Understanding the effects of weighting on parameter estimation for multilevel models

Olusegun Afis Ismail, BSc, MSc.

Submitted for the degree of Doctor of Philosophy
Applied Social Statistics
Lancaster University

September 2020
Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

Olusegun Afis Ismail
Abstract

Creating a model from hierarchical data with missing data without addressing the missingness in the data may lead to poor parameter estimates. Hence the overall aim of this thesis is to investigate the effects of weighting adjustment methods on parameter estimates of weighted multilevel linear model to address unequal sampling selection and non-response in the continuous response variable. The significance of this thesis is that it seeks to fill the gaps in the existing body of work on complex survey data analysis on the identification of best weighting adjustment and conditions suitable to achieve reliable estimates from a weighted multilevel linear model.

To achieve the aim of this thesis, the Jamaica Survey of Living Conditions 2007 was used in a simulation study on a weighted multilevel linear model of the annual household expenditure of meals purchased away from home to investigate four weighting adjustments methods using two scenarios: Missing at Random (MAR) and Missing not at Random (MNAR) at 20%, 40% & 60% rate of missing respectively in the outcome variable. In order to fully investigate the effects of the weighting adjustments, the missingness in the outcome variable were tested as a function of a continuous, categorical and combination of both continuous and categorical variables. The simulation study was also extended to the scaling of the weights to identify changes in the effects on the parameter estimates.

The weighting adjustment with the most reliable estimates were applied to the modelling of reported income from the China Health and Nutrition Survey (CHNS 1989 & 2011) data as well as the household expenditure of meals purchased away from home in the Jamaica Survey of Living Conditions 2007 respectively. In the application of the weighting adjustments, different scenarios on different multilevel models were investigated to identify any changes in the parameter estimates.

The findings from the simulation study on the random effect multilevel model revealed that sampling weight adjusted for missing data using item non-response weight produced the most reliable estimates of the fixed component of the linear multilevel models at the 20% rate of missing.
Acknowledgements

I would like to express my sincere gratitude to Professor Peter Neal who came on board, supervised, and provided guidance that made this journey memorable. I would also like to thank Dr. Novie Younger-Coleman for her supervision, ideas and words of encouragement over the years. I would like to extend special gratitude to the late Professor Damon Berridge and Dr. Gillian Lancaster for their supervision at the beginning of my journey at Lancaster to the point of their departure. Special thanks to Dr. Kanchan Mukherjee for addressing all the challenges faced during my studies at Lancaster. Special thanks to my employer, the University of Technology, Jamaica for offering me a scholarship to embark on this Journey which has sustained me over the years. I would also like to thank my colleagues at the School of Mathematics and Statistics, Faculty of Science and Sports, University of Technology, Jamaica especially, Mr. Kevin Williams, Mrs. Ave McIntosh, Mr. David Finlay, Mr. Gustwell Weir, Mr. Leslie Young, and Mr. Errol Rowe for their understanding and support. I would also like to express thank you to Mr. Frederick Gordon of the Planning Institute of Jamaica (PIOJ), Mr. Douglas Forbes, former Director of Surveys at the Statistical Institute of Jamaica (STATINJA) and Ms Revene Evans for redrawing the maps.

Finally, special thanks to my family and friends for their love, support, encouragement and enduring with me when I was missing in action busy studying. And to Almighty God for giving me the strength to complete this journey.
Dedicated to
My Family
Contents

Acknowledgement ........................................................................................................... iv

List of Figures ................................................................................................................. ix

List of Tables .................................................................................................................... x

Chapter 1 Introduction ..................................................................................................... 1

1.1 Overview of Survey Data Sampling History ......................................................... 1
1.2 Census and Surveys in Jamaica ............................................................................. 2
1.3 Role of complex survey in development ............................................................... 2
1.4 Aims and objectives of the thesis ......................................................................... 4
1.5 Study research questions ....................................................................................... 4
1.6 Thesis structure ......................................................................................................... 5
Chapter 2 Literature Review 6

Chapter Outline
2.1 Understanding Complex Survey and Complex Survey Data…………………7
2.2 Nonresponse, Causes and Associated Problems .................................8
2.3 Unequal Probability of Sample Selection in Complex Sampling Design….10
2.4 Addressing Missing Data Problems................................................11
2.5 Context of Weighting Adjustment Approaches: Design or Model-Based… 11
2.6 Understanding Weighting Adjustments..........................................12
2.7 Multilevel Modelling : The number of Clusters and Intraclass Correlation ..13
2.8 Imputation as an Alternative to Weighting Adjustments......................14
2.9 Major Contribution of the Study....................................................16

Chapter 3 Exploratory Data Analysis 17

Chapter Outline
3.1 Brief History of Jamaica............................................................17
3.2 Jamaica Survey of Living Conditions (JSLC2007)..............................19
3.3 Introduction of Study Variables and Context -JSLC2007....................23
    3.3.1 Outcome Variable – Annual Expenditure of Meal Purchased Away
        from Home.............................................................................26
    3.3.2 Exploring the relationships between the covariates and the response
        variable..................................................................................27
    3.3.3 Exploring the missingness in the response variable due to the
        Covariates..............................................................................27
3.4 China Health and Nutrition Survey 1989 & 2011 .............................34
    3.4.1 CHNS 1989 & 2011 Selected Variables..................................35
    3.4.2 Exploring effects of the covariates on the response variable.........36
Chapter 4 Statistical Methodology

Chapter Outline
4.1 Multilevel Modeling ................................................................. 39
4.2 Weighting Adjustments for the Correction of Sample Selection............ 41
4.3 Logistic Regression Model and Item Non-Response Weight.................... 42
4.4 Parameter Estimation.................................................................. 43
4.4 Summary of the Statistical Methodology............................................ 50

Chapter 5 Simulation Study

Chapter Outline
5.1 Rationale and Purpose of the Simulation Study................................. 51
5.2 Specific Objectives of the Simulation Study...................................... 52
5.3 Review of the Survey Weights in the Simulation Study......................... 53
5.4 Simulation Plan............................................................................ 57
5.5 Missing at Random (MAR) and Missing not at Random (MNAR)............ 57
5.6 Simulation Statistical Assessment Tools.......................................... 58
5.7 Results of the Simulation Study..................................................... 60
5.8 Summary of the Simulation Study.................................................. 65

Chapter 6 Analyses of JSUCL2007 and CHNS 1989 & 2011

Chapter Outline
6.1 Aim and Scope of the Analysis..................................................... 80
6.2 Objectives and Rationale for the Analysis....................................... 81
6.3 Multilevel Models in the Analysis.................................................. 82
6.4 Estimating Response Probabilities for each Data Set and Item Non-Response Weight ................................................................. 88
Chapter 7 Discussion, Conclusion and Limitation 116

7.1 Discussion......................................................................................116
7.2 Conclusion.....................................................................................118
7.3 Limitation.......................................................................................118
7.3 Future Research...............................................................................118

Bibliography

List of Appendices
Appendix A
Appendix B
Appendix C
Appendix D
Appendix E
Appendix F
Appendix G
List of Figures

Figure 3.1: Spread of the tertiary institutions in Jamaica

Figure 3.2: Spread of the Enumeration Districts to illustrate the level of development by percentage of urban and rural areas

Figure 3.3: Sample Enumeration District Map-St. Andrew East Rural

Figure 3.4: Annual expenditure of meal purchased away from home

Figure 3.5: Annual Expenditure of meal purchased away from away from Home by age group of the household head

Figure 3.6: Annual Expenditure of meal purchased away from away from Home by the household size

Figure 3.7: Annual Expenditure of meal purchased away from away from Home by the sex of the household head

Figure 3.8: Annual Expenditure of meal purchased away from away from Home by the education of the household head

Figure 3.9: Annual Expenditure of meal purchased away from away from Home by the employment status of the household head

Figure 3.10: Annual Expenditure of meal purchased away from away from Home by the occupation of the household head

Figure 3.11: Pattern of missingness by predictor variables

Figure 3.12: Distribution of response variable before and after transformation using logarithm

Figure 6.1: Model II-A

Figure 6.2: Model II-B

Figure 6.3: Model III-A

Figure 6.4: Model IV-A

Figure 6.5: Model II-D
List of Tables

Table 3.1: Percentage spread of the Enumeration District (ED) by Parish for Census
Table 3.2: JSLC2007 Distribution of Surveyed Household and population by Parish for the Census 2001
Table 3.3: Jamaica inter-census population for 2012
Table 3.4: List of study variables from the JSLC2007
Table 3.5: List of study variables from the JSLC2007
Table 3.6: Descriptive statistics of the expenditures by the categories in the covariates (Age and Household Size)
Table 3.7 Descriptive statistics of the expenditure by the categories in the covariates (Gender, Level of Education, Employment Status and Occupation)
Table 3.8: Selected covariates and response variables from the CHNS 1989 and 2011
Table 3.9: Descriptive statistics of the individual income without imputation for each category in the covariate
Table 4.0: Summary of household propensity to respond
Table 5.1: Sampling and non-response weights
Table 5.2: Root mean square error (RMSE) for the ED selection weight (PS), Household selection weight (HS), Combination of ED and Household selection weight (PHS) and sampling weight adjusted for item non response weight(SWAM).
Table 5.3: Parameter Estimate of a weighted three level model incorporating different values of ICC under MAR Mechanism at 20% proportion of missing
Table 5.4: Parameter Estimate of a weighted three level model incorporating different values of ICC under MAR Mechanism at 20% proportion of missing
Table 5.5: Summary of the simulation results when the missing mechanism is MAR with missing rate of 20%, 40% and 60% for each variable for sex, age and multivariable covariates models
Table 5.6: Summary of the simulation results when the missing mechanism is MNAR with missing rate of 20%, 40% and 60% for each variable for sex, age and multivariable covariates models
Table 5.7: Parameter estimates from weighted models (scaled method A) from the missing mechanism is MAR with missing rate of 20%, 40% and 60% for each variable for sex, age and multivariable covariates models
Table 5.8: Parameter estimates from weighted models (scaled method A) from the missing mechanism is MNAR with missing rate of 20%, 40% and 60% for each variable for sex, age and multivariable covariates models
Table 5.9: Random Component –(E) estimates from weighted models from the missing mechanism is MAR with missing rate of 20%, 40% and
60% for each variable for sex, age and multivariable covariates models

Table 5.10: Parameter estimates from weighted multilevel models using sampling selection weight adjusted for missing data (SWAM) from the Missing At Random missing mechanism with missing rates of 20%, 40% and 60% for data simulated from age-only, sex-only, sex-age, multivariable covariate models

Table 5.11: Parameter estimates from weighted multilevel models using sampling selection weight adjusted for missing data (SWAM) from the Missing Not At Random missing mechanism with missing rates of 20%, 40% and 60% for data simulated from age-only, sex-only, sex-age, multivariable covariate models

Table 5.12: Parameter estimates from weighted multilevel models using scaling method A to scale sampling selection weight adjusted for missing data (SWAM) from the Missing At Random missing mechanism with missing rates of 20%, 40% and 60% for data simulated from age-only, sex-only, sex-age, multivariable covariate models

Table 5.13: Parameter estimates from weighted multilevel models using scaling method A sampling selection weight adjusted for missing data (SWAM) from the Missing Not At Random missing mechanism with missing rates of 20%, 40% and 60% for data simulated from age-only, sex-only, sex-age, multivariable covariate models

Table 5.14: Parameter estimates from weighted multilevel models using scaling method B to scale sampling selection weight adjusted for missing data (SWAM) from the Missing At Random missing mechanism with missing rates of 20%, 40% and 60% for data simulated from age-only, sex-only, sex-age, multivariable covariate models

Table 5.15: Parameter estimates from weighted multilevel models using scaling method B sampling selection weight adjusted for missing data (SWAM) from the Missing Not At Random missing mechanism with missing rates of 20%, 40% and 60% for data simulated from age-only, sex-only, sex-age, multivariable covariate models

Table 6.1: Distribution of the sampled and total ED per Parish
Table 6.2: Estimated ICC Values for the Levels in the Models-JSCLC2007
Table 6.3: Parameter estimates for Two-Level Random Coefficient Model using Item Non-response Weight Average at the Parish level
Table 6.3.1: Parameter Estimates for Two-Level Random Coefficient Model using Item Non-response Weight Average at the Parish level incorporating Parish as the fixed effect variable
Table 6.4: Parameter estimates for Two-Level Random Coefficient Model using Item Non-response Weight Average at the ED level
Table 6.5: Parameter estimates for Three-Level Random Coefficient Model using Item Non-response Weight Average at the ED level
Table 6.6: Profile of individuals in the Extracted Data from CHNS1989& 2011
Table 6.7: Parameter Estimates for Four-Level Model of Reported Income without Missing Data -Year 1989 for Subjects 16 to 65 years
Table 6.8: Parameter Estimates for Four-Level Model of Reported Income with Missing Data -Year 1989 for Subjects 16 to 65 years
Table 6.9: Parameter Estimates for Three-Level Model of Reported Income without Missing Data -Year 2011 for Subjects 16 to 65 years
Table 6.10: Parameter Estimates for Three-Level Model of Reported Income with Missing Data - Year 1989 for Subjects 16 to 65 Years
Table 6.11: Variance and Intraclass Estimates for the CHN Data 1989 and 2011
Abbreviations

CHNS  China Health and Nutrition Survey
CHNS1989  China Health and Nutrition Survey 1989
CHNS2011  China Health and Nutrition Survey 2011
ED  Enumeration District
JSLC  Jamaica Survey of Living Conditions
JSLC2007  Jamaica Survey of Living Conditions 2007
HS  Household selection weight
IR  Item non-response weight
MAR  Missing at Random
MNAR  Missing not at Random
PS  ED selection weight
PHS  ED and Household selection weight
STATIN  Statistical Institute of Jamaica
SWAM  Sampling Weight Adjusted for Item non response
Chapter 1

Introduction

1.1 Overview of Survey Data Sampling History

The methods of sampling in surveys have been found to be an important in information gathering about a population (Skinner & Wakefield, 2017). The theory behind sampling can be traced to the papers presented by the early statisticians such as Nayman (1934); Hansen and Hurwitz (1934); Sukhatme (1935); and Bowley (1936). These studies enable further development on sampling theory in the nineteen forties and early fifties (Cochran, 1942; Madow & Madow, 1944; Horvitz & Thompson, 1952). These developments on sampling methods continued to enhance the estimates from sample survey about the population. However, the data collected from surveys used to derive these estimates are not immune from non-response problems which are ubiquitous in every survey data collected either via simple or complex sampling techniques (Hansen & Hurwitz, 1946).
1.2 Census and Surveys in Jamaica

The Statistical Institute of Jamaica (STATIN), a Government Agency, has the responsibility for conducting census and household surveys in the island in order to collect national information and country statistics. Since its inception in 1946, the Agency has produced publications such as the consumer price indices reports, demographic statistics, employment and income levels, reports on the environmental statistics and state of the environment report, external trade reports, labour force statistics, national income and product, pocketbook of statistics, population census reports, production statistics, quarterly gross domestic product and statistical reviews (www.statinja.gov.jm).

