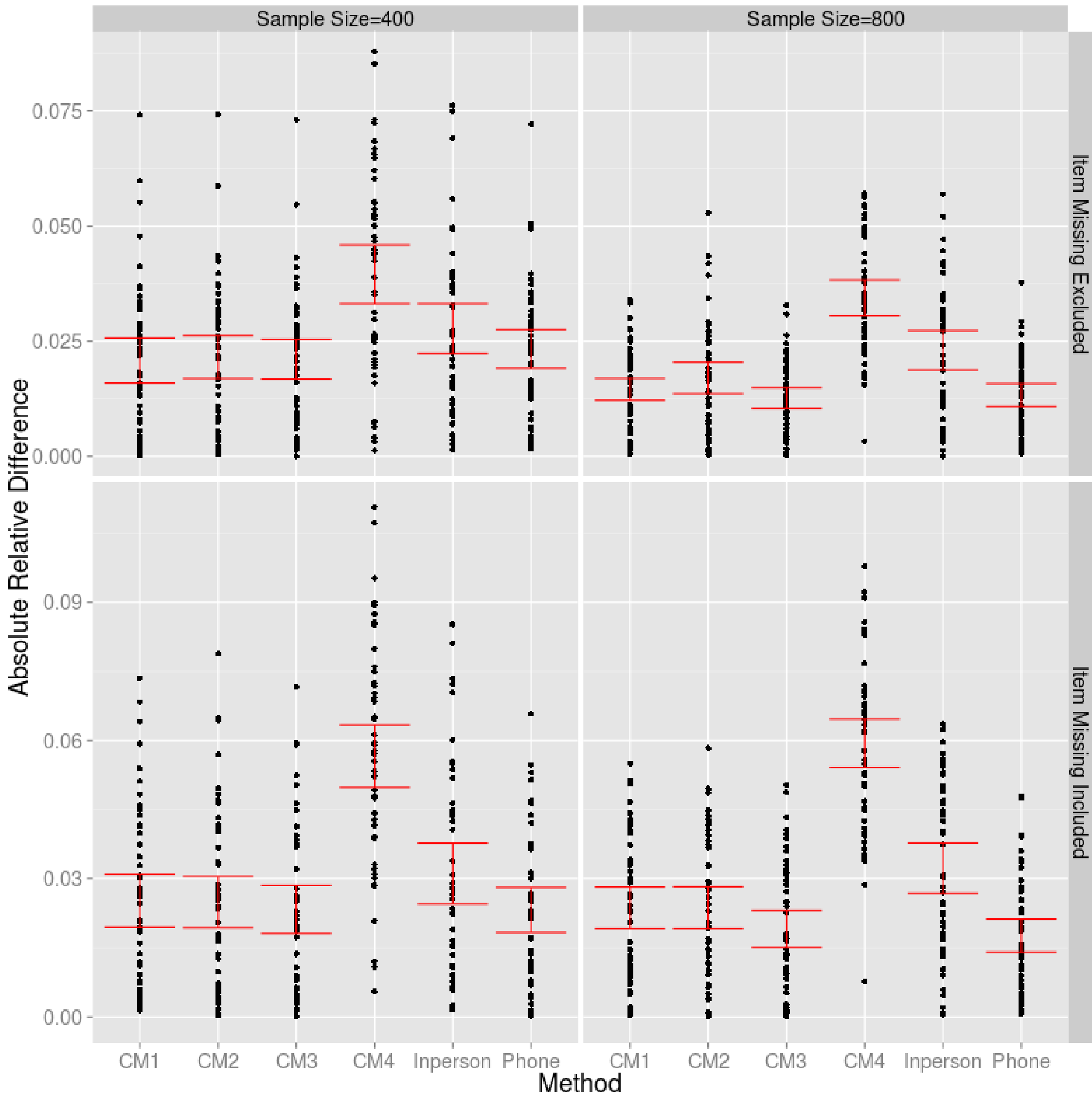


Figure 5.2 – Under the ignorable mode choice imputation model absolute relative differences ($\text{AbsRelDiff}_{CM} = \left(\left| \bar{y}_{CM_i} - \bar{y}_{IRS} \right| \big/ \bar{y}_{IRS} \right)$**, where** CM_i **for** $i = 1, 2, 3, 4$ **is the combination method) in 50 samples[6] of estimates of wage and salary income mean with the four alternative methods of estimation (CM1, CM2, CM3, and CM4=Standard Combination Method) and mode-specific imputed data (Telephone and In-person) by item missing treatment procedure. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean absolute relative difference.**

[6] The model parameters are not estimated in one replicate in sample size=400 simulations due to zero sample size cells.

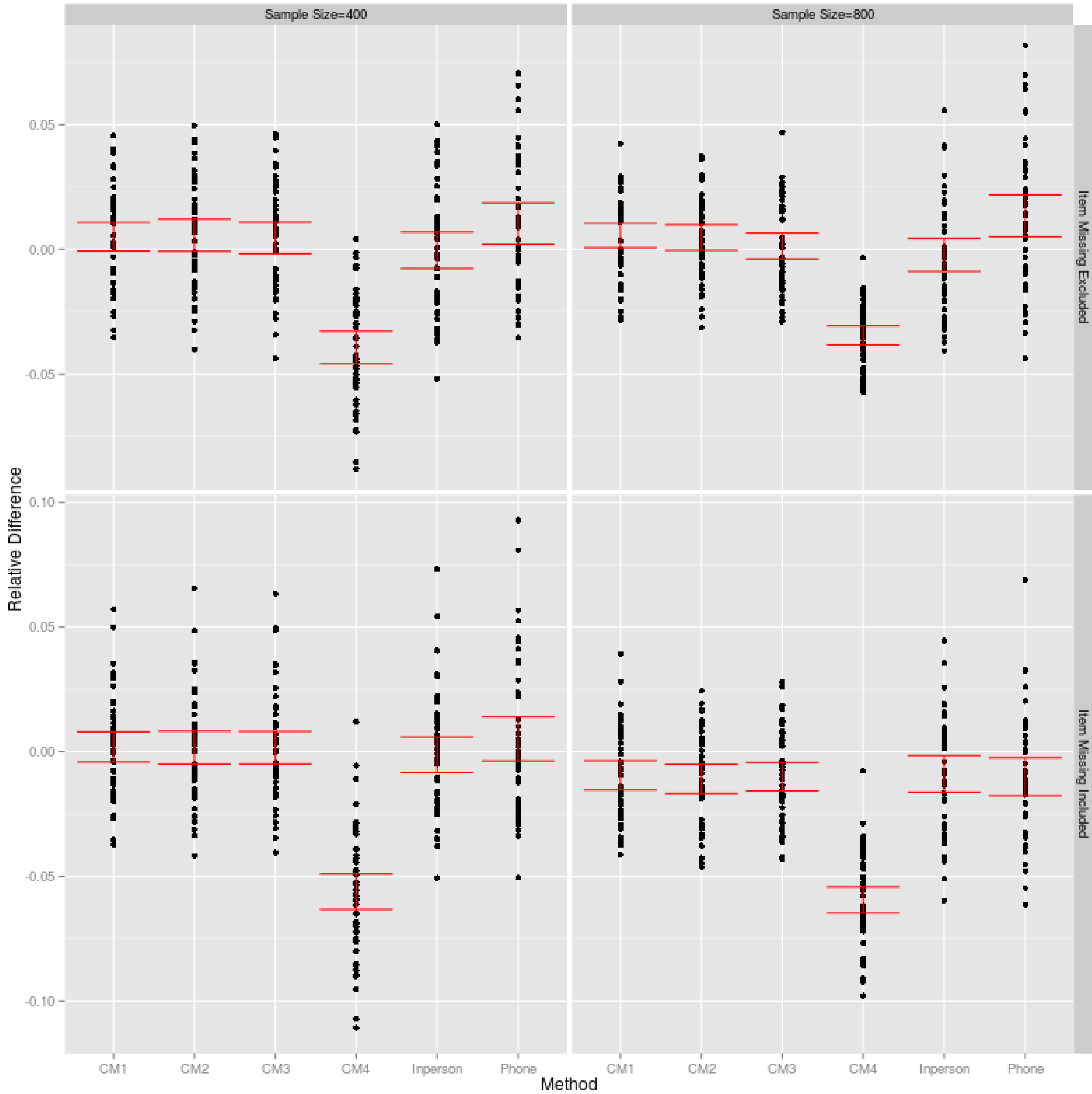


Figure 5.3 – Under the nonignorable mode choice imputation model relative differences ($\mathrm{RelDif}_{CM} = \left(\bar{y}_{CM_i} - \bar{y}_{IRS} / \bar{y}_{IRS}\right)$**, where** CM_i **for** $i = 1,2,3,4$ **is the combination method)** [7] **in 50 samples** [8] **of estimates of wage and salary income mean with the four alternative methods of estimation (CM1, CM2, CM3, and CM4=Standard Combination Method) and mode-specific imputed data (Telephone and In-person) by item missing treatment procedure. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean relative difference.**

---

[7] Same formula is used for the mode specific mean estimates in which *CM* is replaced by the telephone and in-person estimates.

[8] The model parameters are not estimated in three replicate in sample size=400 simulations due to zero sample size cells. The model parameters are not estimated in three replicates in sample size=800 exclude item missing and five replicates in include item missing simulations due to zero sample size cells.

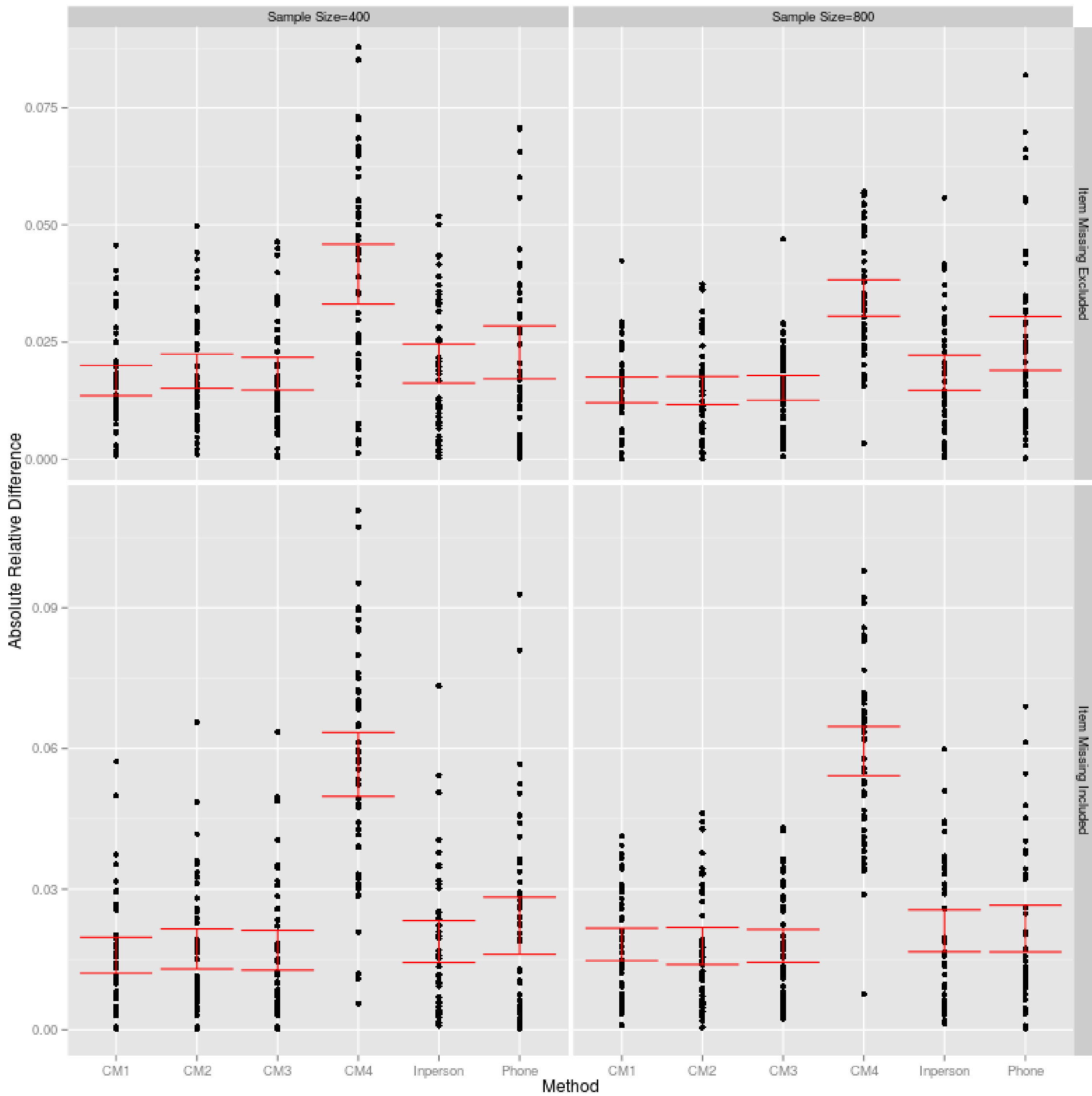


Figure 5.4 – Under the nonignorable mode choice regression model absolute relative differences ($\text{AbsRelBias}_{CM} = \left(\left| \bar{y}_{CM_i} - \bar{y}_{IRS} \right| / \bar{y}_{IRS} \right)$, where CM_i for $i = 1,2,3,4$ is the combination method)[9] in 50 samples[10] of estimates of wage and salary income mean with the four alternative methods of estimation (CM1, CM2, CM3, and CM4=Standard Combination Method) and mode-specific imputed data (Telephone and In-person) by item missing treatment procedure. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean absolute relative difference.