In order to monitor the demographic dynamics and statistics of the population, STATIN conducts a census every ten years. The thirteenth census was conducted in 2001 and the fourteenth was conducted in 2011. Historically, the first census was conducted in 1844, and the Act of Parliament for Population and Housing Census was passed in July 30, 2001 (Population Census, 2001). The Population and Housing census provide information on the demographics and housing spread of the population. It also serves as the population reference for researchers in selecting samples from the population.

1.3 Role of Complex Survey in Policy Development

In developed and developing countries, complex surveys are non-uniform in the sampling design and are becoming a popular tool for data collection in policy formulation. In Jamaica, the Statistical Institute of Jamaica (STATINJA) has the mandate to collect data to inform policy development and the Institute is noted for a series of complex surveys. The Labour Force Survey (LFS) is one of such surveys. The Labour Force Survey publications usually informed the Government of Jamaica on employment and unemployment rate on quarterly and annual basis. This publication enables the Ministry of Labour to formulate policy that guide labour matters such as the issuing of work permit and job opportunities. The Jamaica Survey of Living Conditions (JSLC) is another complex survey that is
published by the Planning Institute of Jamaica while the Statistical Institute of Jamaica is responsible for the data collection. The JSLC is published on an annual basis to inform the Government of Jamaica on the prevalence of poverty. The JSLC enables the Ministry of Labour and Social Security to adequately plan for the beneficiaries on the Programme of Advancement through Health and Education (PATH). The STATINJA also collect data for the Household Expenditure Survey (HES) used to produce the consumer price index on an annual basis. In Jamaica, other institutions seldom conduct complex surveys for decision making, the Jamaica Life Style Survey is an example of such survey conducted by the Caribbean Health Institute within the University of the West Indies. The University of Technology, Jamaica also conduct Adult Population Survey for the Global Entrepreneurship Monitor to measure nascent entrepreneurs and other type of entrepreneurial activities among Jamaica adult population.

In North America especially the United States and Canada, there are a range of complex surveys that inform the policy of the Government in both Countries. In the United States, the centre for disease control is known for the National Health and Nutrition Examination Survey (NHANES) which enables the American Government to monitor the nutrition and health status of American citizens inclusive of sub populations with the USA.

The U.S. Census Bureau, the Government Agency responsible for census, household and business surveys and often use a multi-stage approach in the sample design. The Statistics Canada also conducts series of surveys that informs the policy of the Canadian Government. The Statistics Canada Labour Force Survey is another example of a complex survey.
1.4 Aim and objectives of the thesis

The aim of this thesis is to assess weighting adjustments from different scenarios in multilevel models of continuous outcomes with the ultimate view of identifying the least bias weighting adjustment for compensating for unequal sample selection and missingness due to non-response problems. The continuous variable in the complex survey data of interest is the annual household expenditure of meals purchased away from home in the JSCLC2007 and the reported income in the CHNS 1989 and 2011. This thesis seeks to achieve four objectives relating to weighting methodological issues especially when compensating for unequal sample selection and missing data due to non-response in the modelling of a continuous outcome.

The first objective is to develop a series of weighting methods that can address unequal sample selection in design of complex surveys and missing data due to non-response problems. The second objective is to conduct exploratory data analysis of the JSCLC2007 and CHNS 1989 and 2011 in preparation for building a series of unweighted and weighted linear multilevel models using continuous outcome variable to assess the prediction of the variable in the presence of missing data and unequal sample selection in the complex survey data. These models will also incorporate survey design information to enable comparisons of parameter estimates. The third objective is to conduct a simulation study on weighted multilevel models so as to identify the best weighting adjustment method by comparing parameter estimates to assess the bias from the weighting methods. The fourth objective seek to illustrate the application of the best weighting adjustment method to the analysis of the JSCLC2007 and CHNS 1989 and 2011.

The goal of this research is to investigate the strength and limitations of various weighting approaches for addressing unequal sample selection and missing data issues that arise because of non-response when modeling a continuous outcome variable in complex survey data.
1.5 Study research questions

The study objectives were developed to address the following questions:

1. Which weighting adjustments is the least bias when addressing unequal sample selection in complex survey design and missing data problem due to item nonresponse from survey respondents?

2. What are the conditions such as the proportion of missing and missing mechanism which the least bias estimates occur for the identified weighting adjustment method?

3. What are the effects of Intraclass Correlation Coefficient (ICC) on different parameter estimation methods?

4. Does the number of clusters in a level 1 or level 2 units of unweighted or weighted linear multilevel model affects the final predictors in the model?

1.6 Thesis structure

This thesis contains seven chapters. The Literature Review on complex survey and associated issues are provided in Chapter 2. Exploratory analysis of the Jamaica Survey of Living Conditions (JSLC2007) and China Health and Nutrition Survey data are provided in Chapter 3. Chapter 4 focuses on the statistical methodology used in the thesis. The methodology includes detailed explanation of the different weighting adjustments, parameter estimator for the linear mixed effects models and logistics regression model. The simulation study for the multilevel linear model and the incorporation of survey information are provided in Chapter 5. Chapter 6 application of weighting adjustment method to Jamaica Survey of Living Conditions and China Health and Nutrition Survey. Chapter 7 contains discussions, conclusions and future work.
Chapter 2

Literature Review

Chapter Outline

In Section 2.1, a review of what constitutes a complex survey and complex survey data is presented. In Section 2.2, the reviews of non-response, causes and associated problems are presented. In Sections 2.3 and 2.4, the reviews of unequal probability of sample selection in complex sampling design and how to address missing data are presented respectively. In Sections 2.5, 2.6, 2.7 and 2.8 the reviews of weighting adjustment approaches viz-a-vis design-based or model-based, type of weighting adjustments, multilevel modeling in relation to cluster size and intraclass correlation coefficient, and imputation are presented respectively. -And finally, Section 2.9 contains the major contribution of this thesis to the existing body of knowledge.
2.1 Understanding Complex Survey and Complex Survey Data

There is a momentum on the analysis of complex survey data in the literature. However, this momentum has led to the need for more clarifications on several issues surrounding complex sample design and the associated problems. Nearly every discipline now conducts studies involving data with the characteristics of a complex survey and in some instances, an analytical inference in the form of multilevel modelling.

In survey analysis related studies, the word “complex survey” often refers to a survey in which the sample designs involve some form of clustering and/or stratification in stages of the sample selection (Carle, 2009; Skinner and Wakefield, 2017). A complex survey may be characterized by: unequal probability in the sample selection, multi-stages in the sample selection, clustering, stratification, non-response issues and confidentiality of the survey respondents (Lee and Forthofer, 2006; Skinner, Holt and Smith, 1989; Hansen and Hurwitz, 1943; Pfeffermann, 2011; Heeringa et.al, 2017). These features have made surveys with complex sampling useful in obtaining a representative sample of “hidden and hard-to-reach population” as cited in Wirth and Tchetgen (2014).

The highlighted characteristics of a complex survey are also found in large surveys conducted by National Statistical Offices and some Private Institutions that requires a nation-wide survey. These features differentiate a complex survey from other type of surveys, especially when they are household-based, nation-wide and requires some form of nesting/hierarchy in the sampling structure. A typical example, is the Jamaica Survey of Living Conditions (JSLC) series, which is key data set of interest in this thesis. Similar surveys with complex sample designs include the National...
Health and Nutrition Examination Survey (NHANES, 2000) and National Health Interview Survey (NHIS) in the United States of America (USA). Other notable surveys with these features include the Family Expenditure Survey (FES) in the United Kingdom and the China Health and Nutrition Survey conducted in China by the University of North Carolina, USA. According to Johnson and Elliot (1998), conducting these types of surveys often support efficient use of resources and may be one of the reasons for the preference by the National Statistical Offices and some Private Institutions. The United Nations Guidelines on Household Surveys also supported the use of complex survey design (UN, 2005) for nationwide household-based surveys. In reference to the highlighted examples, the sampling design is usually in stages. Furthermore, the clustering of the target population may be followed by the stratification or vice versa as the case may be in some preferred approach during the sampling processes. For example, Kish (1965, pg. 359) demonstrates how complex survey designs can be constructed in practice using the United States. Dwellings reside within counties which can be used to construct complex surveys at a state or national level.

2.2 Nonresponse, Causes and Associated Problems

In survey sampling, Kish (1995, pg.532) defines non-response as many sources of failure to obtain response or measurement on selected sample. This definition is comparable to all the non-response definitions found in the literature (Bethlehem, 2009, pg. 1, Singer 2006). The non-response is viewed as a major problem facing researchers in social and medical sciences and Official Statistics (Durrant and Steele, 2013). This problem is worldwide that Durrant and Steele (2013) further cited the decreasing response rate for surveys from Martin and Matheson (1999) and also elsewhere according to Dette (1999) and Steeh et.al (2001). Singer (2006) cited Groves and Couper (1998) to affirmed that the aspect of
nonrespondent in a survey that are of greatest concern to survey methodologists are noncontact and refusal. Singer (2006) further trace the history on early research on non-response to polling in the 1930s and stated that first non-response article in JSTOR statistical journals was from 1945 and such article in the Public Opinion Quarterly was from 1948. Singer (2006) concluded by stating that “one of the important scientific challenges facing survey methodology at the beginning of this century is determining the circumstances under which nonresponse damages inferences to the target population”. Singer (2006) also highlighted the second challenge which “is the identification of methods to alter the estimation process in the face of nonresponse to improve the quality of the sample statistics”. In the attempt to study non-response rates and non-response bias in household survey, Groves (2006) expressed non-response bias as the departure of the expected value from an estimate from its true value.

The two categories of non-response in the literature are unit and item non-response (Lahor, 2009, pg.329; Rao, 2000 pg.216). Kish (1995, pg.532) outline causes which may lead to unit non-response such as not-at-home situation, out right refusal, incapacity or inability due to illness, not found, and lost schedules. In addition to the list from Kish (1995), non-response can also be due to poorly worded survey items or poor interviewing skills in conducting a survey (Scheaffer et.al, 1996). Thus, creating partial or total missing data and departure from the expected response rate that may ultimately lead to a bias in the sample estimate if not address. In this study, the Unit non-response is not contemplated for the fact that the profile information of the respondents in the sample frame and reasons for not participating in the survey of the respondents are not available to secondary users. These highlighted issues clearly support continued investigation on non-response especially on the context of rate of missing from non-response causes with regards to hierarchical modeling which will complement the highlighted reviews.
2.3 Unequal Probability of Sample Selection in Complex Sampling Design

In complex survey, the sampling processes are not uniform as stated earlier on the characteristics. Owing to the multistage in the sampling schemes, selected samples often have unequal probabilities of selection at some or all stages of the sampling process (Pfeffermann et.al,1998). These unequal selection probabilities at any stage of the sampling in complex survey may induce bias in standard estimation in the development of multilevel models (Pfeffermann et.al,1998). Despite the views from Pfeffermann et.al, (1998) not all studies relating to hierarchical modelling embrace the adjustment of estimates for unequal selection probabilities as demonstrated in Farrell and Ludwig (2008). For some analysts from other disciplines, the accuracy of estimates in the use of hierarchical maximum likelihood estimates compared with parameter estimation from classical single -level model serves as the motivation besides the adjustment for unequal sample selection in the data. This is highlighted the field of psychology by Rouder et.al (2005). The need for continuing investigation on the importance of incorporation of weight adjustments to address the unequal sample selection probabilities to improve parameters estimation from multilevel model cannot be overemphasize especially for non -survey methodologist. However, survey statisticians suggested that weighting adjustments should be used to address the bias in the estimates that is due to unequal selection probability in statistical inference (DuMouchel and Duncan, 1983; Pfeffermann,1993). This suggestion did not provide comparisons for each adjustment proposed in the sampling process to ascertain at which adjustment and what conditions in terms of missingness do these adjustments reduced or eliminate the bias in the estimate.
2.4 Addressing Missing Data Problems

To address the missing data problem in the post data collection phase, two methods are widely cited in the literature, this includes weighting adjustments and imputation (Little & Rubin, 2002). Furthermore, use of weighting adjustment methods for non-response problems in the context of weighted model estimates using complex survey data are well documented in Holt and Elliot (1991), Himelein (2013); Little and Rubin (2002); Korn and Graubard (1995); Pfefermann (1993); Kish and Frankel (1974); Hansen and Hurwitz (1946). However, findings from these studies usually provide comparisons between weighted and unweighted estimates in most instances along with the type of estimators but were found to be limited in terms of comparative analysis on the percentage of missing data for the weighted methods in which these methods are effective.

2.5 Context of Weighting Adjustment Approaches: Design or Model -Based

The context in which weighting adjustment methods are used in parameter estimation from complex survey data are yet to be fully explored and has led to a continuing debate on the types of estimation methodology, especially in hierarchical modelling in the context of design-based or model-based approaches (Snijders and Bosker, pg.217, 2012). According to Skinner and Wakefield (2017), the model-based approach is common in the main stream statistical curriculum while the design-based approach may require specialized training which continue to gain momentum but has been mostly limited to survey designers and practitioners. The focus of this thesis is the application of weighting adjustments in the context of hierarchical model which is purely a design-based approach. This approach will enable
further investigation on the suitable conditions for weighting adjustments in
the inference for multilevel models.

2.6 Understanding Weighting Adjustments

There are several approaches to weighting in the literature. These
approaches continue to generate further investigation depending on the
context and purpose of the adjustment of the estimates. Of all the weighting
adjustment approaches, the post stratification is the most popular in practice
for adjusting estimates for population total using known auxiliary variables
(Little, 1993; Lohr, 2010; Rao, 2000). The post stratification approach is
common practice by the National Statistical Offices (NSO) as a tool to post
stratify estimates to reflect population figures. In the literature, the post
stratification weight continues to be identified with different names and
under different context with slight modifications. In some studies, post
stratification is described in the context of “raking”, “weighting class
adjustment” and “calibration”, these terms have been used interchangeably
(Holt and Smith, 1979; Smith, 1991; Deville and Sarndal, 1992; Deville,
Sarndal & Sautory, 1993; Kalton and Flores-Cervantes, 2003; Little &
Rubin, 2002; Valliant, Dever & Kreuter, 2013). In most cases, the context
of use of post stratification adjustments is to address the unit non-response
(Heziza & Lesage, 2016). However, some difficulty may arise whenever the
census population figures or needed information for which the proportion
for homogenous group is to be created and this may create challenges to
researchers who may want to create post stratification weight.

Unlike the weighting adjustment by post stratification that requires an
external data in most cases, the weighting adjustment for the item non-
response often uses the available data in the adjustment and this also enable
the application of Little & Rubin (pg.19, 2002) missing mechanisms. The
common item non-response weight in the literature is usually created by the inverse response probability developed through a logistic model of response variable (Skinner and Arrigo, 2011, Bethlehem et.al, 2011, Barbara and Stephen, 2001). Understanding of the performance of the item non-response weight under different rate of missing cases is limited in the literature. Available studies often neglect to state what percentage of missing cases the weight is being applied.

From the highlighted weighting adjustments, the focus of this study shall be on the theory of weighting adjustments to address unequal sample selection due to differential responses in the sampling units and also item non-response. Similarly, these weights will be incorporated in the estimation of the parameter to highlight the effect of the weight in the estimation process.