[9] Same formula is used for the mode specific mean estimates in which CM_i is replaced by the telephone and in-person estimates.

[10] The model parameters are not estimated in three replicate in sample size=400 simulations due to zero sample size cells. The model parameters are not estimated in three replicates in sample size=800 exclude item missing and five replicates in include item missing simulations due to zero sample size cells.

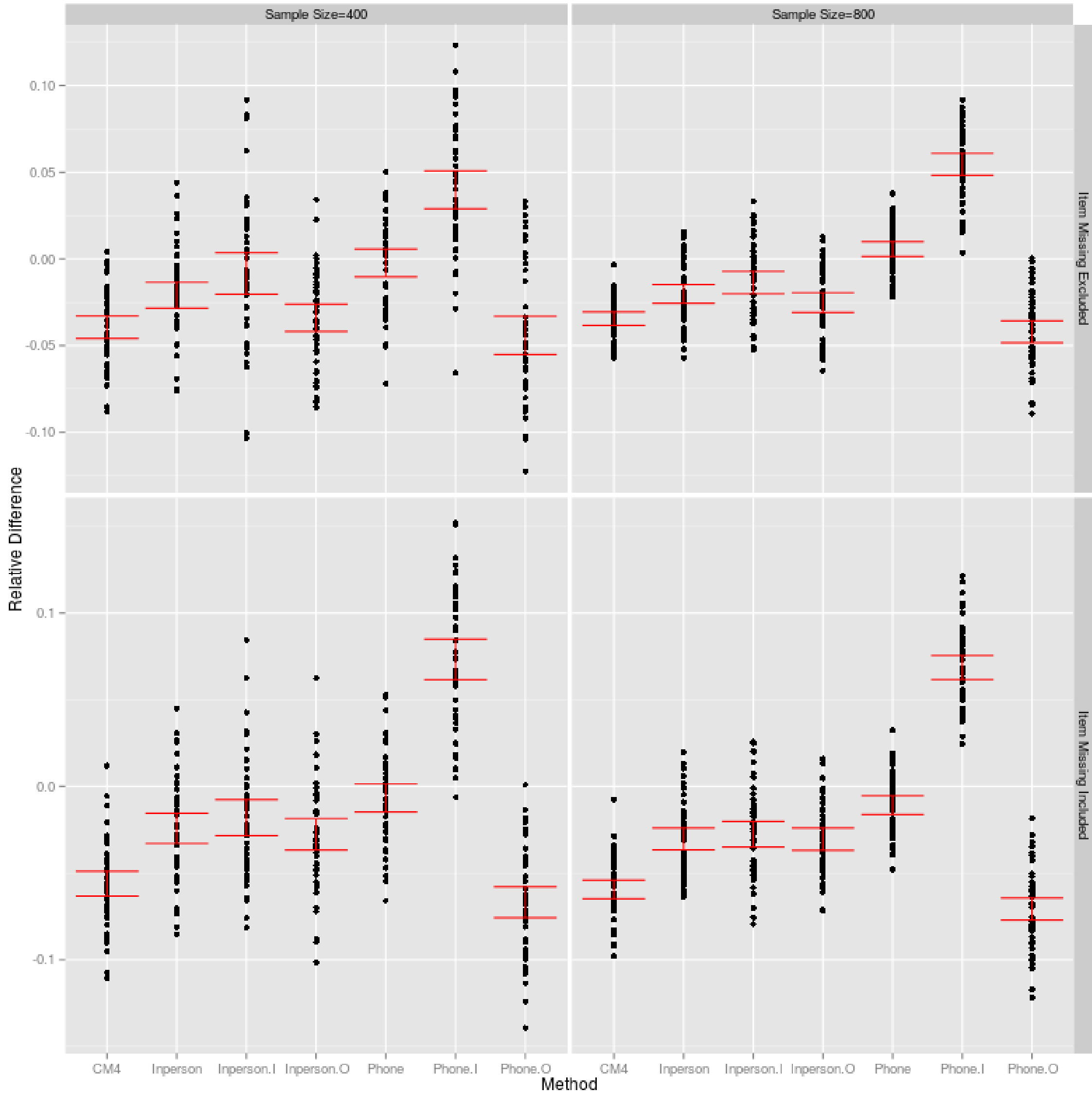


Figure 5.5 – Under the ignorable mode choice imputation model relative differences ($\text{RelDif} = \bar{y} - \bar{y}_{IRS} / \bar{y}_{IRS}$) in 50 samples[11] of estimates of wage and salary income mean with the CM4 (Standard Combination Method) and mode-specific completed (Telephone and In-person) data by item missing treatment procedure. Telephone.O and Inperson.O data represent relative differences for mean wage and salary income based on the observed data. Telephone.I and Inperson.I data represent relative differences for mean wage and salary income based on the imputed data. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean relative difference.

[11] The model parameters are not estimated in one replicate in sample size=400 simulations due to zero sample size cells

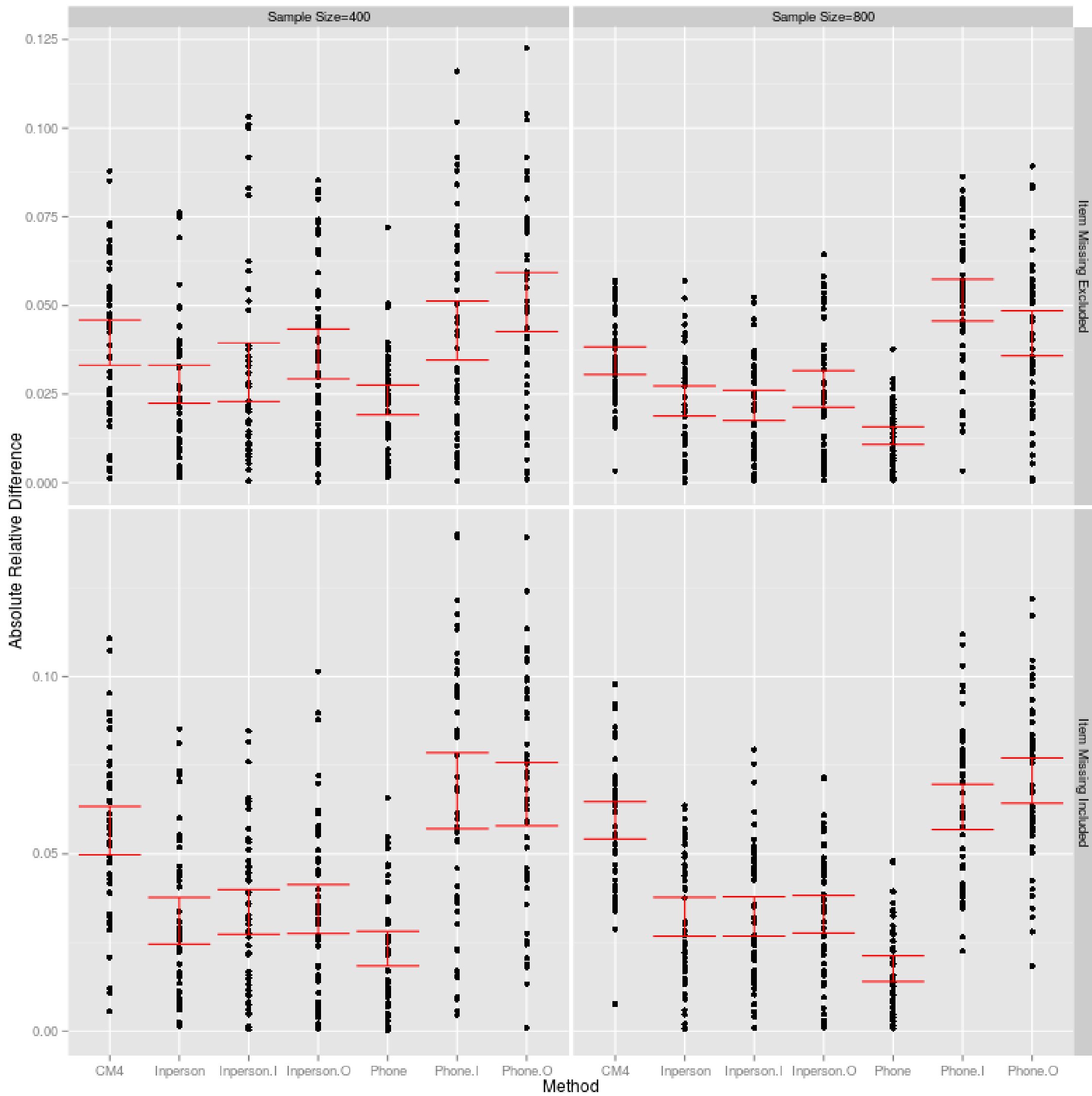


Figure 5.6 – Under the ignorable mode choice imputation model absolute relative differences ($\text{AbsRelDif}=\left|\bar{y}-\bar{y}_{IRS}\right|/\bar{y}_{IRS}$ **) in 50 samples[12] of estimates of wage and salary income mean with the CM4 (Standard Combination Method) and mode-specific completed (Telephone and In-person) data by item missing treatment procedure. Telephone.O and Inperson.O data represent absolute relative differences for mean wage and salary income based on the observed data. Telephone.I and Inperson.I data represent absolute relative differences for mean wage and salary income based on the imputed data. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean absolute relative difference.**

[12] The model parameters are not estimated in one replicate in sample size=400 simulations due to zero sample size cells