2.7 Multilevel Modelling: The Number of Cluster and Intraclass Correlation

There are several issues surrounding the application of multilevel modelling in real life situations. One of the issues is the use of the multilevel modelling when the intraclass correlation is very small without proper definition of what is considered small. Another problem is the number of clusters in the model. These have led to numerous studies on these issues. Based on observations and also some literatures reviewed, the majority of these concerns are emanating from users of statistics especially the field of psychology and sociology to a greater extent compared to the perspective of survey statisticians. For example, McNeish and Stepleton (2014) conducted a review on previous studies on multilevel models with small number of clusters and also provided an illustrative simulation to demonstrate how a simple model becomes adversely affected by small clusters. McNeish and Stepleton (2014) cited Kreft (1996) who postulated a cluster size of 30, also cited Snijders and Bosker (2012) who recommended 20 clusters and Hox (1998, 2010) recommending 50 to 100 clusters. McNeish and Stepleton (2014) claimed that a review of 13 journals
from education, psychology and sociology did not meet these recommendations. The findings of McNeish and Stepleton (2014) provided minimum number of clusters varying from 5 to 50 for continuous outcomes with indication consideration for the adjustment of unequal probabilities of selections of the clusters in the model nor adjustment for missing data. Some researchers such as Huang(2018) highlighted the circumstances surrounding intraclass correlation and the justification of multilevel modelling. Huang(2018) cited Nezlek (2008, p.856); Hayes (2006); and Thomas and Heck (2001) as studies spreading the myth that ‘when the intraclass correlation is low, multilevel modelling is not needed’ . In this study one of the objectives is to investigate the changes in the predictor variables as the intraclass correlation coefficient changes to see if there is any truth in this myth. Similarly, another objective is to investigate the effect number clusters on the predictor variables to identify any possible changes as the cluster numbers varies in the level 1 and level 2 units to provide solutions and recommendations from the perspective survey statistician.

2.8 Imputation as an Alternative to Weighting Adjustments

The use of imputation technique to address missing data in statistical analysis has been around for a long time. The use of imputation method is also common in the literature as a tool for addressing item non-response problem, especially the hot and cold deck imputation, as well as the substitution of the mean and multiple imputation (Little and Smith,1987; Rubin and Little, 2002; Andridge and Little, 2010; Lohr, 2010; Little & Rubin, 2002; Rao, 2000; Buuren, 2012; Schafer & Schenker 2000).

According to Buuren (2012) the statistical method to replace a missing value were developed by Allan and Wishart (1930) this clearly suggest how long imputation method had been in existence. Buuren (2012) further posited that the methods were further enhanced by Yates (1933) and others
such as Dempster et.al, 1977; Madow et.al., 1983; Little and Rubin (1987). The work of Little and Rubin (1987) is now recognized as the leading procedure for imputation techniques. Little and Rubin (2002) described imputations as “means or draws from a predictive distribution of the missing values that requires a method of creating a predictive distribution for the imputation based on the observed data”.

There are two popular approaches to generating predictive distribution for imputation that is based on the observed data (Little and Rubin, 2002). This includes explicit and implicit modelling approaches. According to Little and Rubin (2002), the implicit methods include the hot deck imputation, substitution, cold deck imputation and a composite approach. On the contrary, the explicit method includes mean imputation, regression imputation and stochastic regression imputation.

According to Tabachnick and Fidell (2000), the mean substitution form of single imputation has one major disadvantage which is the decrease in variability between individuals’ responses and biases correlations. In some National Statistical Offices (NSO) such as the Census Bureau in the USA, the hot deck single imputation is still in practice (Reilly & Pepe, 1997) as cited in Patrician (2002), despite the disadvantages. The advances in the imputation method have led to multiple imputations over single imputation when the idea started. The multiple imputation method was developed to address the disadvantages in the single imputation method (Patrician, 2002). Buuren (2012) attributed the multiple imputation to Rubin’s work in the 1970s and continued to gain momentum over single imputation with the work of Allison (2000) and Schafer (1997, 2000).

Whereas the multiple imputation is very straightforward for addressing missing data in non-hierarchical data this is not the case for the clustered data because the software involved and need for advanced statistical knowledge (Grund, Ludtke and Robitzsch, 2016). Available software programmes are based on different procedures. According to Grund et.al (2018) the
procedures can be grouped into two broad paradigms: the joint modelling approach (JM) and fully conditional specification (FCS) approach. For the JM approach, a single model is specified for all variables with missing data and imputations are simultaneously generated from this model for all variables with missing data (Grund et. al, 2018). This is unlike the FCS approach where data are imputed separately for each variable with missing data while conditioning on some or all the other variables in the data set (Grund et.al, 2018). However, Grund et.al (2018) further stated despite the difference in approach, both approaches use similar covariance structures at the individual and group level and can be used interchangeably as demonstrated by Carpenter & Kenward (2013), p.220; Ludtke et.al, 2017; and Mistler (2015). These approaches can be found in software packages such as Mplus and in R packages jomo (Quartagno & Carpenter, 2016). In summary, Multiple Imputation can be explored using any of these approaches in the software packages to address missing data in multilevel data and also for comparison purposes with estimates from weight adjustment methods.

2.9 Major Contribution of the Study

The four main contributions of this study to the statistical analysis of weighting adjustment methodology. The first contribution is the identification of weighting adjustments that best address bias in estimates when there are missing data that is due to nonresponse specifically item non-response and unequal sampling selection when developing hierarchical model with a linear continuous response. The second contribution is on the investigation of the limitation of the weighting adjustments under varying percentages of missingness in the data. The third contribution is a comparative analysis of results from models from small number of clusters versus large number of clusters. The fourth and final contribution is the comparative analysis of results of hierarchical models with imputed data and weighted adjustments for item nonresponse to guide future developments in survey data analytical inference.
Chapter 3

Exploratory Data Analysis

Chapter Outline

This Chapter seeks to provide the exploratory analysis of selected variables from the Jamaica Survey of Living Conditions 2007 (JSLC2007) and China Health and Nutrition Surveys for the year 2009 and 2011 (CHNS2009 and CHNS2011) respectively. The exploratory analysis will serve as the preliminary understanding of the study variables prior to the multilevel modelling activities in Chapter 5. It will also provide an insight on the degree of relationships between these variables from each population. The first section in the Chapter provides a brief history of Jamaica then proceed to the second section on Jamaica Survey of Living Conditions (JSLC) and associated sub-sections. The follow-up sections are devoted to the exploratory analysis of the secondary data: CHNS2009 and CHNS2011. The Chapter is concluded with the summary of the exploratory analysis.

3.1 Brief History of Jamaica

The island of Jamaica is divided into fourteen (14) parishes. Each of the parishes has an administrative capital, except for Kingston and St. Andrew which has joint administrative capital, downtown Kingston and a Mayor responsible for both parishes. Further to the census of 2001 for which the survey of living conditions of 2007 sample was drawn, the parish of Kingston (100%-urban; Table 3.1) had the highest proportion of urban Enumeration Districts (EDs) among the parishes followed by the parish of St. Andrew (88% -urban; Table 3.1).

Some of the manufacturing companies and banks had their headquarters in the City of Kingston or the greater area known as the Kingston Metropolitan Area (KMA).
which includes the urban area of St. Andrew. As a result of the proximity of the parish of St. Catherine to both Kingston and St. Andrew, urbanization of the parish has seen rapid increase especially the City of Portmore. The population for the parish of St. Catherine grew by 7 percent while that of St. James grew by 5 percent between 2001 and 2011 census (2011 Census of Population and Housing – Table 3.1). However, there are also increases in the number of urban districts for some parishes and decrease in some instances. The contrast is evident in the distribution of the number of rural and urban districts boundaries for 2011 Census (Table 3.1). There are also two international airports which significantly contributed to the development of Kingston and the parish of St. James. There are more tertiary and non-tertiary institutions in the KMA compared to the other parishes as well as the opportunities for employment. The KMA comprises of the parishes of Kingston and St. Andrew. The parish of St. Andrew has two public universities and two private universities, which also accounts for 21.25 percent of the island's population while Kingston accounts for 3.30 percent (2011 Census of Population and Housing). These combined factors, places the population of the KMA to 24.55 percent. The lower prevalence of poverty among the household sampled in the Survey of Living Conditions for 2007 could likely be attributed to the employment and access to education in the KMA.

High levels of poverty are widespread in the parishes with high proportions of rural enumeration districts (Tables 3.1). These parishes also had less tertiary type institutions as illustrated in Figure 3.1 but are dominated by mineral extraction and agricultural type activities (Clarendon, St. Catherine, St. Thomas and Trelawny). It is noteworthy to state that the map in Figure 3.1 was drawn in ArcGIS 9.2 after collating the number of tertiary institutions within each Parish. The parishes of St. James, St. Ann and Westmoreland are renowned for tourism. In 2007, agriculture, forestry and fishing contributed 5 percent to the island gross domestic product when compared with the goods producing industry and mining and quarrying of natural resource, a leading contributor to the gross domestic product, 8.5 percent. The gains from the export of natural resources (for example bauxite) and brown sugar have suffered significantly from the continued rise in the cost of crude oil which is the main component in the energy requirement for the semi processing of
the bauxite and sugar cane. Figure 3.2 illustrates the island development by enumeration districts (ED).

Figure 3.1: Spread of the registered tertiary institutions in Jamaica by the University Council of Jamaica

Table 3.1: Percentage (%) spread of Enumeration Districts (ED) by Parish for Census 2001 & 2011

<table>
<thead>
<tr>
<th>Parish</th>
<th>Capital</th>
<th>2001 Urban %</th>
<th>2001 Rural %</th>
<th>2011 Urban %</th>
<th>2011 Rural %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kingston</td>
<td>KMA</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>St. Andrew</td>
<td>Morant-Bay</td>
<td>88</td>
<td>12</td>
<td>87</td>
<td>13</td>
</tr>
<tr>
<td>St. Thomas</td>
<td>Bull-Bay</td>
<td>25</td>
<td>75</td>
<td>28</td>
<td>72</td>
</tr>
<tr>
<td>Portland</td>
<td>Port-Maria</td>
<td>22</td>
<td>78</td>
<td>24</td>
<td>76</td>
</tr>
<tr>
<td>St. Mary</td>
<td>Saints-Ann’s Bay</td>
<td>24</td>
<td>76</td>
<td>27</td>
<td>73</td>
</tr>
<tr>
<td>St. Ann</td>
<td>Falmouth</td>
<td>16</td>
<td>84</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td>Trelawny</td>
<td>Montego-Bay</td>
<td>57</td>
<td>43</td>
<td>56</td>
<td>44</td>
</tr>
<tr>
<td>St. James</td>
<td>Negril</td>
<td>10</td>
<td>90</td>
<td>11</td>
<td>89</td>
</tr>
<tr>
<td>Hanover</td>
<td>Savana-far-mar</td>
<td>22</td>
<td>78</td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td>Westmoreland</td>
<td>Black River</td>
<td>13</td>
<td>87</td>
<td>15</td>
<td>85</td>
</tr>
<tr>
<td>St. Elizabeth</td>
<td>Mandeville</td>
<td>32</td>
<td>68</td>
<td>33</td>
<td>67</td>
</tr>
<tr>
<td>Manchester</td>
<td>May Pen</td>
<td>28</td>
<td>72</td>
<td>31</td>
<td>69</td>
</tr>
<tr>
<td>Clarendon</td>
<td>Spanish Town</td>
<td>70</td>
<td>30</td>
<td>73</td>
<td>27</td>
</tr>
</tbody>
</table>
3.2 Jamaica Survey of Living Conditions (JSLC)

The Government of Jamaica had introduced the Survey of Living Conditions in 1988 to assess the living conditions of the Jamaican population, in order to develop social programmes (JSLC, 1988). Since 1988, several surveys had been conducted to continuously understand the social problems facing the population which had led to the development of several initiatives by the Government. A notable initiative is the Programme of Advancement Through Health and Education (PATH). The Government of Jamaica has also established the Social Development Commission (SDC) and the Jamaica Social Investment Fund (JSIF).
This study uses the data from the survey which was conducted in 2007. It is worth mentioning that similar analyses could have been conducted for data from other years. However, the data for 2007 was released and approved for the study. The survey is a cross-sectional study with several outcomes on demographic characteristics, household consumption, health, education, housing, social welfare & related programmes, and issues pertaining to persons of prime working age in Jamaica. The survey is conducted on a yearly basis.

3.2.1 The Setting of the Study and Design

The Jamaica Survey of Living Conditions, 2007 (JSLC 2007) was conducted in April 2007 and focused mainly on households. The distribution of the households by parish in the JSLC2007 is presented in Table 3.2. The population distribution in Table 3.2 revealed that there are more people residing in the parish of St. Andrew and St. Catherine. The parishes have more households than other parishes and this is the reason for the higher selection of the enumeration districts in these parishes in Table 2.2.

(Source: http://statinja.gov.jm/Demo_SocialStats/populationbyparish.aspx)

Table 3.2: JSLC 2007 Distribution of Surveyed Household and population by Parish for the Census 2001

<table>
<thead>
<tr>
<th>Parish</th>
<th>Number of EDs (PSUs)</th>
<th>Number of Households</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Urban</td>
<td>Rural</td>
<td>Total</td>
</tr>
<tr>
<td>Kingston</td>
<td>6</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>St. Andrew</td>
<td>34</td>
<td>4</td>
<td>38</td>
</tr>
<tr>
<td>St. Thomas</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Portland</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>St. Mary</td>
<td>1</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>St. Ann</td>
<td>6</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>St. James</td>
<td>6</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Trelawny</td>
<td>5</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Hanover</td>
<td>0</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Westmoreland</td>
<td>2</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>St. Elizabeth</td>
<td>2</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>Manchester</td>
<td>5</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>Clarendon</td>
<td>4</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>St. Catherine</td>
<td>22</td>
<td>9</td>
<td>31</td>
</tr>
<tr>
<td>Total</td>
<td>97</td>
<td>73</td>
<td>170</td>
</tr>
</tbody>
</table>
3.2.2 Selection of Enumeration Districts

The JSLC 2007 survey is a complex survey that was designed in two stages. In the first stage, Primary Sampling Units (PSU) are selected from a list as illustrated in Table 3.2. In Jamaica, the Enumeration Districts (ED) are the Primary Sampling Units. The ED is an independent geographic unit sharing common boundaries with a contiguous ED. Except for the parish of Kingston, that has all the EDs being classified as urban, every other parish is comprised of both rural and urban EDs. The minimum number of dwellings for EDs in rural areas is 100, while those in urban areas is 150. The second stage of the survey involves the selection of the dwellings using a systematic sampling method with a fixed interval.

3.2.3 JSCLC2007 Data Collection

The survey utilizes a face-face to interviewing technique in the data collection. Each interviewer is assigned a specific number of EDs to survey. A map of each ED is provided to aid the interviewer’s location of the selected dwelling. (For example, the red boundary in Figure 3.2 illustrates a typical ED map used in the JSLCS and others conducted island wide by STATIN). The STATIN always assigned a team of supervisors to oversee the interviewers on a regular basis.

Figure 3.3: Sample Enumeration District Map - St. Andrew East Rural
3.2.4 JSLC Questionnaire

The JSLC Questionnaire consists of ten modules inclusive of the roaster. The modules are: Health ($A$), Education ($B$), Children items ($C$), Programme for Advancement Through Education and Health, PATH, ($D$), Daily Expenses ($E$), Food Expenses ($F$), Consumption Expenditure ($G$), Non-Consumption Expenditure ($H$), Housing and Related Expenses ($I$).

The food components are found in modules $E$, $F$ and $G$. In module $E$, the head of the household was asked if they had bought coal, kerosene oil, and wood. Module $E$ also includes questions regarding meals away from home such as meat, poultry, fish meals bought away from home, sandwiches, burgers, milk, other milk-based product, breakfast beverages, fruits, juices, vegetables and drinks in box, bottle or other packaging, and other form of meals such as soups, fresh or frozen beef, pork, fish, chicken purchase and vegetarian meals. Module $F$ consists of items of food that was received as gift. The non-food items are found in module $G$. These include personal care items, laundry items and kitchen supplies. The socio-demographic questions on household members and head are located in the roaster module of the questionnaire. Each household is assigned a serial number and this number is applied to all the members of the household.

3.2.5 JSCL2007 Data Source

The data for this study was obtained from the archives of the Sir Arthur Lewis Institute of Social and Economic Studies (SALISES), the University of the West Indies, Mona. SALISES is the archive for Survey data of JSLC Surveys conducted by the Statistical Institute of Jamaica (STATIN) for Academic Research in the West Indies. The permission for the use of the data was approved on behalf of the STATIN by the relevant officers at SALISES. The study data for the JSLC 2007 was extracted from two sources. The first source was the Annual Expenditure data of 1994 households. The second source was the roaster file containing the demographic information of 6,613 household members. The Annual Expenditure data consists of the overall household expenditure on food consumption and other
household expenditures. In addition, this Annual Expenditure is also a major component in the baseline analysis for determining the poverty status of a household in Jamaica by the Planning Institute of Jamaica (PIOJ).

3.3 Introduction of the JSJC 2007 Study Variables and Context

Study variables such as the household annual expenditure database include the annual household expenditure of meals purchased away from home (t meal), the size of the household, the constituency, the enumeration district, the parish and the serial number for the household were extracted from the JSJC 2007 Annual Expenditure file. The serial number for each household was used to select each household head and demographic information about the head of the household from the 6,613 household members. The selected demographic variables include of the household head include age, level of education, occupation, employment status and sex.

Tables 3.4 and 3.5 contains the list of the variables, the sample and relevant percentages for each category in the variable. In Table 3.4, of the 1994 households in the sample, 82.6% or 1648 cases provided data on the annual expenditure of meals purchased away from home suggesting that 17.4% cases were missing. Consequently, there are more females than males among the household heads and varying age groups with more middle age individuals in the sample. The household heads are engaged in nine occupational groups. The explorative analysis in the Tables 3.4 and 3.5, also shows that the household heads are employed but the majority did not provide their level of education.
Table 3.4: List of study variables from the JSCL2007

<table>
<thead>
<tr>
<th>Variables</th>
<th>n</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Expenditure of Meal Purchased Away from Home</td>
<td>1648</td>
<td>82.6</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 – 19</td>
<td>16</td>
<td>0.8</td>
</tr>
<tr>
<td>20 – 24</td>
<td>78</td>
<td>3.9</td>
</tr>
<tr>
<td>25 – 29</td>
<td>136</td>
<td>6.8</td>
</tr>
<tr>
<td>30 – 34</td>
<td>202</td>
<td>10.1</td>
</tr>
<tr>
<td>35 – 39</td>
<td>233</td>
<td>11.7</td>
</tr>
<tr>
<td>40 – 44</td>
<td>248</td>
<td>12.4</td>
</tr>
<tr>
<td>45 – 49</td>
<td>227</td>
<td>11.4</td>
</tr>
<tr>
<td>50 – 54</td>
<td>159</td>
<td>8.0</td>
</tr>
<tr>
<td>55 – 59</td>
<td>154</td>
<td>7.7</td>
</tr>
<tr>
<td>60 – 64</td>
<td>120</td>
<td>6.0</td>
</tr>
<tr>
<td>65 – 69</td>
<td>136</td>
<td>6.8</td>
</tr>
<tr>
<td>70 – 74</td>
<td>103</td>
<td>5.2</td>
</tr>
<tr>
<td>75 – 79</td>
<td>88</td>
<td>4.4</td>
</tr>
<tr>
<td>80 – 84</td>
<td>60</td>
<td>3.0</td>
</tr>
<tr>
<td>85+</td>
<td>34</td>
<td>1.7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1994</td>
<td>100</td>
</tr>
<tr>
<td><strong>Household Size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>451</td>
<td>22.6</td>
</tr>
<tr>
<td>2</td>
<td>374</td>
<td>18.8</td>
</tr>
<tr>
<td>3</td>
<td>352</td>
<td>17.7</td>
</tr>
<tr>
<td>4</td>
<td>269</td>
<td>13.5</td>
</tr>
<tr>
<td>5</td>
<td>229</td>
<td>11.5</td>
</tr>
<tr>
<td>6</td>
<td>142</td>
<td>7.1</td>
</tr>
<tr>
<td>7+</td>
<td>177</td>
<td>8.9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1994</td>
<td>100</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>1070</td>
<td>53.7</td>
</tr>
<tr>
<td>Females</td>
<td>924</td>
<td>46.3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1994</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 3.5: List of study variables from the JSLC2007

<table>
<thead>
<tr>
<th>Level of Education</th>
<th>n</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>High School Level (CXC, CAPE and GCE)</td>
<td>201</td>
<td>10.1</td>
</tr>
<tr>
<td>University Level (Degree and other)</td>
<td>120</td>
<td>6.0</td>
</tr>
<tr>
<td>None and not stated</td>
<td>1673</td>
<td>83.9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1994</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Employment Status</th>
<th>n</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employed</td>
<td>1683</td>
<td>84.4</td>
</tr>
<tr>
<td>Unemployed</td>
<td>152</td>
<td>7.62</td>
</tr>
<tr>
<td>Outside Employment</td>
<td>159</td>
<td>7.97</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1994</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Occupation</th>
<th>n</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skilled Agricultural and Fishery Worker</td>
<td>364</td>
<td>18.25</td>
</tr>
<tr>
<td>Non-classified</td>
<td>315</td>
<td>15.8</td>
</tr>
<tr>
<td>Craft and Related Trade Workers</td>
<td>276</td>
<td>13.84</td>
</tr>
<tr>
<td>Elementary Occupations</td>
<td>275</td>
<td>13.79</td>
</tr>
<tr>
<td>Service workers, Shop and Market Sales Workers</td>
<td>267</td>
<td>13.39</td>
</tr>
<tr>
<td>Professionals</td>
<td>108</td>
<td>5.42</td>
</tr>
<tr>
<td>Plant &amp; Machine Operators &amp; Assemblers and Elementary Occupations</td>
<td>106</td>
<td>5.32</td>
</tr>
<tr>
<td>Technicians and Associate Professionals</td>
<td>97</td>
<td>4.86</td>
</tr>
<tr>
<td>Legislators, Senior Officials, Managers and Professionals</td>
<td>94</td>
<td>4.71</td>
</tr>
<tr>
<td>Clerks</td>
<td>92</td>
<td>4.61</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1994</td>
<td>100</td>
</tr>
</tbody>
</table>

**3.3.1 Outcome Variable – Annual Expenditure of Meal Purchased Away from Home**

In the JSCLC 2007, a total of 1994 household records were sampled. Of the 1994 records only 1,648 had reported the annual expenditure of meal purchased away from home, whilst 346 cases expenditure were missing for different reasons. The highest expenditure of meals purchased away from home in the 1,648 cases was J$1,449,571 with three persons in the household. It is noteworthy that the lowest expenditure reported was J$1564.29 for a household with one person.
In order to understand the spread of the expenditure, Figure 3.3 illustrates the before and after the log of the expenditure was taken as part of the exploratory analysis.

Figure 3.4: Annual expenditure of meal purchased away from home

3.3.2 Exploring the relationships between the covariates and the response variable

The descriptive statistics in Tables 3.6 and 3.7 for the reported expenditure based on the categories in the covariates suggest that some of these variables may have the potential to predict the response variable. For example, a significant difference ($F=3.629$, $p=0.00$) was found for the amount of expenditure for the age group which the head of the household belongs. This is illustrated in the box plot (Figure 3.3.1). The household sizes also influence the expenditure as illustrated by the box plot in Figure 3.3.2. The differences in the expenditure due to the household size was found to be
significant (F=19.186, p=0.00); Unlike the household size, the gender of the household heads did not influence the household expenditure of meal purchased away from home (t=0.441, p=0.66). This is evident in Figure 3.3.3. The level of education was recoded into “educated” and “not stated” to effectively observe the effect of education on the household expenditure of the meal away from home. Figure 3.3.4 shows that there is significance difference (F=21.38, p=0.00) in the expenditure between the household that has an educated head and household which the head did not state level of education was merged with not stated because this category will not have an effect on the analysis. The three employment categories were recoded to two categories “employed” and “unemployed” to sufficiently understand its influence on the expenditure of meal purchased away from home. Employed household heads had higher expenditure than the unemployed as illustrated Figure 3.3.5 and the test statistic (t=8.270, p=0.00). The fact that some of the occupation categories are closely related and statistical power in the future modeling will be enhanced, the categories were collapsed to three. The first new category is the “Non-Office Related Work (NORW)” consisting of plant and machine operators and assemblers. Household heads who did not state the occupation remains in the “Not Classified” category while those in the Office Related Work were grouped into the third category. The Office related work (ORW) consist of Legislators, Senior Officials and Managers; Clerks; and Professionals. The influence of the recoded occupation vary was found to be significant (F=29.93, p=0.00) and illustrated in Figure 3.3.6.
Table 3.6: Descriptive statistics of the expenditure by the categories in the covariates (Age and Household Size)

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Sample (n)</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 – 19</td>
<td>16</td>
<td>10.92</td>
<td>0.81</td>
</tr>
<tr>
<td>20 – 24</td>
<td>78</td>
<td>10.99</td>
<td>0.97</td>
</tr>
<tr>
<td>25 – 29</td>
<td>136</td>
<td>11.12</td>
<td>0.98</td>
</tr>
<tr>
<td>30 – 34</td>
<td>202</td>
<td>11.17</td>
<td>0.82</td>
</tr>
<tr>
<td>35 – 39</td>
<td>233</td>
<td>11.13</td>
<td>0.92</td>
</tr>
<tr>
<td>40 – 44</td>
<td>248</td>
<td>11.03</td>
<td>0.91</td>
</tr>
<tr>
<td>45 – 49</td>
<td>227</td>
<td>11.01</td>
<td>0.99</td>
</tr>
<tr>
<td>50 – 54</td>
<td>159</td>
<td>11.02</td>
<td>0.92</td>
</tr>
<tr>
<td>55 – 59</td>
<td>154</td>
<td>10.94</td>
<td>0.96</td>
</tr>
<tr>
<td>60 – 64</td>
<td>120</td>
<td>10.76</td>
<td>0.96</td>
</tr>
<tr>
<td>65 – 69</td>
<td>136</td>
<td>10.75</td>
<td>1.04</td>
</tr>
<tr>
<td>70 – 74</td>
<td>103</td>
<td>10.54</td>
<td>1.04</td>
</tr>
<tr>
<td>75 – 79</td>
<td>88</td>
<td>10.69</td>
<td>1.06</td>
</tr>
<tr>
<td>80 – 84</td>
<td>60</td>
<td>10.72</td>
<td>1.03</td>
</tr>
<tr>
<td>85+</td>
<td>34</td>
<td>11.11</td>
<td>0.60</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1994</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **Household Size** | | |
|--------------------|------------|-------|------|
| 1                  | 316        | 10.64 | 1.10 |
| 2                  | 270        | 10.80 | 0.95 |
| 3                  | 294        | 10.92 | 0.92 |
| 4                  | 254        | 11.10 | 0.86 |
| 5                  | 216        | 11.16 | 0.84 |
| 6                  | 135        | 11.32 | 0.84 |
| 7+                 | 163        | 11.38 | 0.80 |
| **Total**          | **1648**   | 10.98 | 0.96 |
Table 3.7: Descriptive statistics of the expenditure by the categories in the Covariates (Gender, Level of Education, Employment Status and Occupation)

<table>
<thead>
<tr>
<th>Gender</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>1070</td>
<td>10.99</td>
<td>0.96</td>
</tr>
<tr>
<td>Females</td>
<td>924</td>
<td>10.97</td>
<td>0.96</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1994</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Level of Education       |      |      |     |
| High School Level (CXC, CAPE and GCE) | 201  | 10.93| 0.97|
| University Level (Degree and other) | 120  | 11.46| 0.83|
| None and not stated      | 1673 | 11.41| 0.62|
| **Total**               | **1994** |      |     |

| Employment Status        |      |      |     |
| Employed                 | 1683 | 11.05| 0.93|
| Unemployed               | 152  | 10.52| 1.04|
| Outside Employment       | 159  | 10.39| 0.98|
| **Total**               | **1994** |      |     |

| Occupation Categories    |      |      |     |
| Skilled Agricultural and Fishery Worker | 364  | 10.42| 0.98|
| Non-classified           | 315  | 11.14| 0.87|
| Craft and Related Trade Workers | 276  | 11.17| 0.83|
| Elementary Occupations   | 275  | 11.52| 0.84|
| Service workers, Shop and Market Sales Workers | 267  | 10.88| 0.92|
| Professionals            | 108  | 11.19| 0.86|
| Plant & Machine Operators & Assemblers and Elementary Occupations | 106  | 11.36| 0.73|
| Technicians and Associate Professionals | 97   | 11.33| 0.82|
| Legislators, Senior Officials, Managers and Professionals | 94   | 11.35| 0.86|
| Clerks                   | 92   | 10.50| 0.97|
| **Total**               | **1994** |      |     |
Figure 3.5: Annual expenditure of meal purchased away from home by age group of the household head

Figure 3.6: Annual expenditure of meal purchased away from home by the household size
Figure 3.7: Annual expenditure of meal purchased away from home by the sex of the household head

![Different boxplots for each SEX](image)

Figure 3.8: Annual expenditure of meal purchased away from home by the education of the household head

![Different boxplots for each EDUCATION](image)
Figure 3.9: Annual expenditure of meal purchased away from home by the employment status of the household head.

Figure 3.10: Annual expenditure of meal purchased away from home by the occupation of the household head.
3.3.3 Exploring the missingness in response variable due to the covariates

In this section, the focus is on the missingness in the response variable due to each covariate using Figure 3.4 to illustrates the spread of the missing data in the response variable based on the covariates. The spread of the missingness among the age group revealed that as the age group increases the missingness in the response variable also increases especially as the household head approaches age of retirement (65 – 69 years and above) when compared to the household that is headed by 55-60yrs and less. This suggest that the age covariate has the potential to be a predictor for a missing variable.

The spread of missingness among the categories in gender variable is almost the same for male and female headed households while the spread in the other variables as illustrated in Figure 3.8 were not as defined in the pattern as compared with the age covariate.