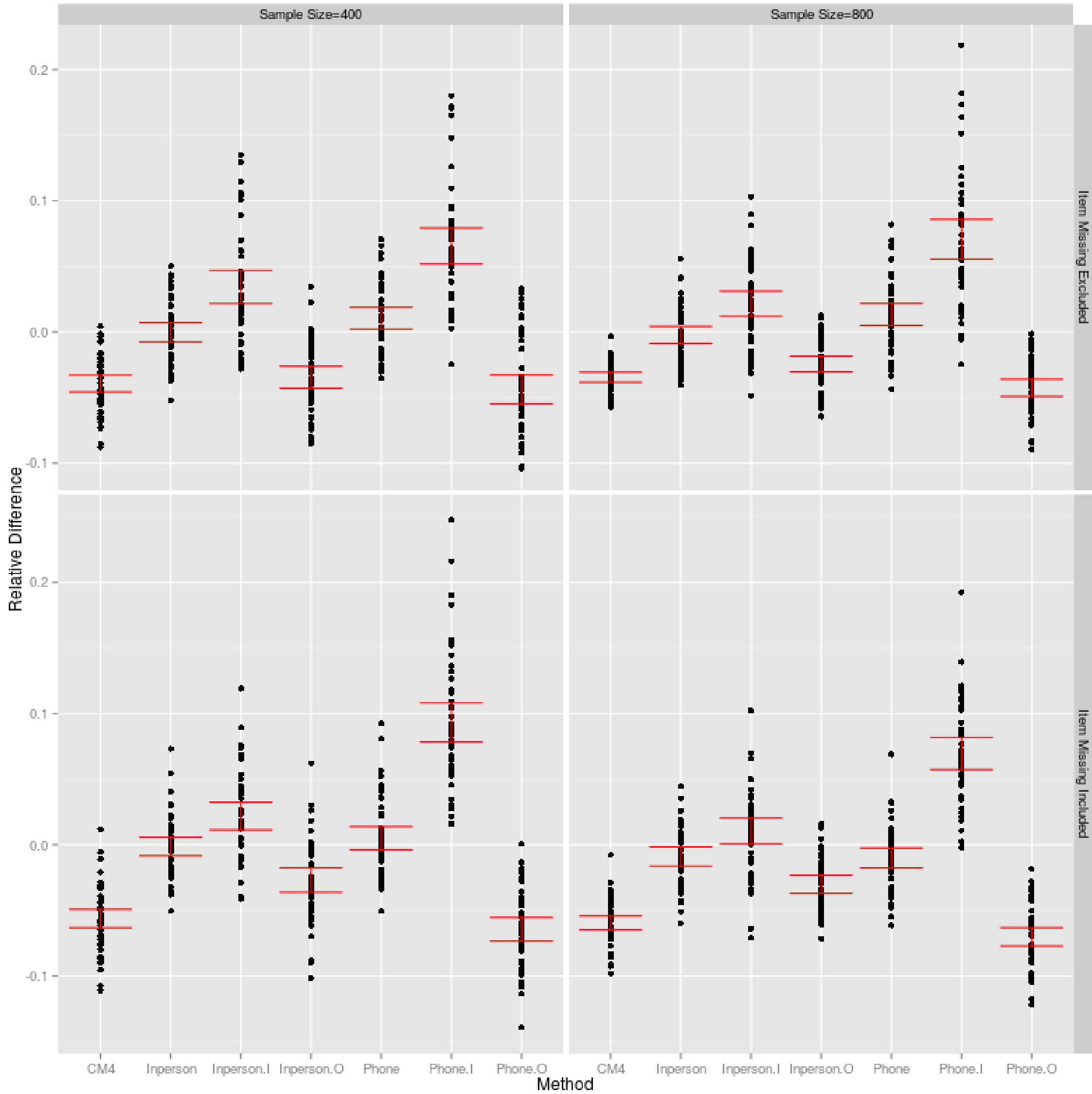


Figure 5.7 – Under the nonignorable mode choice imputation model relative differences (RelDif= $\bar{y} - \bar{y}_{IRS} / \bar{y}_{IRS}$ **) in 50 samples[13] of estimates of wage and salary income mean with the CM4 (Standard Combination Method) and mode-specific completed (Telephone and In-person) data by item missing treatment procedure. Telephone.O and Inperson.O data represent relative differences for mean wage and salary income based on the observed data. Telephone.I and Inperson.I data represent relative differences for mean wage and salary income based on the imputed data. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean relative difference.**

[13] The model parameters are not estimated in one replicate in sample size=400 simulations due to zero sample size cells

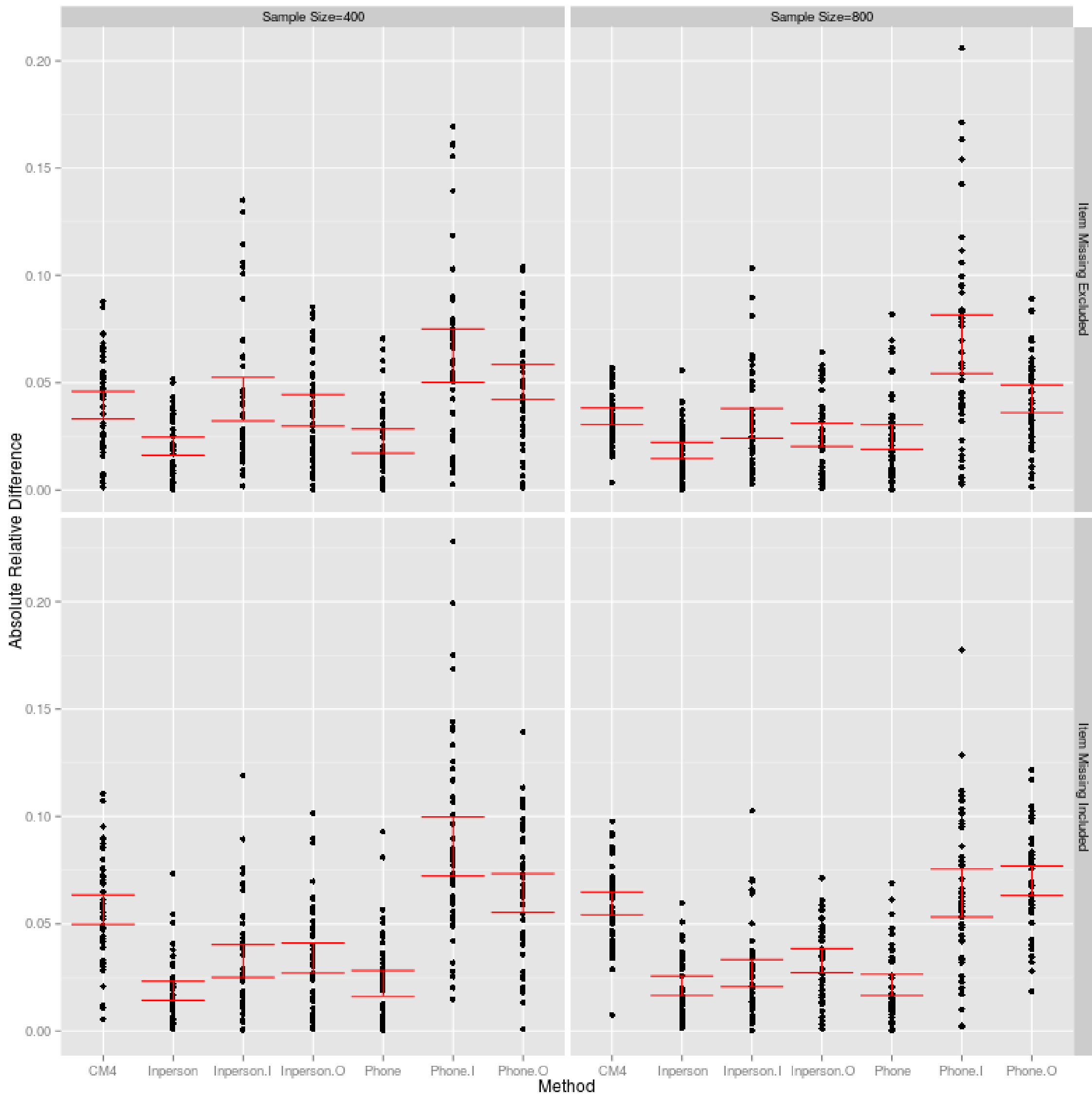


Figure 5.8 – Under the ignorable mode choice imputation model absolute relative differences (AbsRelDif=$\left|\bar{y}-\bar{y}_{IRS}\right|/\bar{y}_{IRS}$ **) in 50 samples[14] of estimates of wage and salary income mean with the CM4 (Standard Combination Method) and mode-specific completed (Telephone and In-person) data by item missing treatment procedure. Telephone.O and Inperson.O data represent absolute relative differences for mean wage and salary income based on the observed data. Telephone.I and Inperson.I data represent absolute relative differences for mean wage and salary income based on the imputed data. Sample sizes are 400 and 800 each for the samples; five imputations were performed for each sample, the red error bars represent the 95% confidence intervals for the mean absolute relative difference.**

[14] The model parameters are not estimated in one replicate in sample size=400 simulations due to zero sample size cells

5.5. Discussion and Extensions

The first of our research questions was whether mode choice could be ignored for estimation. Related to the first research question, we found only in one simulation variation that the absolute relative differences of the in-person mode-specific estimates were higher than those of the telephone estimates. On the other hand, the error sources in the mode-specific estimates are not known without studying the error sources in a randomized experiment, the differences in the relative differences for the mode-specific means were only evaluated by controlling the available covariates analytically. The sample size=400 simulation did not have the power to capture the significant differences. Also the nonignorable mode choice imputation model eliminated the differences between telephone and in-person.

Our second research question was whether improved estimators could be developed that accounted for the possibility that modes might have different biases. In addressing the second research question, all the proposed combination methods outperformed the standard method in all of the simulations. For the combined estimator (3.22), α_{opt} is given in (3.23). α_{opt} depends on $\rho_{TI} = corr(\bar{Y}_T^*, \bar{Y}_I^*)$, v_T and v_I, and $\bar{B}_{UT}$ and $\bar{B}_{UI}$. As defined in the Section 3.3.2.3, $v_T = \mathrm{var}_M(\bar{Y}_T^*)$, $v_I = \mathrm{var}_M(\bar{Y}_I^*)$, i.e. v_T and v_I are variances for the mode-specific estimates with respect to the statistical error model. $\bar{B}_{UT}$ and $\bar{B}_{UI}$ are average of mode specific systematic reporting errors as defined in the measurement error model in (3.1). The model inspection of the error term variance suggested that $v_T = v_I$. Furthermore, earlier analysis suggested that $\bar{B}_{UT}$ could be very close to $\bar{B}_{UI}$. In that case, α_{opt} reduces to ½. So, CM1 (which is a simple average), CM2 (which weights inversely according to v_T and v_I) and CM3 (which weights inversely according to mean square error) all implicitly use α_{opt} conditioned on ρ_{TI}. Consequently, they are theoretically expected to outperform the standard method that ignores mode effects. Ignorable mode effects and equal variance properties yield a special case of the combination weights in CM_3 that minimizes the mean square error of combined estimator. Empirical findings supported this expectation.

Some households in the CPS/IRS income study were missing the value of income. In this investigation, a separate selection mechanism was not considered in the imputation for households with missing incomes. In one variation of the simulations, the item missing was treated the same as the other mode responses and were imputed by the same multiple selection models. We observe that the differences in relative differences increased for the simulations in which item missing are imputed as well. This suggests that there may be a different mechanism that needs to be included for the treatment of item missing in the mode-specific imputations.

Chapter 6
Simulation Evaluations of Mixed-Mode Survey Inference Methods

6.1. Introduction

Chapter 6 includes two studies: (1) a simulation study that investigates the performance of the proposed and the standard methods (Section 6.2), and (2) empirical comparison analyses that study possible differences in the estimates generated by the proposed and the standard methods (Section 6.3). In the simulation study, benchmarks are available so that the performance of the methods can be evaluated with respect to the benchmarks. On the other hand, empirical comparison analysis is more appropriate for a more common situation where no benchmarks are available. The simulation study uses 1973 CPS match data to create hypothetical populations as described in Section 6.2.2. The performance of the proposed and the standard methods are evaluated by relative bias as discussed in Section 6.2.5. The empirical comparison analysis uses the bootstrap replicates from 2012 CPS March data to compute mean personal income and percent health insurance coverage by each method. Replicate mean personal income and percent health insurance coverage per method are compared using repeated measurement ANOVA. Empirical comparison analyses results are discussed in Section 6.3.5.

6.2. Simulation Study

6.2.1. Motivation

In the first part of the chapter, results from the simulations using hypothetical populations based on the 1973 CPS match data are discussed. The 1973 CPS match data were limited in evaluating the performance of the proposed method in the presence of nonignorable mode effects. Consequently, hypothetical populations were created using the 1973 CPS match data by introducing known mode effects to the CPS-based hypothetical population data on a range from severe underreporting to severe overreporting for the in-person responses. To eliminate the sampling error, samples were not drawn to apply the proposed methods in contrast to the other simulation studies. The alternative methods of constructing estimators were applied to each full, hypothetical population rather than to samples from the full population.