Figure 3.11: Pattern of missingness by predictor variables
In this study, the goal is to investigate the effects of weighting approaches on parameter estimates from multilevel model with emphasis on missing data in the continuous outcome variable. The China Health and Nutrition Survey data have some unique features that facilitates this type of analysis and will serves as the secondary analysis to apply the results to a secondary data. The unique features are: (1) it is a publicly available dataset that scholars from across the world can access; (2) the survey used a multistage, random cluster design in the selection of a stratified probability sample (Zhang et. al, 2014, Popkin et.al, 2009).
3. 4.1 CHNS 1989 and 2011 Selected Variables

For the purpose of comparison, two different years were selected for analysis, the year 1989 and 2011 datasets were selected with the focus on the dynamics on the reported income of the individuals in the survey as the response variable. In addition, there are two variables on reported income of the individuals in the 1989 and 2011 dataset that were very useful in the analysis (reported income with imputation and without imputation). The survey covers provinces such as Liaoning, Jiangsu, Shandong, Henan, Hubei, Hunan, Guangxi, Guizhou, while in 2011, additional provinces were added such as Beijing, Heilongjiang, Shanghai and Chongqing in 2011 survey. The distribution of the response variables before and after the logarithm of the income variables is illustrated in Figure 3.5 demonstrating the attempts to meet the normal distribution assumption.

In both CHNS1989 and 2011, a total of six similar variables were selected in each data set. The variables consist of reported income with missing data, reported income without missing data, age, gender, marital status, and years of schooling. In the 1989 dataset, 3,052 individuals in the age range of 16yrs to 64yrs were selected from the sample compared with 2,109 sample for 2011 as stated in Table 3.8.

Table 3.8: Selected covariates and response variables from the CHNS 1989 and 2011

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1606</td>
<td>52.6</td>
<td>984</td>
<td>46.7</td>
</tr>
<tr>
<td>Female</td>
<td>1446</td>
<td>47.4</td>
<td>1125</td>
<td>53.3</td>
</tr>
<tr>
<td>Total</td>
<td>3052</td>
<td>100</td>
<td>2109</td>
<td>100</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>2662</td>
<td>87.2</td>
<td>1955</td>
<td>92.7</td>
</tr>
<tr>
<td>Not married</td>
<td>390</td>
<td>12.8</td>
<td>154</td>
<td>7.3</td>
</tr>
<tr>
<td>Total</td>
<td>3052</td>
<td>100</td>
<td>2109</td>
<td>100</td>
</tr>
<tr>
<td>Age</td>
<td>35.03</td>
<td>10.14</td>
<td>45.95</td>
<td>11.94</td>
</tr>
<tr>
<td>Years of schooling</td>
<td>15.31</td>
<td>8.81</td>
<td>24.01</td>
<td>7.6</td>
</tr>
<tr>
<td>Income without imputation</td>
<td>1371.93</td>
<td>1652.96</td>
<td>28733.88</td>
<td>32007</td>
</tr>
<tr>
<td>Income with imputation</td>
<td>1376.01</td>
<td>1655.09</td>
<td>29383.51</td>
<td>32431.9</td>
</tr>
</tbody>
</table>
3. 4.2 Exploring effects of the covariates on the response variable - CHNS

The descriptive statistics in Table 3.9 revealed differences in the reported income of the individuals for the selected covariates. The years of schooling has considerable differences in the income of the individuals with 16 years or more than the individuals with less years of education. This affirms the widely view about years of education and income. The income spread by among the age group categories also support the years of schooling effect on the income variable. The high-income earners are in the age group of 30 to 44 years with the expectation that these individuals will have 16 years of more of schooling. The fact that some differences exist in the reported income due to the effect of the covariates, it is expected that these variables are potential predictor in the follow up analysis in Chapter 6.
Table 3.9: Descriptive Statistics for the individual income without imputation for each category in the covariate.

<table>
<thead>
<tr>
<th>Variable</th>
<th>1989</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>1606</td>
<td>6.78</td>
</tr>
<tr>
<td>Females</td>
<td>1446</td>
<td>6.67</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>2662</td>
<td>6.74</td>
</tr>
<tr>
<td>Not married</td>
<td>390</td>
<td>6.62</td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-19</td>
<td>146</td>
<td>6.13</td>
</tr>
<tr>
<td>20 - 24</td>
<td>321</td>
<td>6.45</td>
</tr>
<tr>
<td>25 - 29</td>
<td>510</td>
<td>6.68</td>
</tr>
<tr>
<td>30 - 34</td>
<td>590</td>
<td>6.84</td>
</tr>
<tr>
<td>35 - 39</td>
<td>520</td>
<td>6.90</td>
</tr>
<tr>
<td>40 - 44</td>
<td>409</td>
<td>6.91</td>
</tr>
<tr>
<td>45 - 49</td>
<td>265</td>
<td>6.79</td>
</tr>
<tr>
<td>50 - 54</td>
<td>169</td>
<td>6.70</td>
</tr>
<tr>
<td>55 - 59</td>
<td>84</td>
<td>6.41</td>
</tr>
<tr>
<td>60 - 64</td>
<td>38</td>
<td>6.49</td>
</tr>
<tr>
<td>65 -69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of schooling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-15</td>
<td>840</td>
<td>6.68</td>
</tr>
<tr>
<td>16+</td>
<td>1599</td>
<td>6.82</td>
</tr>
</tbody>
</table>
Chapter 4

Statistical Methodology

Chapter Outline

This Chapter provides a broad understanding of the Statistical Methodology used in this study. The Statistical Methodology includes concepts such as the theory of parameter estimates for unweighted and weighted linear mixed-effects models, weighting adjustments and associated theory to correct for unequal probability of sample selection and missing data bias in estimates. To fully appreciate these concepts, the first section starts with an elaborate development of multilevel models framework which can be extended to any level. The second section explains how weights are to be incorporated in the estimation of parameters for weighted multilevel models. The Chapter finishes with a summary section on how the developed weights and methods of parameter estimates would be utilized in Chapters 5 and 6, respectively.

4.1 Multilevel Model

One of the motivations for this research, is the investigation of the effects of weighting adjustments on parameter estimates on linear models consisting of fixed and random effects of continuous outcome with missing data. Despite the fact that multilevel data analysis continues to evolve with novel research for example the work of Bryan and Jenkins (2016), West et.al (2015), Steele and Durrant (2011), Farrell and Ludwig (2008), Zhou et.al (2007), there are more opportunity for continued research in this field, especially the effect of different weighting adjustments on estimates.
a simple hierarchical model to illustrate the principles with further generalization to any hierarchy in the data structure.

In the JSCLC2007, \( J \) denote the Parishes as the highest Unit for \( j = 1, 2, 3, \ldots, m \) and within Parish \( j \), \( k^{th} \) denote the Enumeration District (ED) from \( 1, \ldots, k_j \). For the CHNS 1989 & 2011 data structure, the notation is repeated by denoting \( J \) Provinces as the highest Unit. Within Province \( j \), similarly \( k^{th} \) denote the County /City by \( (j, k) \) within the province and within each Council /City \( l^{th} \) denote the Community by \( (j, k, l) \), the lowest Unit.

In the proposed multilevel model, there are \( n \) observations split into \( J \) Parishes and within \( j^{th} \) Parish, there are \( n_j \) observations labelled \( k = 1, 2, \ldots, n_j \). The response and covariate variables associated with the observations are denoted as \( y_{jk} \) and \( x_{jk} \) respectively from \( k^{th} \) ED within \( j^{th} \) Parish. The random errors for the \( j^{th} \) Parish and \( k^{th} \) ED are represented by \( u_j \sim N(0, \sigma_u^2) \) and \( \epsilon_{jk} \sim N(0, \sigma_{\epsilon}^2) \) respectively. The assumption is that the data satisfy the model below:

\[
y_{jk} = x_{jk} \beta + u_j + \epsilon_{jk}
\]  \hspace{1cm} (4.1)

The model in 3.1 is considered as a hierarchical model and may also be refer to a mixed-effect model because of the fixed and random components. The model may be extended to three or more groups in the application to the analysis of the JSCLC2007 and CHNS 1989 & 2011.
4.2 Weighting Adjustments for the Correction of Sample Selection

To understand the effects of the weights, this section seeks to elaborate on the weighting adjustments that will be used in the analysis to address sampling selection and missing data issues which affects the parameter estimates in the model. In this section, the theory of weight adjustment to correct the unequal probability in the selection of a defined Unit and sub populations within the Unit are discussed. The Unit selection weight is the first weight to be developed in this section. A Unit is an independent group or cluster in the hierarchy of the sampling process of non-uniform sampling technique. For example, a school and a classroom could be considered as examples of independent Units in the survey of students from a non-uniform sampling method. Another example is the Parish and Enumeration District in the JSRC 2007 survey. Below are the steps involved in deriving the probability of selecting a Unit and corresponding weight in an independent single stage sampling adapted from Little and Rubin (2002).

In a single stage sampling scheme when \( n_j \) units are selected from group \( j \).

Denoting the total number of units in group \( j \) as \( N_j \), the probability of this selection may be expressed as:

\[
\pi_j = n_j / N_j
\]

(4.2)

Thus, the weight \( w_j \) to address the unequal selection of the Unit is defined as the inverse of the probability \((\pi_j^{-1})\).

\[
w_j = \pi_j^{-1} = N_j / n_j
\]

(4.3)

However, in a two-stage sampling design where further sampling can take place within a Unit, another probability and corresponding weight will also be required to correct for unequal selection within this Unit. The Valliant, Dever & Kreuter (2013) approach is presented in the section below:

In the first stage of independent sampling, there are \( N \) Unit, the total observation, which is split into \( N_j \) Unit in group \( j \) and \( N_{jk} \) Unit in \( k^{th} \)
subgroup within group \( j \). Assuming that in the first stage is the same with the previous selection with similar weight as stated in equation (4.3), while in the second stage, \( n_{jk} \) units are selected from \( N_{jk} \) units with the probability of the second stage of selection defined as:

\[
\pi_{k|j} = \frac{n_{jk}}{N_{jk}}
\]  

(4.4)

The weight to address the unequal selection is the inverse of the probability of the second selection and define as:

\[
w_{k|j} = \pi_{k|j}^{-1} = \frac{N_{jk}}{n_{jk}}
\]  

(4.5)

Hence, the overall weight to address the first and second selection in the sampling process may be expressed as the product of the inverse probabilities from each stage of the selection as follows:

\[
w_{jk} = \pi_{k|j}^{-1} \pi_{j|k}^{-1} = \frac{N_{j} N_{jk}}{n_{j} n_{jk}}
\]  

(4.6)

The two-stage independent sampling developed above is synonymous to the selection of \( n_{j} \) schools from a total of \( N_{j} \) schools in the first stage of a selection while in the second stage, \( n_{jk} \) classrooms are selected from \( N_{jk} \) classrooms.

### 4.3 Logistic Regression Model and Item Non-Response Weight

The logistic regression model has been recommended for modelling binary response outcomes (McCullagh & Nelder, 1989, pg 110). In this study, the logistic regression model will be used to model the response and non-response taking into account the covariates. Below is an illustration of how the parameters will be estimated.

Let \( Z_{jki} = 1 \) if the \( i^{th} \) individual with the Enumeration District \( k \) did not respond. Suppose that the probability of non-response depends on a covariate \( x_\_\{jki\} \) with
\[ p_{jki} = P(Z_{jki} i = 1 | X = x_{jki}) \quad (4.7) \]

Then the logistic regression model relating the outcome variable and the predictor variable can be expressed as:

\[ \log e \left( \frac{p_{jki}}{1 - p_{jki}} \right) = \phi_0 + \phi_1 x_{jki} \quad (4.8) \]

Further to the estimation of the model parameters \( \phi_0, \phi_1 \) the probability is thus expressed as:

\[ p_{jki} = \frac{e^{\phi_0 + \phi_1 x_{jki}}}{1 + e^{\phi_0 + \phi_1 x_{jki}}} \quad (4.9) \]

Thus, item non-response weight (IR) for each individual \( i \) from group \( k \) within group \( j \)

\[ w_{jki} = 1 / \hat{p}_{jki} \quad (4.10) \]

### 4.4 Parameter Estimation

In this Section, parameter estimation will be via maximum likelihood estimation. I start with a simple linear regression model and build up to maximum likelihood estimation for multilevel models and the use of weights in the likelihood process.

#### 4.4.1 Parameter Estimation of Simple Linear Regression Model

Consider a linear relationship between a pair of response \( y_i \) and covariate \( x_i \) variables for \( i^{th} \) observations from a simple random setting which satisfy the regression model below:

\[ y_i = \beta_0 + \beta_1 x_i + \epsilon \quad (4.11) \]
where $\beta_0, \beta_1$ represent the intercept and slope of the model respectively.

The random components $\varepsilon_i$ are i.i.d with $N(0, \sigma^2)$ and this leads to the $y_i$ independent with the distribution as given. The random error may be expressed for the $i^{th}$ observation as:

$$
\varepsilon_i = y_i - \beta_0 - \beta_1 x_i 
$$

(4.12)

The likelihood function of a given observation $(x_i, y_i)$ and for unknown parameter $\theta = (\beta_0, \beta_1, \sigma^2)$ based on the model in (4.11) is:

$$
L(\theta | y_i) = \prod_i^n f(y_i ; \theta)
$$

(4.13)

$$
L(\beta_0, \beta_1, \sigma^2 | y_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-(\beta_0+\beta_1 x_i))^2}{2\sigma^2}}
$$

(4.14)

Then the log-likelihood is

$$
\log_e L(\theta | y_i) = \sum_{i=1}^n \log f(y_i ; \theta)
$$

$$
\log_e L = -\frac{n}{2} \log_e 2\pi - \frac{n}{2} \log_e \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i-(\beta_0+\beta_1 x_i))^2
$$

(3.16)

By taking the partial derivative of the log-likelihood, a set of equations are generated and set to zero and to find the maximum likelihood estimates of $\beta_0, \beta_1$ and $\sigma^2$ as follows:

$$
\frac{\partial \ln L(y_i ; \beta_0, \beta_1, \sigma^2)}{\partial \beta_0} = -\frac{1}{\sigma^2} \sum_{i}^n (y_i - \beta_0 - \beta_1 x_i) = 0
$$

(4.17)

$$
\frac{\partial \ln L(y_i ; \beta_0, \beta_1, \sigma^2)}{\partial \beta_1} = -\frac{1}{\sigma^2} \sum_{i}^n (y_i - \beta_0 - \beta_1 x_i) x_i = 0
$$

(4.18)
\[
\frac{\partial \ln L(y_i; \beta_0, \beta_1, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = 0 \tag{4.19}
\]

This gives the maximum likelihood estimates (MLEs):

\[
\hat{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = (X'X)^{-1} X'Y \tag{4.20}
\]

\[
\hat{\sigma}^2 = \frac{1}{n} (Y - X \hat{\beta})' (Y - X \hat{\beta}) \tag{4.21}
\]

In some instances, there are situations where the observation errors are independent but not identically distributed that is they do not meet the assumption that \( \epsilon_i \sim N(0, \sigma^2) \), a required assumption for regression models. Therefore, the introduction of a weight (\( w_i = 1/\sigma_i^2 \)), the reciprocal of the observation’s error term variance may become necessary to address this problem (Neter, Wasserman and Kutner, 1985, pg.169). This is the basis for the Weighted Least Squares (WLS) Regression Model.