6.2.2. Creation of Hypothetical Populations: Models used to Generate the Hypothetical Populations

Each simulation included two hypothetical populations corresponding to telephone and in-person responses for a given person. Hypothetical populations were generated using the CPS-IRS data. The mode by which each person responded in the CPS was retained. This is a special case of the measurement error model described in (2.9).

Income values were generated using models built on the IRS AGI. One income value was generated for each person in each hypothetical population, depending on the mode in which the person had actually responded in the CPS and the assigned degree of in-person mode effect perturbation. If the person responded by telephone, the relationship between CPS-reported income and IRS AGI was used. If a person responded in-person, the relationship of the generated income value and the IRS AGI was controlled, as described below. The true income value for each person *j* was the IRS AGI, $Y_j^{(AGI)}$. In the first step, an initial artificial predicted value was computed for each person as $Y_{pj}^{*} = \beta_p^{(AGI)} Y_j^{(AGI)}$ where $\beta_p^{(AGI)}$ is a slope parameter. For CPS telephone respondents, the slope from the actual survey data was used. For CPS in-person respondents, the slope was varied from 0.1 to 2, which corresponds to relative reporting errors of -0.9 to 1. This technique creates one mode effect for the telephone values and several mode effects for in-person values, depending on the slope that was used.

To add variation to the generated income values while retaining the relationship to various demographics, an artificial income value was generated as $\hat{Y}_j = X_j^{(Y)}\hat{\beta}^{(Y)} + e_j$ where the $\hat{\beta}^{(Y)}$ and $\hat{\sigma}^2$ were estimated by regressing Y_{pj}^* , from the first step, on a set of demographic covariates collected in the CPS. The covariates are listed in Table 6.1. Tables 4.1 and 4.2 present the distributions for these variables. The error in the model was distributed as $e_j \overset{iid}{\sim} N(0, \hat{\sigma}^2/c)$ with $c \in (0.5,10,15)$ to reflect different degrees of model fit from good to poor.

Table 6.1 – Covariates included in the Response Models (Y-Model)

Covariate	**Definition**
EDUCATION	Number of years of education completed by the household head
EDUCATION2	EDUCATION squared
EXPERIENCE	AGE - EDUCATION
EXPERIENCE	EXPERIENCE squared
V1040	Sex of household head, 1- male
WHITE	Unity if the race of the household head is white; zero otherwise
V1026	SMSA residence, 1- Not in SMSA, 2- in SMSA: Central City, 3- in SMSA: Ring
V1029	Tenure of occupied residence, 1: Rented for cash, 0: Other
LVNGQRTS	Living quarters, 1- Trailer, permanent, 0- Other
V1023	Region, 1- Northeast, 2- North Central, 3- South, 4- West
WORKERCLASS	Household head's occupation, 1- Other, 2- Professional, 3- Sales, 4- Craft, 5- Laborer
INDUSTRY	Industry the household head is employed, 1- Other, 2- Agriculture, 3- Manufacturing, 4- Transportation, 5- Trade, 6- Service
CPSINCOMETYPE	Income type, 1- None, 2- Wages only, 3- Self-employed only, 4- Wages and Self-employed, 5- Other and self-employed, 6- Other and wages, 7- Wages, self-employed and other
V1001	Month in sample
V1067	Part time/ Full time status in 1972, 1- Full year, full time, 2- Part year, full time, 3- Full year, part time, 4- Part year, part time
WELFARE_R	Welfare receipt status
KDSP	Spouse's and Kids' working status, 1- Single, 2- Married, spouse not working, no kids, 3- Married, spouse working, no kids, 4- Married, spouse not working, kids not working, 5- Married, spouse working, kids not working, 6- Married, spouse not working, kids working, 7- Married, spouse working, kids working
MARCHRESP	Unity if the household head is the respondent in March 1973

Table 6.2 – Covariates included in the Mode Choice Models (R-Model)

Covariate	Definition
V1001	Month in sample
EDUCATION	Number of years of education completed by the household head
INDUSTRY	Industry the household head is employed
KDSP	Spouse's and Kids' working status
V1023	Region
V1026	SMSA residence
V1029	Tenure of occupied residence
V1001 x EDUCATION	Month in sample x Number of years of education completed by the household head
V1001 x INDUSTRY	Month in sample x Industry the household head is employed
V1001 x KDSP	Month in sample x Spouse's and Kids' working status
V1001 x V1023	Month in sample x Region
V1001 x V1026	Month in sample x SMSA residence
V1001 x V1029	Month in sample x Tenure of occupied residence

6.2.3. Hypothetical Populations Characteristics

Through hypothetical populations, we created a situation where the size of the mode effects (for in-person respondents) is explicitly controlled. The simulation data were created using real data and regression analysis was constructed based on the real data associations. To understand the impact of introducing mode effects on the response data, distributions for the hypothetical populations were examined. Figure 6.1 shows an example of the simulated distributions of survey income responses by telephone and in-person modes and IRS AGI data. For the telephone responses, the distribution for the condition beta=0.97 closely follows the IRS AGI data distribution. For the in-person responses, there is a shift in the location of the distribution as a result of introducing mode effects (beta=0.7). For both telephone and in-person responses, changing the constant in the error term variance spreads the distribution wider as expected. In addition to response distributions, the mode choice mechanism with respect to income distribution was preserved. Figure

6.2 shows the mean income for simulated populations by telephone and in-person response mode. The means of total family income for telephone and in-person respondents were computed as $(\bar{Y}_T \mid P = T)$ and $(\bar{Y}_I \mid P = I)$ in Figure 6.2. Figure 6.2 shows that, as expected, the lower mean income for in-person respondents was preserved.

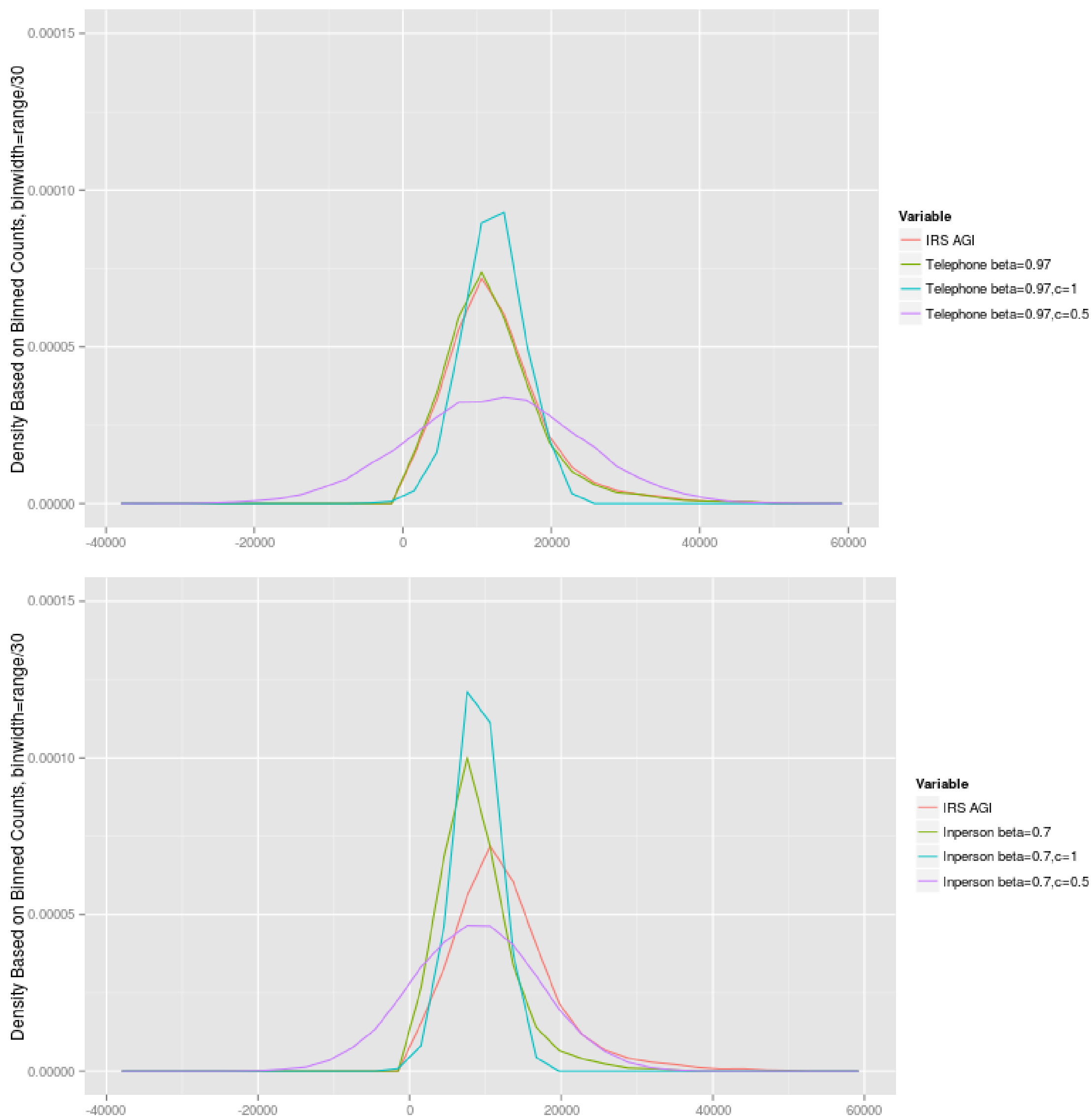


Figure 6.1 – Density Plots of Survey Response Data for Telephone and In-person Hypothetical Populations, and AGI data. beta=(0.97, 0.7) for telephone and in-person, respectively, in $Y_{pj}^{*} = \beta_p^{(AGI)} Y_j^{(AGI)}$, and $c \in (1, 0.5)$ in $\hat{Y}_j = X_j^{(Y)} \hat{\beta}^{(Y)} + e_j$, where

$e_j \overset{iid}{\sim} N(0, \sigma^2 / c)$.

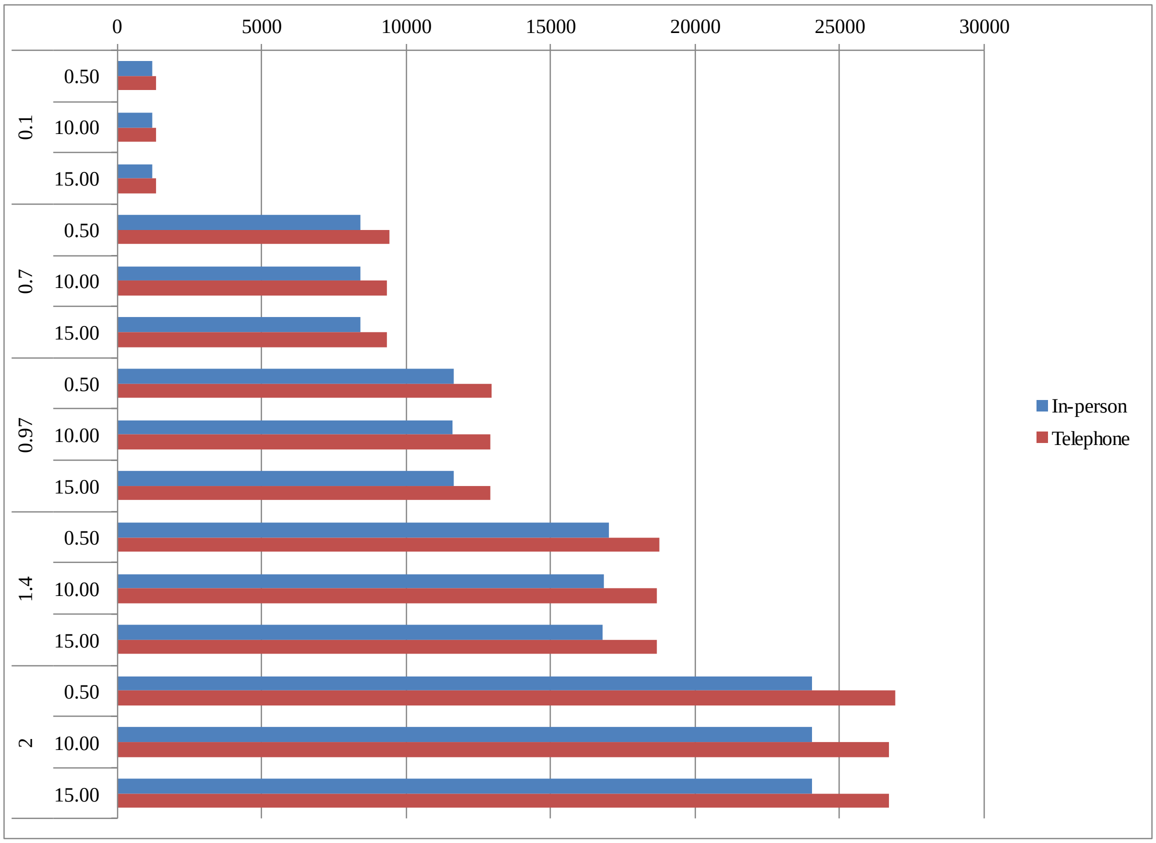


Figure 6.2 – Telephone and In-person Hypothetical Population Income Means for Varying Beta and Constant by Response Mode. The vertical axis shows combinations of $\beta_I^{(AGI)}$ and $c \in (0.5, 10, 15)$.