The estimation of the parameters when the weights are introduced is illustrated in the scenario below.

Consider \( W_i \) as the diagonal matrix \( W(n \times n) \) of weights required for each observation.

\[
W = \begin{bmatrix}
w_1 & 0 & \cdots & 0 \\
w_2 & w_1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & w_n
\end{bmatrix} \tag{4.22}
\]
Initially, the likelihood without the weight was (3.13)

\[ L(\theta | y_i) = \prod_{i=1}^{n} f(y_i; \theta) \]

With the introduction of the weight the likelihood becomes:

\[ L(\theta | y_i) = \prod_{i=1}^{n} f(y_i; \theta)^{w_i} \] (4.23)

\[ \log_e L_w(\beta_{0w},\beta_{1w},\sigma_w^2|y_i) = \sum_{i=1}^{n} w_i \log f(y_i; \theta) \] (4.24)

\[ \log_e L_w(\beta_{0w},\beta_{1w},\sigma_w^2|y_i) = \sum_{i=1}^{n} w_i \log \left[ \sqrt{\frac{w_i}{2\pi}} \exp \left( -\frac{(y_i-(\beta_0-\beta_1x_i))^2}{2\sigma_w^2} \right) \right] \] (4.25)

\[ \log_e L_w = \frac{nw_i}{2} \log w_i - \frac{nw_i}{2} \log 2\pi - \frac{1}{2\sigma_w^2} \sum_{i=1}^{n} w_i (y_i-(\beta_0-\beta_1x_i))^2 \] (4.26)

The partial derivative of the log-likelihood are generated as follows:

\[ \frac{\partial \ln L(y_i; \beta_{0w},\beta_{1w},\sigma_w^2)}{\partial \beta_{0w}} = -\frac{1}{\sigma_w^2} \sum_{i=1}^{n} w_i (y_i-\beta_{0w}) = 0 \] (4.27)

\[ \frac{\partial \ln L(y_i; \beta_{0w},\beta_{1w},\sigma_w^2)}{\partial \beta_{1w}} = -\frac{1}{\sigma_w^2} \sum_{i=1}^{n} w_i (y_i-\beta_{0w}-\beta_{1w}x_i)x_i = 0 \] (4.28)

\[ \frac{\partial \ln L(y_i; \beta_{0w},\beta_{1w},\sigma_w^2)}{\partial \sigma_w^2} = -\frac{n}{2\sigma_w^2} + \frac{1}{2\sigma_w^4} \sum_{i=1}^{n} w_i (y_i-\beta_{0w}-\beta_{1w}x_i)^2 = 0 \] (4.29)

The Weighted Least Squares regression parameter estimators for \( w_i \neq 1 \) are:
\[ \hat{\beta}_w = \left( \begin{array}{c} \hat{\beta}_{0w} \\ \hat{\beta}_{1w} \end{array} \right) = (X'WX)^{-1} X'WY \]  
\[ \sigma^2 = \frac{1}{n}(Y - X \hat{\beta})' W (Y - X \hat{\beta}) \] 

4.3.2 Parameter Estimation of Multilevel Model

The parameter estimation is illustrated with a hierarchical model involving the Parish and the Enumeration District (ED) via the likelihood approach. In the JSLC 2007, there are fourteen Parishes denoted by \( j = 1 \ldots m \) and also, there are \( n_j \) EDs within parish \( j \). The EDs are denoted by \( k = 1 \ldots n_j \). In the development of a model \( y_{jk} \) represents the continuous response from \( k^{th} \) Enumeration District within the \( j^{th} \) Parish and \( x_{jk} \), the covariate for every \( y_{jk} \) observation. The random errors for the \( j^{th} \) Parish and \( k^{th} \) ED as \( u_j \sim N(0, \sigma_u^2) \) and \( \varepsilon_{jk} \sim N(0, \sigma_e^2) \) respectively for the purpose of developing a mixed-effect model that fits the structure of the data. The proposed model is:

\[ y_{jk} = x_{jk} \beta + u_j + \varepsilon_{jk} \]  

For the model in (4.32), the likelihood will be expressed as a joint distribution for the observed data in the Parishes and EDs with two sources of randomness since the values of \( u \) and \( y \) are known as follows:

\[ f(y|u, \beta, \sigma^2, \sigma_u^2) = \prod_{j=1}^{m} \prod_{k=1}^{n_j} \frac{1}{\sqrt{2\pi\sigma}} \exp \left( -\frac{1}{2\sigma^2} \left( y_{jk} - x_{jk} \beta - u_j \right)^2 \right) \]  
\[ f(u|\beta, \sigma^2, \sigma_u^2) = \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi\sigma_u}} \exp \left( -\frac{1}{2\sigma_u^2} u_j \right) \] 

47
By combining 3.34 and 3.35, the likelihood becomes:

\[
\begin{align*}
= \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi\sigma_U}} \exp \left(-\frac{1}{2\sigma^2} u_j^2 \right) \prod_{k=1}^{n_j} \frac{1}{\sqrt{2\pi\sigma}} \exp \left(-\frac{1}{2\sigma^2} \left(y_{jk} - x_{jk} \beta - u_j \right)^2 \right) du_j
\end{align*}
\]

(4.35)

\[
= \prod_{j=1}^{m} \left\{ \frac{1}{\sqrt{2\pi\sigma_U}} \exp \left(-\frac{1}{2\sigma_U^2} u_j^2 \right) \prod_{k=1}^{n_j} \frac{1}{\sqrt{2\pi\sigma}} \exp \left(-\frac{1}{2\sigma^2} \left(y_{jk} - x_{jk} \beta - u_j \right)^2 \right) du_j \right\}
\]

(4.36)

To further simplify the likelihood, the unknown parameters are denoted as \( \theta = (\beta, \delta, \delta_U) \), \( \delta = 1/\sigma^2 \), \( \delta_U = 1/\sigma_U^2 \), for \( N = \sum_{j=1}^{m} n_j \)

By substituting in 3.36 the revised likelihood given \( u \) is:

\[
L(\theta; y, u) = \prod_{j=1}^{m} \frac{\sqrt{\delta_U}}{\sqrt{2\pi}} \exp \left(-\frac{\delta_U}{2} u_j^2 \right) \prod_{k=1}^{n_j} \frac{\sqrt{\delta}}{\sqrt{2\pi}} \exp \left(-\frac{\delta}{2} \left(y_{jk} - x_{jk} \beta - u_j \right)^2 \right)
\]

(4.37)

Then the log-likelihood will be:

\[
l(\theta; y, u) = K + \frac{m}{2} \log \delta_U - \frac{\delta_U}{2} \sum_{j=1}^{m} u_j^2 + \frac{N}{2} \log \delta - \frac{\delta}{2} \sum_{j=1}^{m} \sum_{k=1}^{n_j} \left(y_{jk} - x_{jk} \beta - u_j \right)^2
\]

(4.38)

If \( K = -\frac{1}{2} \log(2\pi) (m + N) \), the log-likelihood further reduces to:

\[
= K + \frac{m}{2} \log \delta_U - \frac{\delta_U}{2} \sum_{j=1}^{m} u_j^2 + \frac{N}{2} \log \delta - \frac{\delta}{2} \sum_{j=1}^{m} \sum_{k=1}^{n_j} \left(y_{jk} - x_{jk} \beta \right)^2 - u_j \left(y_{jk} - x_{jk} \beta \right) + u_j^2
\]

(4.39)

This is now applied in the Expectation - Maximization algorithm to find Maximum Likelihood for the unknown parameters \( \theta = (\beta, \delta, \delta_U) \). The first step in the process is the identification of the distribution of \( u_j \) as:

\[
f(u_j | \theta, y, u_{-j}) \propto \exp \left(-\frac{\delta_U}{2} u_j^2 \right) \exp \left(-\frac{\delta}{2} \sum_{k=1}^{n_j} \left(y_{jk} - x_{jk} \beta - u_j \right)^2 \right)
\]

(4.40)

\[
\propto \exp \left(-\frac{\delta_U}{2} u_j^2 - \frac{\delta_n}{2} u_j^2 + \delta u_j \sum_{k=1}^{n_j} \left(y_{jk} - x_{jk} \beta \right) \right)
\]

(4.41)
\[
= \exp \left( - \frac{1}{2} \left[ u_j^2 \left( \delta_U + \delta n_j \right) - 2u_j \sum_{k=1}^{n_j} \left( y_{jk} - x_{jk} \beta \right) \right] \right) \tag{4.42}
\]

Hence, distribution thus can be conveniently expressed as follows:

\[
U_j | \theta, \gamma, u_{-j} \sim N \left( \sum_{k=1}^{n_j} \left( y_{jk} - x_{jk} \beta \right) / \left( \delta_U + \delta n_j \right), 1 / \left( \delta_U + \delta n_j \right) \right) \tag{4.43}
\]

In order to find the values that maximizes the log-likelihood, the E-step in the EM-algorithm was started with initial value of \( u_j = 0 \) for all \( j \) in (3.38). This is followed by the computation of \( E \left[ u_j | \theta, \gamma, u_{-j} \right] \) and \( E \left[ u_j^2 | \theta, \gamma, u_{-j} \right] \) in place of \( u_j \) and \( u_j^2 \) in (3.38) for the M-step. This is repeated for the re-computation of the E-step and M-step until convergence is achieved.

### 4.3.4 Pseudo Maximum Likelihood Estimation Method

The JSLC2007 scenario is used to illustrate the calculation of the weight in the parameter estimation while addressing unequal sample selection and for non-response. The Pseudo Maximum Likelihood has been proposed as the estimator for the weighted parameter (See Rozi, 2013; Skinner., 1998; Binder, 1983). This procedure will be used in this study. One of the issues in the non-uniform sampling is the unequal sampling or non-response. The following section is an illustration on the use of the weight in the correction of the unequal sampling of the \( k^{th} \) ED within the \( j^{th} \) Parish. Consider \( w_{k|j} \) as the weight for the correction of unequal sampling of the \( k^{th} \) ED within the \( j^{th} \) Parish. The weight is defined as
\( w_{k|j} = \pi_{k|j}^{-1} = N_{j,k} / n_{j,k} \). Using the earlier parameterisation \( \theta = (\beta, \delta, \delta_U) \).

This can be expressed as the pseudolikelihood:

\[
L_p(\theta; y, u) = \prod_{j=1}^{m} \prod_{k=1}^{n_j} f(y, u; \theta)^{w_{ki}}
\] (4.45)

Recalling the solution for the log-likelihood in 3.39, similar derivation will be express for the log-pseudolikelihood in 3.46 below

\[
l_p(\theta; y, u) = K + \frac{m}{2} w_{k|j} \log \delta_U - \frac{\delta_U}{2} \sum_{j=1}^{m} w_{k|j} u_j^2 + \frac{N}{2} \sum_{j=1}^{m} \sum_{k=1}^{n_j} \left\{ y_j - x_j \beta - u_j \right\}^2
\]

If \( w_{ki} = 1 \), the log-pseudolikelihood will equal the log-likelihood.

Hence, the parameter estimates that will maximize the log-pseudolikelihood, are identified from the EM algorithm explained earlier in this section.

4.4 Summary of the Statistical Methodology

In this Chapter, the weighting adjustments and associated parameter estimators for unweighted and weighted multilevel model are elaborated and will be utilized in the simulation study and the analyses of JSCLC2007 and CHNS 1989 &2011. In Chapters 5 and 6.
Chapter 5

Simulation study

Chapter Outline

In this chapter, I present a simulation study to assess the various weighting methods in obtaining a reliable parameter estimate. Existing techniques were used, along with a variety of choices to estimate parameters, including a range of statistical assessment tools to assess the reliability of the parameter estimate. The chapter is concluded with a discussion and recommendations for addressing non-response and sample selection bias.

5.1 Rationale and Purpose of the simulation study

The rationale for the simulation study is the assessment of the weighting method to address non-response and sampling bias problems in a survey data. The use of simulation studies to evaluate statistical models and methods is a common practice (see e.g., Asparouhov, 2006; Burton et al., 2006; Chiou & Muller, 2005; Yucel et al., 2018). This simulation experiment is designed for the primary purpose of providing answers to the following questions:

(a) What is the performance of weighting methods for sample selection and non-response problems?

(b) Under what condition(s) does a weighting method perform best?
5.2 Specific objectives of the simulation study

The specific objectives of this simulation investigation are to:

(a) Simulate the properties of a complex survey data using a multilevel linear model with a continuous outcome so that the best performed weighting methods may be identified.

(b) Evaluate the performance of the weighting methods by using the parameter estimates from different scenarios in the simulation study.

(c) Identify similarities and dissimilarities in the parameter estimates from weighting methods for different missing mechanisms and different proportions of omissions to identify the conditions for the weighting adjustment method.

5.3 Review of the Survey Weights in the Simulation Study

In the simulation study, four weights will be evaluated. This will include, Enumeration District (ED) selection weight ($W_j$); Household selection weight from the ED ($W_{kj}^{ij}$) within a Parish, design weight ($W_{jk}^{i}$), the item non-response weight ($W_{jkl}^{i}$). These weighting adjustments are illustrated in the following sections to complement the review.

5.3.1 Illustration of the weights using a case from the JSCLC 2007

For the purpose of illustration, consider an observation $i$ of a household with a household head of age 71 selected from an ED within the Parish. There are 217 households in the ED of which 16 households were sampled but 14 households completed the survey instrument from a Parish in the island. In this Parish, there are 126 EDs of which 4 EDs were selected.
The summary of the data is provided as follows:

\[ N_j = 126 \] is the total number of EDs in Parish \( j \).

\[ n_j = 4 \] is the number of EDs sampled in Parish \( j \).

Hence, the probability of selecting a given ED in Parish \( j \) is calculated as:

\[ \pi_j = \frac{n_j}{N_j} = \frac{4}{126} = 0.031746 \]

\[ w_{jPS} = p_j^{-1} = \frac{N_j}{n_j} = 31.5 \]

The 31.5 represents the weights for every ED sampled in that Parish.

To calculate the household weight, the procedure is as follows:

Let \( N_{jk} = 217 \) be the total number of the households or dwellings within \( k^{th} \) ED within the \( j^{th} \) Parish.

Let \( n_{jk} = 14 \) be the number of households that responded in \( k^{th} \) ED within the \( j^{th} \) Parish.

\[ \pi_{k|j} = \frac{n_{jk}}{N_{jk}} = \frac{14}{217} = 0.064516 \]

Thus, the weight for the household and subsequently household head selection in the ED within a Parish:

\[ w_{k|j}^{HS} = \pi_{k|j}^{-1} = \frac{N_{jk}}{n_{jk}} = 15.5 \]

The combined weights for ED and Household selections are denoted as \( \text{PHS} \) in the simulation study and calculated as:
In the calculation of the item non-response (IR), the age of the subject, 71 was applied in the logistic model below to deduce the propensity of responding as:

\[ \hat{p}_{jki} = \hat{p}_{jki}(71) = \frac{\exp(X'_{jki} \beta)}{1 + \exp(X'_{jki} \beta)} = 0.6905963 \text{ for an individual.} \]

The above probability was calculated for the household heads in the \( k^{th} \) ED for household \( l = 2,3,4,5,6,8,9,10,11,12,13,14 \) in Table 5.0.