6.2.4. Estimation Methods

As shown in Figure 1.4, the proposed methods impute counterfactual data for the alternative response mode as if they had responded in the other mode. In detail, the estimation methods follow four steps:

1- Parameter estimation: Models were estimated to compute the parameters of beta coefficients for the mode choice and the response models. Two imputation models were implemented, ignorable and nonignorable mode choice models. As described in Section 3.3.3.1, the ignorable mode choice models include only the response regression models. As described in Section 3.3.3.2, the nonignorable mode choice models include both mode choice and response regression models.

2- Imputation: Using the parameter estimates from Step 1, telephone and in-person completed data vectors were created. These completed data vectors include both observed and imputed data values conditioned on the response mode, telephone and in-person. Five completed data vectors were computed.

3- Estimation: Using the completed data vectors, mode-specific means for total family income were computed. Since the mode-effects for in-person mode were introduced explicitly, differences between mode-specific means is not a research interest.

4- Combination of mode-specific means: Mode-specific means were combined using three methods: (1) simple average estimator (CM1), (2) inverse variance weighted estimator (CM2), and (3) inverse MSE weighted estimator (CM3). The details of the methods are presented in Section 3.3.4. Although as a part of the simulation study MSE weighted estimator was feasible, it is not feasible for most of the cases. The relative biases for these combination methods and the standard method were compared as presented in the next section.

6.2.5. Simulation Study Results

As shown in Figure 6.3 and Figure 6.4, the methods were compared in terms of relative bias as the true population mean was available. Figure 6.3 and Figure 6.4 show the relative biases for ignorable and nonignorable mode choice imputation models, respectively. The standard method in which modes are ignored is not model dependent. Consequently, the variations of constant c did not yield a change in the relative biases for a given beta. As shown in Figure 6.3, CM1, CM2, and CM3 outperformed the method where mode is ignored, CM4, when in-person incomes were overreported. Under the underreporting mode effect condition, CM2 consistently produced larger relative biases due to distribution of total family income. As shown in the scatterplot Figure 4.2, minimum of AGI is zero. As a result, the first step of the creation of the hypothetical population produced telephone and in-person responses with zero minimum. When the mode effects were adjusted for underreporting, this yielded a smaller variance for the in-person response distribution. As shown in Figure 6.1, negative values were allowed as a result of second step of the hypothetical population creation.

As Figure 6.3 shows when in-person incomes were overreported, the performance of CM2 varied considerably, depending on the size of the variance parameter *c*. The model fit influences the performance of the alternative methods under the ignorable model effect imputation model. For example, for severe overreporting in the in-person (beta=2) condition, relative biases for CM2 and CM3 increased when the model fit was poor (c=0.5).

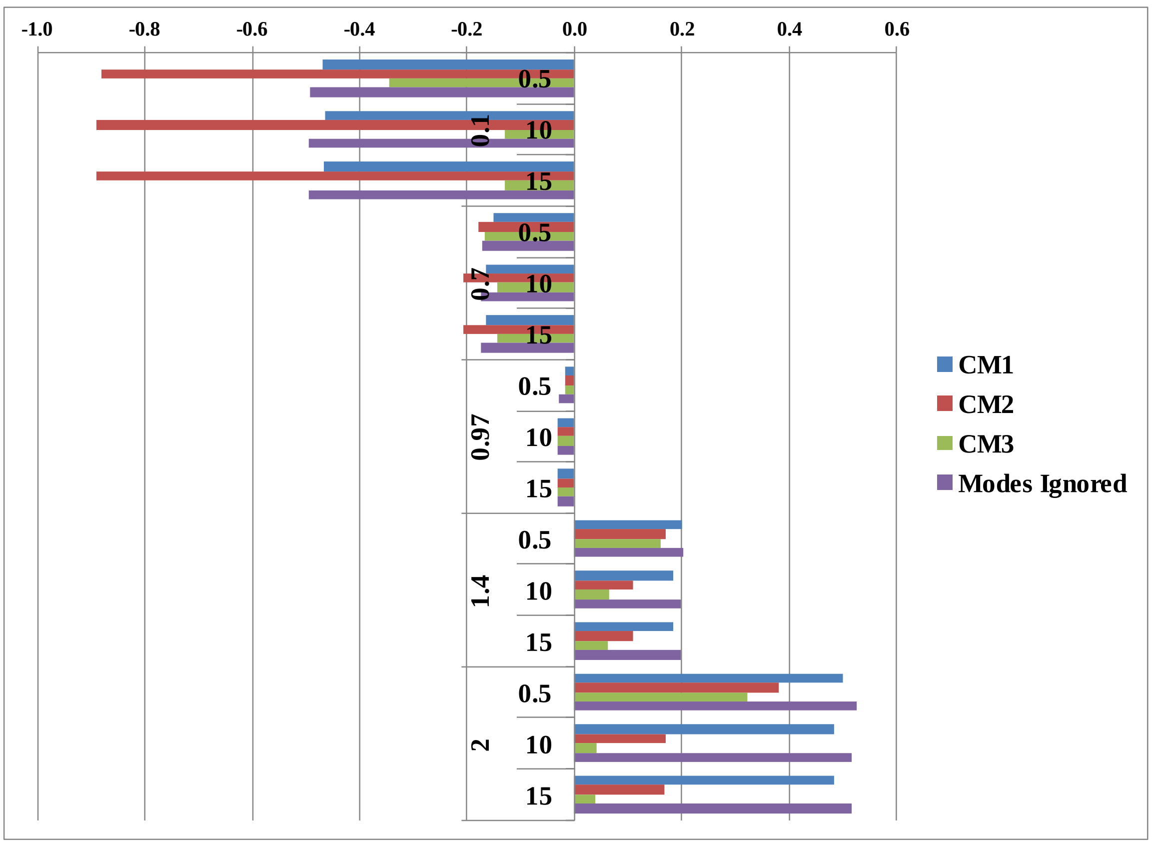


Figure 6.3 – Relative Biases for Alternative Inference Methods under Ignorable Mode Choice Imputation Model

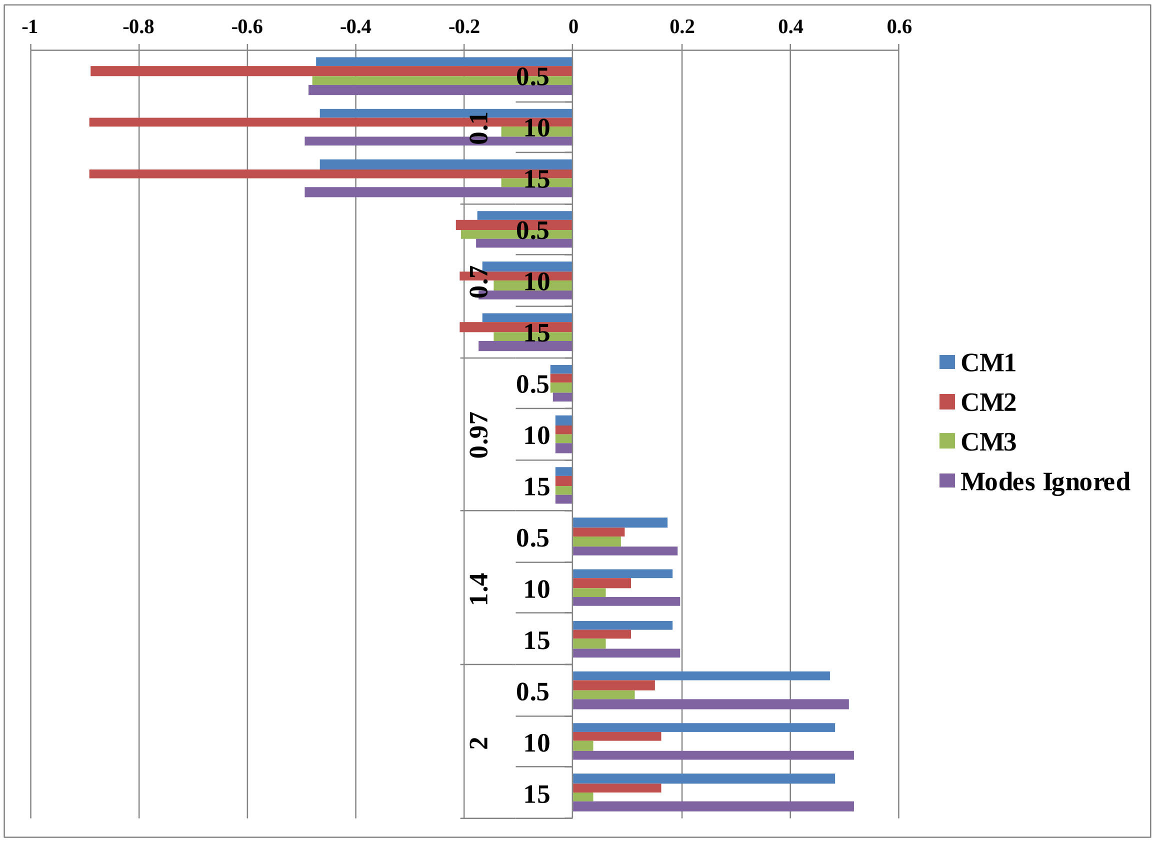


Figure 6.4 – Relative Biases for Alternative Inference Methods under Nonignorable Mode Choice Imputation Model

As Figure 6.4 shows results were similar under both ignorable and nonignorable mode effect imputation models. In Chapter 3, Section 3.3.3.1 and Section 3.3.3.2, the models that were used for ignorable and nonignorable mode choice are parameterized. For the ignorable mode choice model, a normal regression model for response was fit using the covariates listed in Table 6.1. For the nonignorable mode choice model a normal selection model, which included both mode choice (covariates listed in Table 6.2) and response models (covariates listed in Table 6.1), was fit. The ignorable mode choice model included a normal regression model for response as described in Section 4.3.3. In addition to the normal regression model, the normal selection model included a logistic regression model for the mode choice as described in Section 4.3.1. The variable selection was done separately for the mode choice and the response models. Section 4.1.1 describes the distributions of the covariates used in the models.

The nonignorable mode choice imputation model used assumed that total family income was normally distributed. Since the assumption agrees with the mechanism used to generate the data, this approach often yielded smaller relative biases than the model that ignored mode effects. The improvement was most apparent when the imputation model fits the best.

These results suggest that the standard method can be severely biased when the ignorable mode effects assumption is violated. Although the results for the alternative methods employed in this simulation are promising, assumptions in the imputation models play a crucial role. In practice, since there are no benchmarks evaluations of these models are a challenge. Alternatively, empirical comparison analyses can be conducted in the absence of benchmarks. The next section evaluates CPS March 2012 mean personal income and percent health insurance coverage as a case study in which benchmarks are not available.

6.3. Empirical Comparison Analysis: Application of the methods to CPS March 2012 Data

6.3.1. Motivation

In this section, the counter factual imputation-based estimation method described in Section 3.3 is applied to the CPS March 2012 data. The CPS data for March 2012 is described in detail and are compared to the 1973 CPS match data in Section 4.2.2 and Section 4.2.3. The CPS March 2012 data file was created by using the CPS March 2012 public data file. We included the cases with response mode data in the analysis. In the previously presented empirical and simulation studies, benchmark values were available and comparisons could be made accordingly. But this is not the usual case in real-life research conditions. In this section, we present results from a case study for which no benchmarks are available. Possible benchmark values could be provided by other survey data sources or administrative records. Other survey data sources such as American Community Survey (ACS) and National Health Interview (NHIS), which provide data on income and health insurance are mixed-mode surveys. Currently there are no available administrative records to compare the results against. Alternatively, we conducted an empirical comparison analysis on bootstrap replicates. Although this analysis cannot address the question with respect to the magnitude of the mode effects, significant differences would motivate further research on mode effects.

6.3.2. Creation of Sample Replicates

Since the unequal probability adjustment weights were not available separately, the sampling weights were recomputed at the state and month in sample (MIS) level to reflect the unequal probabilities of selection for the 2012 CPS observation. In the bootstrapping computations, units are defined as the Primary Sampling Unit (PSU) and state x MIS are considered to be the strata.

Although the replicate weights were computed and applied using the bootstrap function in R `survey` package, this method should incorporate a more comprehensive approach in the future work (Kennickell, 1991). The current method does not re-estimate the parameters of mode choice and response regression models, $(\hat{\beta}^{(R)}, \hat{\beta}^{(Y)}, \hat{\sigma}^2, \hat{\rho})$ which should be included in the comprehensive approach.

6.3.3. ANOVA for Repeated Measurements

The proposed estimator and the standard estimator that ignores modes use the same dataset to generate estimates. As a result these estimates cannot be evaluated under the independence assumption. The evaluations in estimate differences are conducted under an ANOVA for Repeated Measurements model:

$Y^*_{bCM} = \mu + \gamma_{CM} + \tau_b + e_{bCM}$, where:

CM=1,2 and standard method (CM4);

$b = 1, 2, ..., n_b$ indexes bootstrap replicates, n_b =200;

Assuming independence and common variance structure:

$$\tau_b \overset{iid}{\sim} N(0,\sigma_b^2)\,,\ \varepsilon_{bCM} \overset{iid}{\sim} N(0,\sigma_\varepsilon^2) \text{ and } \tau_b \perp \varepsilon_{bCM}$$

The evaluation will be made based on the null hypothesis of main effects of method using the F-test:

$$H_0:\ \gamma_{CM1} = \gamma_{CM2} = \gamma_{\text{ModesIgnored (or } CM4)} = 0\ .$$

6.3.4. Estimation Methods

Four steps have been applied to the CPS 2012 data to compute mean personal income and percent health insurance coverage:

1- Model Selection: Personal income and health insurance coverage models were fit separately for the mode choice and the response models. The modeling exercise and the final model structures are detailed in Section 4.4.

2- Parameter estimation: Models were estimated to compute the parameters of beta coefficients for the mode choice and the response models. Two imputation models were implemented, ignorable and nonignorable mode choice models. As described in Section 3.3.3.1, the ignorable mode choice models include only the response regression models. . As described in Section 3.3.3.2, the nonignorable mode choice models include both mode choice and response regression models.

3- Imputation: Using the parameter estimates from Step 2, telephone and in-person completed data vectors were created for a given bootstrap replicate sample. These completed data vectors include both observed and imputed data values conditioned on the response mode- telephone and in-person. Five completed data vectors were computed.

4- Estimation: Using the survey weights and the completed data vectors, mode-specific means for personal income and health insurance coverage were computed. These mode-specific means were compared against the means generated by the standard method using a repeated measurement ANOVA model to detect significant differences for possible mode effects.

5- Combination of mode-specific means: Mode-specific means were combined using two methods: (1) simple average estimator (CM1) , and (2) inverse variance weighted estimator (CM2) . These are comparable to the γ_{CM} used in the previous empirical and simulation studies but MSE weighted estimator cannot be used as there are no benchmarks available. These combined estimates were compared using a repeated measurement ANOVA model to detect significant differences for possible mode effects on the estimates.

6.3.5. Empirical Comparison Analysis Results

6.3.5.1. Personal Income

As Table 6.3 and Table 6.4 show, personal income as measured in the CPS 2012 March was sensitive to the methods applied. The differences between the methods and the mode-specific estimates were significant under both ignorable and nonignorable mode choice models. When the nonignorable mode choice model was used, in-person mean was lower and this was reflected in the combined estimates. Although these results cannot address the sources for differences, they may be considered as motivation for further investigation of mode effects.

Table 6.3 – F statistics for Repeated Measurement ANOVA for Personal Income (n_b = 200)

Method Comparison	Model	F-test	df	p-value
CM1, CM2, CM4	Ignorable mode choice	1366485	2/398	<.0001
In-person vs. Telephone	Ignorable mode choice	12618750	1/299	<.0001
CM1,CM2,CM4	Nonignorable mode choice	57115095	2/398	<.0001
In-person vs. Telephone	Nonignorable mode choice	72492226	1/299	<.0001

Table 6.4 – Mean and Standard Errors for Personal Income by Method and Imputation Model (n_b = 200)

Method	Imputation Model	
	Ignorable Mode Choice	Nonignorable Mode Choice
CM1	38,322.91 (178.32)	36,283.1 (174.39)
CM2	38,213.39 (174.24)	35,901.05 (168.15)
CM4 (Modes Ignored)	37,946.6 (267.47)	37,946.88 (267.47)
In-person	37,579.91 (183.37)	34,409.59 (175.82)
Telephone	39,065.92 (239.10)	38,158.23 (239.75)
In-person imputed	41,049.0.9	35,389.33
In-person observed	33,162.01	33,162.01
Telephone imputed	35,706.11	33,643.29
Telephone observed	41,704.23	41,704.23

Table 6.5 shows the constructed 95% confidence intervals for mean personal income. Figure 6.5 and Figure 6.6 visualize the 95% confidence intervals reported in Table 6.5.

Table 6.5 – 95% Confidence Intervals for Personal Income by Method and Imputation Model

Method	Imputation Model	
	Ignorable Mode Choice	Nonignorable Mode Choice
CM1	(37,973.4,38,672.42)	(35,942.3,36,625.9)
CM2	(37,871.88,38,554.9)	(35,571.77,36,230.91)
CM4 (Modes Ignored)	(37,422.64,38,471.12)	(37,422.64,38,471.12)
In-person	(37,220.5,37,939.32)	(34,065.02,34,754.24)
Telephone	(38,597.28,39,534.56)	(37,687.66,38,627.48)

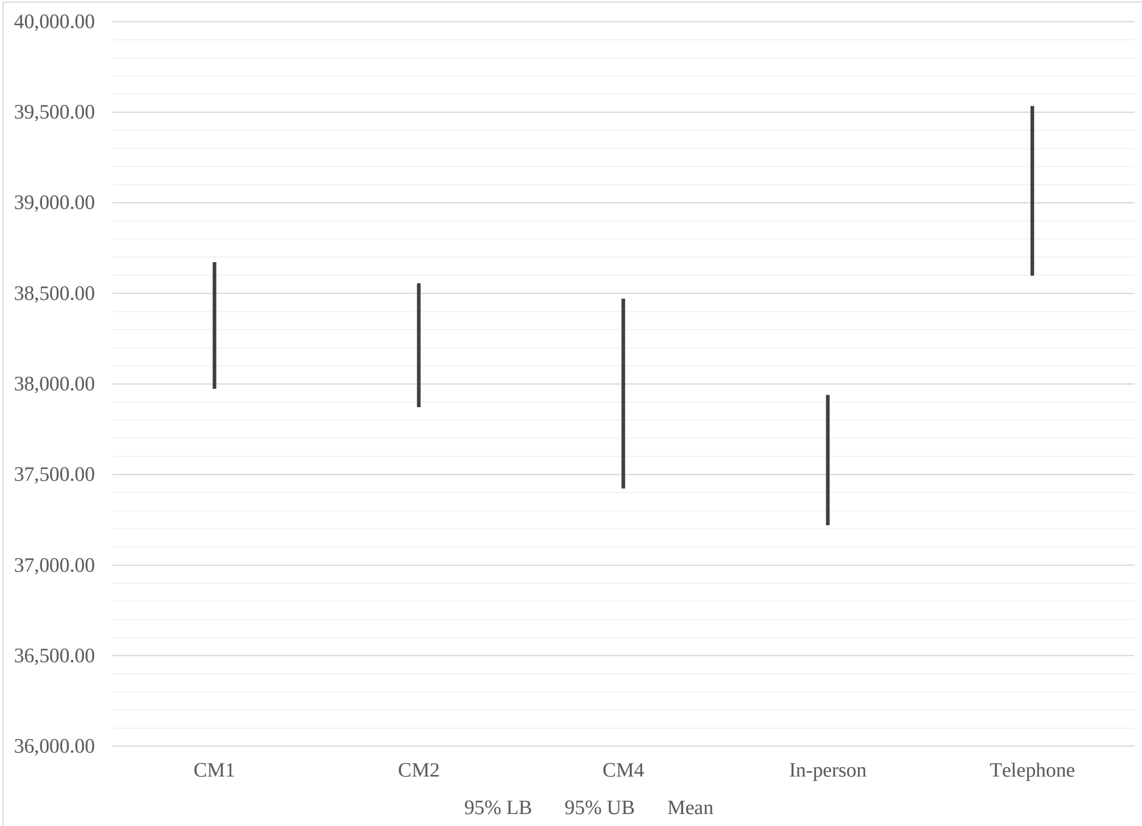


Figure 6.5 – 95% Confidence Intervals for Personal Income by Method under the Ignorable Mode Choice Imputation Model

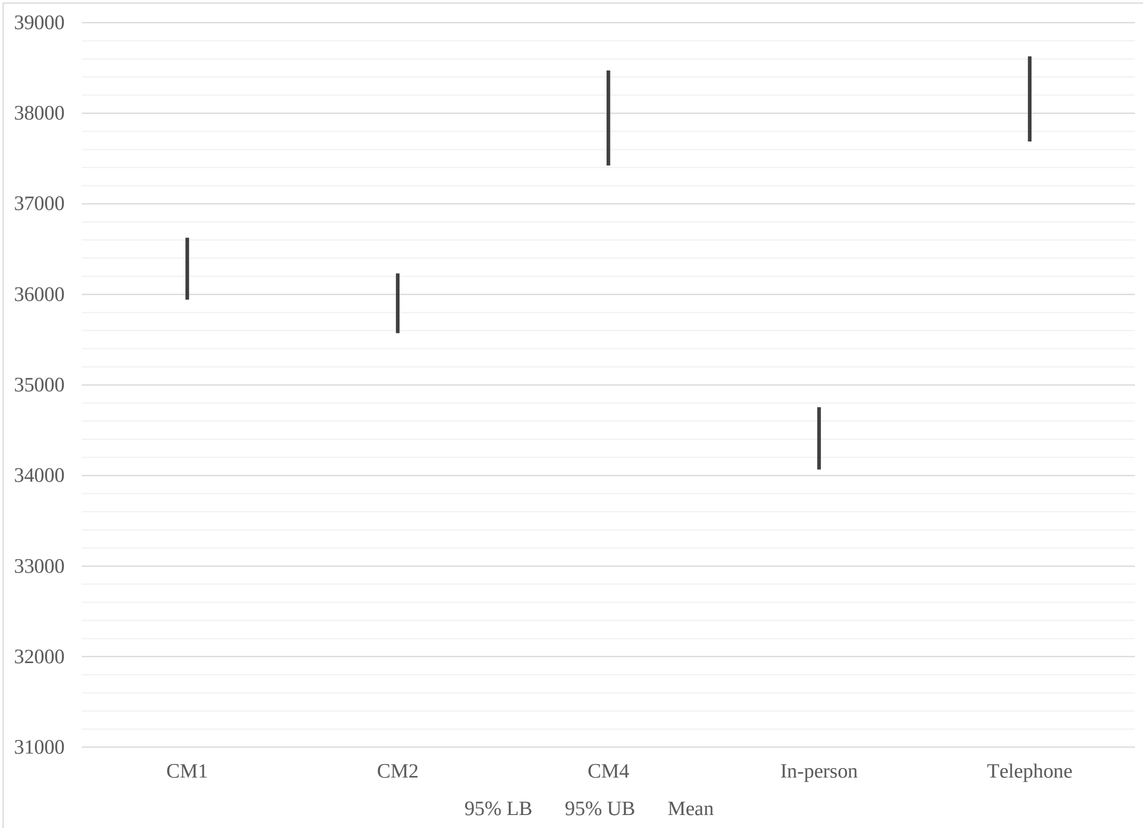


Figure 6.6 – 95% Confidence Intervals for Personal Income by Method under the Nonignorable Mode Choice Imputation Model

If these constructed 95% confidence intervals were used, the differences between the telephone and the in-person mode-specific means were significant. But the differences between the CM1 and CM2 methods were not significant under both imputation models. The standard combining method did not yield different results under the ignorable mode choice imputation model. On the other hand, both CM1 and CM2 yielded lower mean personal income under the nonignorable mode choice imputation model. Since the α is not optimal in either combination methods, these results further motivate the investigation on the α_{opt} and ignorability of mode choice.