### Table 5.0: Summary of household propensity to respond

<table>
<thead>
<tr>
<th>Household ( i )</th>
<th>( age_{jkl} )</th>
<th>( P_{jkl} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71</td>
<td>0.69060</td>
</tr>
<tr>
<td>2</td>
<td>41</td>
<td>0.88502</td>
</tr>
<tr>
<td>3</td>
<td>71</td>
<td>0.6906</td>
</tr>
<tr>
<td>4</td>
<td>39</td>
<td>0.89316</td>
</tr>
<tr>
<td>5</td>
<td>62</td>
<td>0.76392</td>
</tr>
<tr>
<td>6</td>
<td>59</td>
<td>0.78552</td>
</tr>
<tr>
<td>7</td>
<td>Nil</td>
<td>Nil</td>
</tr>
<tr>
<td>8</td>
<td>41</td>
<td>0.88502</td>
</tr>
<tr>
<td>9</td>
<td>66</td>
<td>0.73287</td>
</tr>
<tr>
<td>10</td>
<td>34</td>
<td>0.91131</td>
</tr>
<tr>
<td>11</td>
<td>Nil</td>
<td>Nil</td>
</tr>
<tr>
<td>12</td>
<td>67</td>
<td>0.72472</td>
</tr>
<tr>
<td>13</td>
<td>45</td>
<td>0.86713</td>
</tr>
<tr>
<td>14</td>
<td>44</td>
<td>0.87181</td>
</tr>
<tr>
<td>15</td>
<td>41</td>
<td>0.88502</td>
</tr>
<tr>
<td>16</td>
<td>40</td>
<td>0.88916</td>
</tr>
</tbody>
</table>

Thus, the response probability in for the households in \( k^{th} \) ED within the \( j^{th} \) Parish is:

\[ \bar{p}_{avg, jk} = 0.8197043. \]

The item non response weight (IR) is thus calculated as follows:

\[ W_{jki} = \left( \bar{p}_{avg, jk} \right)^{-1} = 1.219952. \]
The proposed weighting methods is summarized in Table 5.1

*Table 5.1: Sampling and non-response weights*

<table>
<thead>
<tr>
<th>Weight</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sampling Selection Weights</strong></td>
<td></td>
</tr>
<tr>
<td>ED selection weight</td>
<td>PS</td>
</tr>
<tr>
<td>Household selection weight</td>
<td>HS</td>
</tr>
<tr>
<td>ED and Household selection weights</td>
<td>PHS</td>
</tr>
<tr>
<td><strong>Non-Response Weight</strong></td>
<td></td>
</tr>
<tr>
<td>Item non-response weight</td>
<td>IR</td>
</tr>
<tr>
<td><strong>Sampling Selection Weight Adjusted For Missing Data</strong></td>
<td>SWAM</td>
</tr>
</tbody>
</table>

### 5.3.2 Scaled weights

The investigation on weighting scale in multilevel modelling continues to generate a debate in the literature (see Carle, 2009; Rabe–Hesketh & Skrondal, 2006). These studies focus on binary responses while investigating the effects of weight scaling in multilevel modelling. Rabe–Hesketh and Skrondal (2006) concluded that “for small clusters, estimated random -intercept variance was found to be biased” while Carle (2009, pg.6) also concurred that “scaled weighted estimates and standard errors differed slightly from unweighted analyses and observed that the differences were minimal and did not lead to different inferential conclusions”. These conclusions were not available for multilevel models with a continuous outcome.

The findings from these studies serve as the rationale for the inclusion of the effect of weight scaling in multilevel model using a continuous outcome. Accordingly, scaling methods proposed by Carle (2009, pg.9) were adopted for this study.
Scaling methods were applied to the design weights (PHS), item non-response weight (IR) and SWAM. The scaling methods are described as follows:

### 5.3.2.1 Method A

For method A, $w_{ij}^p$ for $p = PHS, IR$ & SWAM is denoted as the original weight to be scaled while $w_{ij}^{p,A}$ is denoted as the scaled weight for a given weight. In this method the weights are scaled to the size of $n_j$, that is, the number of sampled units in cluster $j$. Specifically,

the formula below is used to calculate scaled weight $A$ for $i^{th}$ household in the $k^{th}$ ED within $j^{th}$ Parish.

$$
W_{jkl}^{p,A} = W_{jkl}^p \left( \frac{n_j}{\sum_i W_{jkl}^p} \right)
$$

(5.1)

### 5.3.2.2 Method B

In the scaling process for Method B, the scale weight is denoted as $w_{ij}^{p,B}$. When the original weight of $w_{ij}^p$ was scaled to the sum of the weights in cluster $j$, the results are indicated in the formula below:

$$
W_{jkl}^{p,B} = W_{jkl}^p \left( \frac{\sum_i W_{jkl}^p}{\sum_i \left( W_{jkl}^p \right)^2} \right)
$$

(5.2)
5.4 Simulation Plan

Four simulation scenarios were considered. Each held a combination of 20%, 40% and 60% missing data with missing mechanism MAR and MNAR. The model below will be used to simulate the data using parameters of a prior analysis of the JSLC2007 as a guide:

\[
y_{jki} = x_{jki}^T \beta + u_j + v_{jk} + \varepsilon_{jki}
\]  

(5.3)

In the first scenario, the chosen parameter for the model was \( \beta' = (\beta_0 = 6, \beta_1 = -0.080) \) when age is the independent variable, while \( x_{jki}^T = (1, \text{age}_{jki}) \) of the individuals in the survey. For the second scenario, the parameter changes to \( \alpha' = (\alpha_0 = 6, \alpha_1 = 5) \) when the sex is the independent variable, and \( x_{jki}^T = (1, \text{sex}_{jki}) \). The third scenario model parameters combination of the age and sex covariates while the fourth is the extension of the third model with the inclusion of household size, education, employment and occupation of the individuals in the surveys. In the four models, the random components where \( u_j \sim N(0, \sigma^2_j) \), \( v_{jk} \sim N(0, \sigma^2_k) \) and \( \varepsilon_{jki} \sim N(0, \sigma^2_{ijkl}) \) for parish \( j \), ED, \( k \) and subject \( jkl \) levels respectively. The simulated data generating mechanism were represented as: \( Y_{1,jki}, Y_{2,jki}, Y_{3,jki}, Y_{4,jki} \) for each model. These simulated data sets were used in the recovery of the weighted parameter estimates for reliability of the weighting methods.

5.5 Missing at Random (MAR) and Missing Not at Random (MNAR)

The missing at random (MAR) mechanism was generated using a binomial function \( \text{Miss} \), where \( \text{Miss} \sim \text{Bin}(n, p) \), for \( n = 1994 \) and \( p = 0.2, 0.4, 0.6 \). The logistic model \( \text{loMiss} \) was then applied to create the missing not at random (MNAR) using the R-Software Version 3.5.2
The logistic regression model function used to create the scenario for MNAR is:

\[
loMiss = \frac{e^{(\Delta_0 + \Delta_1 \text{age})}}{1 + e^{(\Delta_0 + \Delta_1 \text{age})}}
\]  

(5.4)

The model parameters were set at different values for \( \Delta_0 = 0.07, 0.88, 1.91 \) and \( \Delta_1 = -0.03 \); while aiming for 20%, 40% and 60% missing respectively, based on the data generation model in the simulation study. These parameters were carefully chosen and varied so as to achieve the rate of missing in relation to the population of interest. A total of \( N=100 \) repetitions was simulated for each weighted multilevel model containing 1994 cases in each scenario where MNAR and MAR represented rate of missing (20%, 40%, and 60%).

5.6 Simulation Statistical Assessment Tools

To evaluate the bias in the parameter estimates, the indicator for performance of the weighting adjustment methods, the error in the estimates were calculated to identify the relative closeness of the estimated and true parameters. The calculation was expanded by estimating the average errors to determine the root mean square error (RMSE) for each weighting method. Yucel et al.’s (2018) percent bias (PB) approach assessed the performance of estimated parameters relative to the true value. The strength of this approach was the availability of large PB value for large error size and vice versa for small error size. Additionally, in order for a method to be effective, the PB should not exceed 5 percent (Demirtas, Freels, & Yucel, 2008).
The RMSE and PB were derived for each weighting methods as follows:

Let $T$ be the true value of the parameter in the data generation models and $E(\hat{T})$ is the average estimated for parameter of interest (PI) from the evaluative models.

The average is estimated thus:

$$E(\hat{T}) = \bar{T} = \frac{1}{N} \sum_{i=1}^{N=100} \hat{T}$$

for $i=1\ldots N=100$.

Ultimately leading to the percent bias defined as:

$$PB = 100 \times \left| \frac{E(\hat{T}) - T}{T} \right|$$

For the RMSE, $\hat{T}_i$ is defined as the estimated parameter from each model and $T_i$ be the true parameter, the errors were calculated as

$$e_i = T_i - \hat{T}_i$$

Hence, the mean square error

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} e_i^2$$

for $i=1,2,3,4\ldots N=100$.

And finally

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} e_i^2}$$
5.7 Results of the Simulation Study

5.7.1 Simulation Study without Incorporation of Varying Intraclass Coefficient

There are series of results originating from the parameter estimates for the fixed and random components in the simulation experiment study. The result of the parameter estimates for the continuous and categorical variable were used as a generalization for all continuous and categorical variables. The recovered parameter estimates for age and sex variables were selected as the illustration of the results. To understand the findings on the recovery of the parameter based on the weighting methods, Table 5.2, an excerpt was created from more detailed findings in Tables 5.5 and 5.6. This illustration is from the series of simulation on parameter estimates for continuous variable using the data generating mechanism $Y_{i,j,k,l}$ and a true value $\beta_i = -0.08$.

Table 5.2.: Root mean square error (RMSE) for the ED selection weight (PS), Household selection weight (HS), Combination of ED and Household selection weight (PHS) and sampling weight adjusted for item non response (SWAM).

<table>
<thead>
<tr>
<th></th>
<th>MAR RMSE VALUES</th>
<th>MNAR RMSE VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20%</td>
<td>40%</td>
</tr>
<tr>
<td>WM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS</td>
<td>0.0721</td>
<td>0.0722</td>
</tr>
<tr>
<td>HS</td>
<td>0.0723</td>
<td>0.0722</td>
</tr>
<tr>
<td>PHS</td>
<td>0.0723</td>
<td>0.0724</td>
</tr>
<tr>
<td>IR</td>
<td>0.0722</td>
<td>0.0722</td>
</tr>
<tr>
<td>SWAM</td>
<td>0.0019</td>
<td>0.0736</td>
</tr>
<tr>
<td>Unweighted</td>
<td>0.0721</td>
<td>0.0722</td>
</tr>
</tbody>
</table>

According to Table 5.2, the root mean square error for the PS, HS, PHS and IR weighting methods were relatively similar to the unweighted method. But this changes when the sampling weights is adjusted for missing data.
using the item nonresponse weight (SWAM), the root mean square error were relatively smaller compared with the other weighing methods without adjustment for missing data.

In Tables 5.5 and 5.6, the detailed results for MAR and MNAR are presented respectively for the PS, HS, & PHS weighting method. The rate of missing for $Y_{3, jk}$ and $Y_{4, jk}$ were set at 20%, 40% and 60%. For the MAR, the weighted parameter estimates from $\hat{\lambda}_{1_{UP}}$, $\hat{\lambda}_{1_{PS}}$, $\hat{\lambda}_{1_{HS}}$, and $\hat{\lambda}_{1_{PHS}}$ all were within acceptable limit of the percent bias (<5%) while the least root mean square error was observed for item non-response weight (IR) for $Y_{3, jk}$. For the overall results, of the three selection weights (PS, HS, PHS), the HS appears to outperform the other selection weights regardless of the scaling methods A or B. All the parameter estimates for age from the weighted model were found to be significant (p<0.005). Consequently, the parameter estimates from the simulation scenario involving the data generating mechanism that contains a categorical covariate variable sex, the parameter estimates from all the different weighting methods $\hat{\lambda}_{2_{UP}}$, $\hat{\lambda}_{2_{PS}}$, $\hat{\lambda}_{2_{HS}}$, $\hat{\lambda}_{2_{PHS}}$, and $\hat{\lambda}_{2_{IR}}$ were also found to be within the acceptable limit of a 5% bias. The result also revealed that IR has the best performed weighting method with the least root mean square error. The findings also revealed that the root mean square error increases as the rate of missing increases. Also, these parameter estimates were not significant (p > 0.05) unlike the coefficient for the continuous variable.

In the $Y_{4, jk}$ data generation model, all the weighting methods were found to be reliable for the continuous covariate because the parameter estimates ($\tau_{1_{UP}}$, $\hat{\tau}_{1_{PS}}$, $\hat{\tau}_{1_{HS}}$, and $\hat{\tau}_{1_{PHS}}$) have percent bias within the acceptable limit. This is in contrast to the parameter estimates for the categorical covariate in the model for sex variable estimates ($\tau_{3_{UP}}$, $\hat{\tau}_{3_{PS}}$, $\hat{\tau}_{3_{HS}}$, $\hat{\tau}_{3_{PHS}}$, and $\hat{\tau}_{3_{IR}}$) which shows that the weighting methods were unreliable. In the MNAR scenario, the results were similar as the MAR scenario with item non-response weight (IR) was found to have the least root mean square at the 20% rate of missing followed by the
household selection weight (HS). In the MAR and MNAR scenarios for $Y_{jki}$ and $Y_{2, jki}$ data generating mechanism were also found to produced reliable parameter estimates for both continuous and categorical covariates (See Tables 5.5 & 5.6)

Tables 5.7 and 5.8 contains the findings of the scaled weights using Method A, while similar findings for Method B can be found in Appendix E. The results from Method A were similar to the findings from Method B for both MAR and MNAR scenarios. The scaled weights in the simulation study included PHSA, IRA, PHSB, IRB, which investigated the effects of scaling. In addition, both Tables 5.7 and 5.8, the weights methods were found to be reliable because the root mean square were found to be minimal while the percent bias was within the acceptable limit of less than 5% based on estimates from $Y_{jki}$, $Y_{2, jki}$, $Y_{3, jki}$. In $Y_{4, jki}$, the weighting methods were unreliable for the estimates of the categorical covariate ($\hat{\tau}_{3, jki}$, $\hat{\epsilon}_{3, jki}$, and $\hat{\epsilon}_{3, jki}$). The item non-response weight (IRA) was observed to have the least root mean square error among all the scaled weights followed by the design weight (PHSA). Similar results are contained in other tables in Appendix E.