6.3.5.2. Health Insurance Coverage

Table 6.6 and Table 6.7 report the F-tests for the comparisons of percent of health insurance coverage. Under both ignorable and nonignorable imputation models, a larger percent is observed for the telephone mode. But estimates are 14-21% less under nonignorable imputation model than they are under the ignorable model. This pattern is consistent across all four groups, 65+ (worker vs. nonworker) and <=65 (worker vs. nonworker). The future research will include simulation studies to investigate the source of the differences. Under the ignorable mode choice imputation model, as a result of lower percent in the in-person mode, the combined estimators are 1% less than the estimates generated by the standard method.

Table 6.6 – F statistics for Repeated Measurement ANOVA for Health Insurance Coverage (n_b = 200)

Method Comparison	Model	F-test	df	p-value
CM1, CM2, CM4 (Modes Ignored)	Ignorable mode choice	14630415	2/398	<.0001
In-person vs. Telephone	Ignorable mode choice	70021706	1/299	<.0001
CM1,CM2,CM4 (Modes Ignored)	Nonignorable mode choice	339586080	2/398	<.0001
In-person vs. Telephone	Nonignorable mode choice	54513353	1/199	<.0001

Table 6.7 – Means and Standard Errors for Health Insurance Coverage by Method and Imputation Model (n_b = 200)

Method	Imputation Model	
	Ignorable Mode Choice	Nonignorable Mode Choice
CM1	0.85 (0.001)	0.68 (0.001)
CM2	0.85 (0.001)	0.70 (0.001)
CM4 (Modes Ignored)	0.86 (0.002)	0.86 (0.002)
In-person	0.84 (0.002)	0.63 (0.002)
Telephone	0.86 (0.002)	0.72 (0.002)

6.4. Discussion

We presented results from one simulation and one empirical study in this chapter. The simulation study uses hypothetical populations that are based on observed associations but with controlled mode effect magnitudes. The relative biases were used as the evaluation criteria. Results are informative in two ways. First, under substantial mode effects, the standard method which ignores mode effects could yield large biases besides theoretically unknown bias properties. Second, modeling assumptions play a crucial role in the imputation estimation methods.

The empirical study used a subset of public CPS March 2012 data. This data set allowed us to implement the imputation method for a continuous, personal income, and a binary, health insurance coverage, variable. Empirical comparison analysis were conducted to detect possible differences as a result of mode effects. The differences under ignorable and nonignorable mode choice imputation models again emphasized the importance of modeling assumptions. Bootstrapping method requires a more comprehensive approach to estimate the variances. The current application of the bootstrapping method does not reestimate the beta coefficients for a given bootstrap replicate.

In these simulation and empirical studies, derived variables such as total family income and health insurance coverage currently ignore the associations between the individual variables that are used in constructing these variables. Following the imputation terminology they were actively imputed. Alternatively, the components of constructed variables could have been imputed by preserving the associations and then constructed variables could have been passively imputed. The properties of passive and active imputation techniques should have been investigated for imputing these derived variables as a part of the future research (van Buuren, 2007).

Chapter 7 Conclusions

7.1. Theoretical Framework

Chapters 2 and 3 of the dissertation outline the theoretical framework for mixed-mode survey inference. Theoretical work includes two layers. In the first layer, Lessler and Kalsbeek's (1992) statistical error model is extended to a mixed-mode survey context. Although the scope of this dissertation is to adjust for mode effects, this statistical error model covers the non-observational survey error components including coverage and sampling based on the Total Survey Error (TSE) framework. For the purpose of this dissertation, the scope is only single-frame mixed-mode surveys, but the statistical error model can be extended to a multi-frame design.

In the second layer, the statistical error model is extended to include a measurement error model which defines mode choice and mode specific systematic reporting errors explicitly. The extended statistical error model is instrumental in studying the bias properties of alternative mixed-mode survey statistical inference methods. The alternative statistical inference methods include the standard method that ignores mode effects, a proposed imputation method, and the existing methods that are used to unconfound mode choice and mode effects. The proposed imputation method of mixed-mode survey inference also attempts to unconfound mode choice and mode effects under some specific modeling assumptions by computing mode-specific estimates for the complete sample. As shown in Section 3.3.2.3, $\mathrm{E}_R\mathrm{E}_M\mathrm{E}_{IMP}\left[\bar{Y}_T^* - \bar{Y}_I^*\right] = \bar{B}_{UT} - \bar{B}_{UI}$. This implies that differences in mode specific estimates under the proposed method can be used to evaluate the average differences in mode effects. But the method does not allow us to estimate mode effects, $\bar{B}_{UT}$ and $\bar{B}_{UI}$ separately.

7.2. Bias of Alternative Estimators

As discussed in Chapter 3, under a measurement error model that includes systematic reporting errors, a mixed-mode survey mean estimator that ignores mode effects yields an expected estimation error that depends on the mode choice function and average systematic reporting errors. Under the conditions where there is no control over the mode choice function, i.e. in which mode sample units respond, the standard mean estimator may yield varying levels of biases in each realization of the measurement. In other words, the bias property of a mean estimator is not known. On the other hand, the estimation error for the mode-specific and combined estimator can be shown to have known bias properties under the imputation method. The imputation method estimators allow to evaluate and adjust for mode effects in a mixed-mode survey context. But it does not allow to estimate the mode specific biases.

7.3. Empirical and Simulation Studies

Empirical and simulation study results conformed to the expectations of the theoretical framework. The first empirical/simulation study on a special subset that included person level benchmarks allowed to compare the relative differences of alternative estimators. The conclusions for the ignorability of mode effects were different under different imputation model assumptions. Under the nonignorable mode choice imputation model, mode effects were concluded to be ignorable. The alternative combination methods did not reveal differences in terms of relative differences and they all outperformed the standard method that ignores mode effects.

The second study conducted simulations on hypothetical populations that were created based on the observed associations. As in the first study, the results conformed to the expectations. In addition, the results reiterated the importance of the modeling assumptions.

The third empirical study was conducted on a dataset where no benchmarks were available. Empirical comparison analyses for a continuous variable, personal income, and a binary variable, health insurance coverage, were conducted. Results pointed to the need for further research on the mode effects for these two variables. In particular, substantial differences in personal income and health insurance coverage draw curiosity whether these differences are due to mode choice or mode effects. The mode effects should be studied under the designs that Tourangeau et al. (2000b) describe. Also substantial differences between the ignorable and the nonignorable mode choice imputation model results for health insurance coverage require further understanding of the modeling technique for binary variable.

7.4. Limitations

The imputation models included household and householder covariates as collected by the mixed-mode survey. These covariates were assumed to be immune to mode effects. Another assumption was ignorable item nonresponse on covariate data. Although these seem to be generally plausible assumptions, these assumptions should be reviewed for a given survey procedure and a survey population. The covariates may also be augmented by the available auxiliary frame variables.

The imputation models studied here apply to a case where the data are collected from a fixed phase x mode sequence for the entire sample, which may not be the case. For example, in an ACS-like sequential mixed-mode survey a telephone follow-up for a mail phase nonrespondent may be a reminder to respond by mail. According to the fixed phase x mode sequence approach, this influence of telephone follow-up will not be captured. Although it is difficult to capture all possible patterns, these patterns could be informational in imputing data.

Also the imputation models do not incorporate the likelihood that an in-person report would be correlated with a telephone report for most persons. Future research should explore multivariate distribution modeling techniques to incorporate possible correlations between the responses in different modes in addition to studying mode effects in explicit experimental designs as described in Tourangeau et al. (2000b). Furthermore, these methods can be extended to panel surveys that switch from one mode to another. This will also provide a case in which between mode responses correlations are estimable. Given the mixed-mode survey data structures described in Table 1.1 and Table 1.2, between-mode response correlations are not estimable.

The results are shown to be sensitive to the modeling assumptions. Although a general measurement model is used in this dissertation, social and cognitive theories may be helpful when formulating models and assumptions.

7.5. Future Research

There are seven extensions of this dissertation research: (1) extend the method to model item nonresponse separately, (2) empirically evaluate α_{opt} for general cases, (3) extend the method to multi-frame designs, (4) empirically evaluate method for a sequential mixed-mode survey, (5) incorporate correlated random error variance/covariance structures for interviewer-administered modes,(6) empirically evaluate the model for multi-phase multi-mode designs and (7) conduct mode effect analyses on existing datasets that include randomized experimental data, such as the Institute for Social Research's Health Retirement Survey (HRS).

References

Aquilino, W. S. (1994). Interview Mode Effects in Surveys of Drug and Alcohol use: A Field Experiment. *Public Opinion Quarterly, 58*(2), 210–240.

Beland, Y., & St-Pierre, M. (2008). Mode effects in the Canadian Community Health Survey: A Comparison of CATI and CAPI. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. De Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in Telephone Survey Methodology* (pp. 297–314). Wiley.

Biemer, P., & Stokes, S. (1991). Approaches to the Modeling of Measurement Errors. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 487–516). Hoboken, New Jersey: Wiley-Interscience.

Boudreaux, M., Ziegenfuss, J. Y., Graven, P., Davern, M., & Blewett, L. A. (2011). Counting Uninsurance and Means-Tested Coverage in the American Community Survey: A Comparison to the Current Population Survey. *Health Services Research*, *46*(1), 210–231.

Brambilla, D. J., & McKinlay, S. M. (1987). A Comparison of Responses to Mailed Questionnaires and Telephone Interviews in a Mixed Mode Health Survey. *American Journal of Epidemiology*, *126*(5), 962–71.

Brick, J., & Lepkowski, J. M. (2008). Multiple Mode and Frame Telephone Surveys. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. Lavrakas, M.W. Link, & R. L. Sangster (Eds.), *Advances in Telephone Survey Methodology* (pp. 149–170). Wiley.

Broyden, C. G. (1969). A New Double-rank Minimization Algorithm. *Notices of the American Mathematical Society*, *16*, 670.

Broyden, C. G., Dennis Jr., J. E., & More, J. J. (1973). On the Local and Superlinear Convergence of Quasi-(Newton) Methods. *Journal of the Institute of Mathematics and its Applications*, *12*, 223–245.

Buelens, B., & Van den Brakel, J. (2011). Inference in Surveys with Sequential Mixed-Mode Data Collection. Retrieved from http://www.cbs.nl/NR/rdonlyres/C70ED9D5-6199-4E27-B37D-792A161D614D/0/2011x1021.pdf

Camillo, F., & D'Attoma, I. (2011). Tackling the Problem of Self Selection in the Integration of Different Data Collection Techniques. Retrieved from http://amsacta.unibo.it/3063/1/1.pdf

Chib, S., & Greenberg, E. (1998). Analysis of Multivariate Probit Models. *Biometrika*, *85*(2), 347–361.

Cobben, F., Schouten, B., & Bethlehem, J. (2006). A Model for Statistical Inference based on Mixed Mode Interviewing. In *European Conference on Quality in Survey Statistics*. Retrieved from http://www.statistics.gov.uk/events/q2006/downloads/T10_Cobben.doc

Couper, M. P. (2008). Web Surveys in a Mixed-Mode Environment. In *Designing Effective Web Surveys* (pp. 353–356).

Couper, M. P. (2011). The Future of Modes of Data Collection. *Public Opinion Quarterly*, *75*(5), 889–908.

David, M., Little, R. J. A., Samuhel, M. E., & Triest, R. K. (1986). Alternative Methods for CPS Income Imputation. *Journal of Americal Statistical Association*, *81*(393), 29–41.

Davis, M. E., & Alexander, Jr., C. H. (1997). The American Community Survey: The Census Bureau's Plan to Provide Timely 21st Century Data. *Missouri Library World, Spring*, Spring 1997.

De Leeuw, E. D. (1992). *Data Quality in Mail, Telephone and Face to Face Surveys.* Amsterdam: T. T. Publikaties. Retrieved from http://edithl.home.xs4all.nl/pubs/disseddl.pdf

De Leeuw, E. D. (2005). To Mix or not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics*, *21*(2), 233–255.

Diffendal, G. (2001). The Hard-To-Interview in the American Community Survey. In *Proceedings of the Annual Meeting of the American Statistical Association, Survey Research Methods Section*.

El Kasabi, M., Suzer-Gurtekin, Z. T., Lepkowski, J. M., Kim, U., Curtin, R., & McBee, R. (forthcoming). A comparison of Address-Based Sampling Mail Surveys and RDD Surveys for Measuring Consumer Attitudes. *International Journal of Market Research*.

Fletcher, R. (1970). A New Approach to Variable Metric Algorithms. *The Computer Journal*, *13*(3), 317–322.

Fowler, F. J., Gallagher, P. M., & Nederend, S. (1999). Comparing Telephone and Mail Responses to the CAHPS (TM) Survey Instrument. *Medical Care*, *37*(3), MS41–MS49.

Genz, A. (1992). Numerical Computation of Multivariate Normal Probabilities. *Journal of Computational and Graphical Statistics*, *1*(2), 141–149.

Genz, A. (1993). Comparison of Methods for the Computation of Multivariate Normal Probabilities. *Computing Science and Statistics*, *25*, 400–405.

Glynn, R. J., Laird, N. M., & Rubin, D. B. (1986). Selection Modeling versus Mixture Modeling with Nonignorable Nonresponse. In H. Wainer (Ed.), *Drawing Inferences from Self-selected Samples* (pp. 115–142). New York: Springer- Verlag.

Glynn, R. J., Laird, N. M., & Rubin, D. B. (1993). Multiple Imputation in Mixture Models for Nonignorable Nonresponse with Follow-ups. *Journal of the American Statistical Association*, *88*(423), 984–993.

Goldfarb, D. (1970). A Family of Variable-Metric Methods Derived by Variational Means. *Mathematics of Computation*, *24*(109), 23–26.

Gray, M., Blake, M., & Campanelli, P. (2011). The Use of Cognitive Interviewing Methods to Evaluate Mode Effects in Survey Questions. In *The Royal Statistical Society, Mode Effects on Survey Measurement*. Retrieved from http://surveynet.ac.uk/sdmi/events.asp

Greene, W. H. (2011). *Econometric Analysis (7th Edition)* (p. 1232). Prentice Hall.

Greenlees, J. S., Reece, W. S., & Zieschang, K. D. (1982). Imputation of Missing Values When the Probability of Response Depends on the Variable Being Imputed. *Journal of the American Statistical Association*, *77*(378), 251–261.

Groves, R. M. (1989). *Survey Errors and Survey Costs*. Hoboken, New Jersey: Wiley Series in Survey Methodology.

Groves, R. M. (1991). Measurement Error Across the Disciplines. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement Errors in Surveys* (pp. 1–25). Hoboken, New Jersey: Wiley-Interscience.

Groves, R. M. (1999). Survey Error Models and Cognitive Theories of Response Behavior. In M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, & R. Tourangeau (Eds.), *Cognition and Survey Research, New York: Wiley* (pp. 235–250). Wiley Series in Probability and Statistics.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. Wiley Series in Probability and Statistics Survey Methodology Section.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd ed.). Hoboken, New Jersey: Wiley Series in Survey Methodology.

Groves, R. M., & Kahn, R. (1979). *Surveys by Telephone: A National Comparison with Personal Interviews*. New York: Academic Press, Inc.

Hansen, M., & Hurwitz, W. (1946). The Problem of Nonresponse in Sample Surveys. *The American Statistician*, *41*(236), 517–529.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica: Journal of the Econometric Society*, *47*(1), 153–161.

Heerwegh, D. (2009). Mode Differences Between Face-to-Face and Web Surveys: An Experimental Investigation of Data Quality and Social Desirability Effects. *International Journal of Public Opinion Research*, *21*(1), 111–121.

Hochstim, J. R. (1967). A Critical Comparison of Three Strategies of Collecting Data from Households. *Journal of the American Statistical Association*, *62*(319), 976–989.

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus Face-to-face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias. *Public Opinion Quarterly*, *67*(1), 79.

Iacus, S. M., King, G., & Giuseppe Porro. (2012). Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis*, *20*(1), 1–24.

Jäckle, A., Roberts, C., & Lynn, P. (2010). Assessing the Effect of Data Collection Mode on Measurement. *International Statistical Review*, *78*(1), 3–20.

Kennickell, A. B. (1991). Imputation of the 1989 Survey of Consumer Finances: Stochastic Relaxation and Multiple Imputation. In *Joint Statistical Meetings - Section on Survey Research Methods* (pp. 1–10).

Kim, J. K. (2011). Parametric Fractional Imputation for Missing Data Analysis. *Biometrika*, *98*(1), 119–132.

Körmendi, E. (1988). The Quality of Income Information in Telephone and Face to Face surveys. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (Eds.), *Telephone Survey Methodology* (pp. 341–356). Wiley Series in Survey Methodology.

Lee, S., & Valliant, R. (2007). Weighting Telephone samples Using Propensity Scores. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. De Leeuw, L. Japec, P. Lavrakas, M.W. Link, & R. Sangster (Eds.), *Advances in Telephone Survey Methodology* (pp. 170–183). Hoboken, New Jersey: Wiley-Interscience.

Lessler, J. T., & Kalsbeek, W. D. (1992). *Nonsampling Error in Surveys*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics Section.

Little, R.J.A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). Hoboken, New Jersey: Wiley Series in Probability and Statistics.

Little, Roderick J.A. (1993). Pattern-mixture Models for Multivariate Incomplete Data. *Journal of the American Statistical Association*, *88*(421), 125–134.

Lugtig, P., Lensvelt-Mulders, G. J. L. M., Frerichs, R., & Greven, A. (2011). Estimating Nonresponse Bias and Mode Effects in a Mixed-mode Survey. *International Journal of Market Research*, *53*(5), 669–686. doi:10.2501/IJMR-53-5-669-686

Lynn, P., Hope, S., Jäckle, A., Campanelli, P., & Nicolaas, G. (2011). The Role of Visual and Aural Communication in Producing Mode Effects on Answers to Survey Questions. In *The Royal Statistical Society, Mode Effects on Survey Measurement*. Retrieved from http://surveynet.ac.uk/sdmi/events.asp

Martin, E., Hunter Childs, J., DeMaio, T., Hill, J., Reiser, C., Gerber, E.,Styles, K, & Dillman, D. (2007). Guidelines for Designing Questionnaires for Administration in Different Modes.

Moore, J. C., Stinson, L. L., & Welniak, Jr., E. J. (2000). Income Measurement Error in Surveys: A Review. *Journal of Official Statistics*, *16*(4), 331–361.

Newman, M. C. (1993). Regression Analyses of Log-transformed Data: Statistical Bias and Its Correction. *Environmental Toxicology and Chemistry*, *12*, 1129–1133.

Nicolaas, G., Campanelli, P., Hope, S., Jäckle, A., & Lynn, P. (2011). Is It a Good Idea to Optimise Question Format for Mode of Data Collection? In *The Royal Statistical Society*, *Mode Effects on Survey Measurement*. Retrieved from http://surveynet.ac.uk/sdmi/events.asp

Peytchev, A., Ridenhour, J., & Krotki, K. (2010). Differences Between RDD Telephone and ABS Mail Survey Design: Coverage, Unit Nonresponse, and Measurement Error. *Journal of Health Communication: International Perspectives*, *15*, 117–134.

Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, *70*(1), 41.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association*, 516–524.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Hoboken, New Jersey: Wiley Classics Library.

Schwarz, N., Strack, F., Hippler, H. J., & Bishop, G. (1991). The Impact of Administration Mode on Response Effects in Survey Measurement. *Applied Cognitive Psychology*, *5*(3), 193–212.

Searle, S., Speed, F., & Milliken, G. (1980). Population Marginal Means in the Linear Model: An Alternative to Least Squares Means. *The American Statistician*, *34*(4), 216–221.

Shanno, D. F., & Kettler, P. C. (1970). Optimal Conditioning of Quasi-Newton Methods. *Mathematics of Computation*, *24*(111), 657–664.

Soulakova, J., Hartman, A., Gibson, J., & Davis, W. (2009). The Impact of Survey and Response Modes on Current Smoking Prevalence Estimates Using TUS-CPS: 1992-2003. *Survey Research Methods*, *3*(3), 123–137. Retrieved from http://w4.ub.uni-konstanz.de/srm/article/viewDownloadInterstitial/2118/3491

Stolzenberg, R. M., & Relles, D. A. (1997). Tools for Intuition about Sample Selection Bias and Its Correction. *American Sociological Review*, *62*(3), 494–507. Retrieved from http://www.jstor.org/stable/2657318

The Census Bureau. (2012). Retrieved from http://www.census.gov/newsroom/releases/archives/american_community_survey_acs/cb12-247.html

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000a). Mode of Data Collection. In *The Psychology of Survey Response* (pp. 289–312). Cambridge, England: Cambridge University Press.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000b). Designs for Assessing the Accuracy of Survey Reports. In *The Psychology of Survey Response* (pp. 267–269). Cambridge University Press.

Tourangeau, R., Shapiro, G., Kearney, A., & Ernst, L. (1997). Who Lives Here? Survey Undercoverage and Household Roster Questions. *Journal of Official Statistics*, *13*(1), 1–18.

Tourangeau, R., & Smith, T. W. (1996). Asking Sensitive Questions the Impact of Data Collection Mode, Question Format, and Question Context. *Public Opinion Quarterly*, *60*, 275–304.

Tourangeau, R., & Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin*, *133*(5), 859–883.

Tucker, C., & Lepkowski, J. M. (2008). Telephone Survey Methods: Adapting to Change. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. De Leeuw, L. Japec, P. J. Lavrakas, M.W.Link, & R. L. Sangster (Eds.), *Advances in Telephone Survey Methodology* (pp. 3–26). Wiley-Interscience.

Van Buuren, S. (2007). Multiple Imputation of Discrete and Continuous Data by Fully Conditional Specification. *Statistical Methods in Medical Research*, *16*(3), 219–242.

Vannieuwenhuyze, J., Loosveldt, G., & Molenberghs, G. (2010). A Method for Evaluating Mode Effects in Mixed-mode Surveys. *Public Opinion Quarterly*, *74*(5), 1027–1045.

Vannieuwenhuyze, J., Loosveldt, G., & Molenberghs, G. (2012). A Method to Evaluate Mode Effects on the Mean and Variance of a Continuous Variable in Mixed-Mode Surveys. *International Statistical Review*, *80*(2), 306–322.

Voogt, R., & Saris, W. (2005). Mixed mode designs: Finding the balance between nonresponse bias and mode effects. *Journal of Official Statistics*, *21*(3), 367–387.

Williams, R. (2006). Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables. *The Stata Journal*, *6*(1), 58–82.