This section contains the results of the performance of the weight with regards to the random components. Three random components were examined: $u_j$ for the parish cluster, $v_{jk}$ for the PSU cluster and $e_{jki}$ for the household level. The true value for these random components were 3, 2, and 1, respectively. The results for the random components $u_j$ and $v_{jk}$ from the MAR and MNAR scenarios revealed that the parameter estimates were unreliable and the weighting methods performed poorly. These results are not illustrated in this section but can be found in Appendix E. In Table 5.15, the estimates from MAR for $e_{jki}$ are presented. A thorough analysis of these results in Table 5.15 revealed that all the weights performed and were reliable except for estimates of the weights from the household selection (HS). Of all the weighting methods, the PS and IR weights were found to produce better estimates than other weights in the estimation for $e_{jki}$. 


5.7.2 Simulation Study with Incorporation of Varying Intraclass Correlation Coefficient

The objective of this simulation is to determine the performance of the weighting adjustment methods under varying intra-correlation coefficient. In this simulation study different values of the random components were selected arbitrarily to deduce different values of the Intra-class correlation coefficients (ICCs) at the Parish, the highest and Enumeration District which is referred to as the Primary Sampling Unit respectively. For this simulation three scenarios were investigated with different values of the random components under Missing at Random (MAR) and Missing Not at Random (MNAR) while proportion of missing cases in the dependent variable was set at 20%. For the first scenario the random components were \( u_j \sim N(0,4^2) \), \( v_{jk} \sim N(0,7^2) \) and \( \varepsilon_{jki} \sim N(0,10^2) \). These values were reversed in the second scenario as \( u_j \sim N(0,10^2) \), \( v_{jk} \sim N(0,7^2) \) and \( \varepsilon_{jki} \sim N(0,4^2) \). In the third scenario, lower values were selected \( u_j \sim N(0,5^2) \), \( v_{jk} \sim N(0,4^2) \) and \( \varepsilon_{jki} \sim N(0,3^2) \). In the earlier simulation study, the random components were \( u_j \sim N(0,3^2) \), \( v_{jk} \sim N(0,2^2) \) and \( \varepsilon_{jki} \sim N(0,1^2) \) for parish \( j \), the Primary Sampling Unit (PSU), \( k \) and subject \( jkl \) levels respectively. Hence the Intra-class correlation coefficient (ICC) values for the Parish, the Primary Sampling Units (PSU) and individual levels remained constant during the simulation.

Findings

The findings Table 5.3 involving 20% proportion missing at random (MAR) revealed that the root mean square error reduces for the estimates from the four weighting adjustments investigated as the ICC values increases. Similar findings were observed in Table 5.4 with same 20% proportion of missing not at random (MNAR). Of the three scenarios investigated, the findings for the 0.4586 or 45.86% ICC for the parish and 0.7738 or 77.38% for the PSU generated based on
the random component for: $u_j \sim N(0,5^2)$, $v_{jk} \sim N(0,4^2)$ and $\varepsilon_{jki} \sim N(0,3^2)$ had the most reduced root mean square errors for each of the estimates investigated. At the 20% proportion of missing for either MAR or MNAR, the findings suggest that the varying ICC values in the weighted multilevel model can affect the performance of the estimator. The incorporation of the ICCs values will be necessary in simulation studies to enable reliable estimates. Further simulation studies with higher proportions of missing under MAR and MNAR mechanisms will be necessary to generalize beyond the 20% missing cases.

Table 5.3: Parameter Estimate of a weighted three level model incorporating different values of ICC under MAR Mechanism at 20% proportion of missing

<table>
<thead>
<tr>
<th>U</th>
<th>V</th>
<th>E</th>
<th>ICC -AVERAGE</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ICC- PARISH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>10</td>
<td>0.0795</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ICC- PSU</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3671</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
<td>0.4586</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.7738</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>4</td>
<td>0.5578</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8803</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Weighting Adjustment Methods (WM); True Value (TV); Average Estimate (AE); Root Mean Square Error (RMSE); Intraclass correlation (ICC)
Table 5.4: Parameter Estimate of a weighted three level model incorporating different values of ICC under MNAR Mechanism at 20% proportion of missing

<table>
<thead>
<tr>
<th>U</th>
<th>V</th>
<th>E</th>
<th>ICC- PARISH</th>
<th>ICC- PSU</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>7</td>
<td>10</td>
<td>0.0845</td>
<td>0.3777</td>
<td>HS</td>
<td>-0.0800</td>
<td>-0.0797</td>
<td>0.0174</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PS</td>
<td>-0.0800</td>
<td>-0.0786</td>
<td>0.0191</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PHS</td>
<td>-0.0800</td>
<td>-0.0786</td>
<td>0.0196</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IR</td>
<td>-0.0800</td>
<td>-0.0790</td>
<td>0.0162</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
<td>0.4280</td>
<td>0.7663</td>
<td>HS</td>
<td>-0.0800</td>
<td>-0.0807</td>
<td>0.0055</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PS</td>
<td>-0.0800</td>
<td>-0.0810</td>
<td>0.0053</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PHS</td>
<td>-0.0800</td>
<td>-0.0809</td>
<td>0.0059</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IR</td>
<td>-0.0800</td>
<td>-0.0807</td>
<td>0.0049</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>4</td>
<td>0.5499</td>
<td>0.8789</td>
<td>HS</td>
<td>-0.0800</td>
<td>-0.0809</td>
<td>0.0067</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PS</td>
<td>-0.0800</td>
<td>-0.0814</td>
<td>0.0072</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PHS</td>
<td>-0.0800</td>
<td>-0.0810</td>
<td>0.0076</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IR</td>
<td>-0.0800</td>
<td>-0.0811</td>
<td>0.0062</td>
</tr>
</tbody>
</table>

**5.8 Summary of the simulation study**

The simulation study provided clearer understanding of parameter estimation from weighted multilevel models with varying rate of missing. The findings in relation to the earlier questions which facilitated the design of the simulation study under the different headings are summarized in three broad headings: performance of weighting adjustment; and conditions under which a weighting adjustment will perform best.

**5.9.1 Performance of the Weighting Methods for Sample Selection Bias**

The findings from the simulation study on the sample selection weights - **PS**, **HS**, and **PHS** - show that the selection weights seldom produced reliable estimates, especially for continuous covariates in any multilevel model. Based on the information provided on the survey design, any of the weights in the study could be used to address sample selection bias. Of the three selection weights used in the study, the **PS** was the first level weight most performed, most reliable and with the
lower root mean square error. This was followed by the HS, which was the second level weight for the majority of the scenarios in the simulation study. However, the design weight, the combination of both PS and HS which addresses unequal selections at both stages of the survey design exhibited higher root mean square error and was found to be unreliable.

5.9.2 Performance of the Weighting Methods for Non-response Problems

The item non-response weight was developed to address non-response bias in the simulation study. The results from the scenarios in the simulation study revealed that, item non-response weight (IR) was found to be reliable estimate from the results. This was confirmed from the root mean square errors obtained from the simulation. One of the advantages of the item non-response is that the weight development relies on the available information in the data unlike the unit non-response adjustment and design weights. Both the unit non-response adjustment and design weights will require the use of survey design information which may not be available for secondary analysis.

5.9.3 Performance of the Sampling Weighting Adjusted for Missing Data using the Item Non-response Weight

Adjusting for missing data at the individual level in addition to the adjustment for the bias due to unequal probabilities of the level one and two Units of selection revealed much lower root mean square error for the recovery of the true parameter estimates. Especially under 20% proportion of missing for simulated data with continuous, categorical data generating mechanisms (Tables 5.11 and 5.12) in comparison with the simulated data from multivariate data generating mechanism. Of all the estimates, adjusting for sample selection probabilities and missing data produce better estimates when the relevant information on the survey design is available.
5.9.4 Performance of Scaled Weight

Three weights (PHS, IR & SWAM) were scaled to the size of the cluster and also to the sum of the weights within the cluster to investigate the effects of scaling. The reliability of the estimated parameters from the scaled weights were found to be similar for the MAR and MNAR scenarios from either method of scaling. The estimates from the scaled item non-response weight were observed to have the least root mean square and judged to be the best of all the methods. This suggests that the scaling added no value to the parameter estimates apart from a reduction in error value.

5.9.5 Identifiable conditions for weighted multilevel models in the presence of missingness/omission

The analyses of the results of the simulation study show that the root mean square error for the 20% rate of missing cases were relatively reduced when compared with the values for 40% and 60% rate of missing cases. This implies that the fewer the cases of missing, the better the performance of the weighting methodology. It was deduced that at the 20% rate of missing, all the parameter estimates were found to be reliable except for the categorical variable in the weighted multivariable model which was not significant.

5.9.6 Performance of the random components

The simulation experiment affirms poor performance of the weights except for besides the item non-response weight in the parameter estimates of the random components. The item non-response weight produced reliable estimates for the individual level random component. Other random components for the other cluster levels such as the parish and ED were poorly estimated by the weights.
5.9.6 Lessons Learnt from the simulation study

The conclusion from this simulation study is that item non-response weighting appears to be reliable across the scenarios of missing at random and missing not at random scenarios for rates of missing under 20%. Similarly, weighting methods for known design information at the cluster levels can be considered for multilevel modelling of outcome variable to address sample selection bias. The weighting adjustment for the unit non-response appears to be less performed in the simulation when compared with the item non-response and sample selection weights at different levels.

5.9.7 Future simulation study on weighting methods

This simulation study may be extended in future. This would be possible through the simulation of multiple imputation of a given outcome variable with missing data from a survey data at a specified varying rate of missing cases. The objective would be to obtain similar root mean square errors from weighting methods to determine the conditions at which both methods will produce same magnitude of errors in the parameter estimates under the missing at random (MAR) and missing not at random (MNAR) scenarios.

Any future study will seek to find root mean square error at which parameter estimates from either method will be similar thereby enabling equivalence in the handling of missing data in the analysis of survey data. This research will be a useful tool to guide modelling of continuous, binary and multinomial variables in multilevel modelling.
Table 5.5: Summary of the simulation results when the missing mechanism is MAR with missing rates of 20%, 40% and 60% for each variable for sex - age and multivariable covariates models

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_{3ijkl}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\lambda_1</td>
<td>UP</td>
<td>-0.08000</td>
<td>-0.07985</td>
<td>0.22500</td>
<td>0.00153</td>
<td>-0.08020</td>
<td>0.25000</td>
<td>0.00155</td>
<td>-0.07933</td>
<td>0.83750</td>
<td>0.00250</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS</td>
<td>-0.08000</td>
<td>-0.07985</td>
<td>0.18750</td>
<td>0.00167</td>
<td>-0.08025</td>
<td>0.31250</td>
<td>0.00181</td>
<td>-0.07946</td>
<td>0.67500</td>
<td>0.00269</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS</td>
<td>-0.08000</td>
<td>-0.07985</td>
<td>0.15000</td>
<td>0.00177</td>
<td>-0.08031</td>
<td>0.38750</td>
<td>0.00177</td>
<td>-0.07930</td>
<td>0.87500</td>
<td>0.00288</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHS</td>
<td>-0.08000</td>
<td>-0.07985</td>
<td>0.08750</td>
<td>0.00183</td>
<td>-0.08032</td>
<td>0.40000</td>
<td>0.00199</td>
<td>-0.07943</td>
<td>0.71750</td>
<td>0.00305</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>-0.08000</td>
<td>-0.07985</td>
<td>0.20875</td>
<td>0.00155</td>
<td>-0.08020</td>
<td>0.24750</td>
<td>0.00156</td>
<td>-0.07934</td>
<td>0.82375</td>
<td>0.00249</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_{4ijkl}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\tau_1</td>
<td>UP</td>
<td>-0.07800</td>
<td>-0.08084</td>
<td>3.10769</td>
<td>0.00195</td>
<td>-0.00791</td>
<td>1.39615</td>
<td>0.00231</td>
<td>-0.00765</td>
<td>1.93282</td>
<td>0.00283</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS</td>
<td>-0.07800</td>
<td>-0.07979</td>
<td>2.24872</td>
<td>0.00233</td>
<td>-0.00785</td>
<td>0.58654</td>
<td>0.00252</td>
<td>-0.00758</td>
<td>2.80897</td>
<td>0.00310</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS</td>
<td>-0.07800</td>
<td>-0.07979</td>
<td>2.24872</td>
<td>0.00202</td>
<td>-0.00803</td>
<td>3.00769</td>
<td>0.00261</td>
<td>-0.00769</td>
<td>1.38923</td>
<td>0.00289</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHS</td>
<td>-0.07800</td>
<td>-0.07979</td>
<td>2.17708</td>
<td>0.00231</td>
<td>-0.00793</td>
<td>1.68000</td>
<td>0.00279</td>
<td>-0.00761</td>
<td>2.49064</td>
<td>0.00336</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>-0.07800</td>
<td>-0.08084</td>
<td>3.04872</td>
<td>0.00196</td>
<td>-0.00791</td>
<td>1.41667</td>
<td>0.00231</td>
<td>-0.00765</td>
<td>1.95388</td>
<td>0.00286</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI = Parameter of interest; WM = Weighting methods; TV = True value; AE = Average Estimate for parameter of interest; PB = Percent bias; RMSE = Root mean square error; UP = Unweighted parameter; PS = ED selection weight; HS = Household selection weight; PHS = ED and Household selection weight; IR = Item non-response weight.

69
Table 5.6: Summary of the simulation results when the missing mechanism is MNAR with missing rates of 20%, 40% and 60% for each variable for sex - age and multivariable covariates models

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>20%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>40%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>60%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DGM = Data generation model; PI= Parameter of interest; WM =Weighting methods; TV = True value; AE=Average Estimate for parameter of interest; PB=Percent bias; RMSE=Root mean square error; UP= Unweighted parameter; PS=ED selection weight; HS =Household selection weight ; PHS=ED and Household selection weight; IR=Item non-response weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.7: Parameter estimates from weighted multilevel models (scaled method A) from the missing mechanism is MAR with missing rates of 20%, 40% and 60% for each data simulated from sex- age and multivariable covariate models

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>20%</th>
<th>40%</th>
<th>60%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
</tr>
<tr>
<td>$Y_{1,jkl}$</td>
<td>$\beta_1$</td>
<td>UP</td>
<td>-0.08000</td>
<td>-0.08016</td>
<td>0.20000</td>
<td>0.07217</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSA</td>
<td>-0.08000</td>
<td>-0.08026</td>
<td>0.32875</td>
<td>0.07219</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRA</td>
<td>-0.08000</td>
<td>-0.08019</td>
<td>0.23250</td>
<td>0.07220</td>
</tr>
<tr>
<td>$Y_{2,jkl}$</td>
<td>$\alpha_1$</td>
<td>UP</td>
<td>5.00000</td>
<td>5.00522</td>
<td>0.10440</td>
<td>0.05698</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSA</td>
<td>5.00000</td>
<td>5.00517</td>
<td>0.10340</td>
<td>0.07482</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRA</td>
<td>5.00000</td>
<td>5.00380</td>
<td>0.07600</td>
<td>0.05910</td>
</tr>
<tr>
<td>$Y_{3,jkl}$</td>
<td>$\lambda_1$</td>
<td>UP</td>
<td>-0.08000</td>
<td>-0.07982</td>
<td>0.22000</td>
<td>0.01535</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSA</td>
<td>-0.08000</td>
<td>-0.07994</td>
<td>0.07875</td>
<td>0.00185</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRA</td>
<td>-0.08000</td>
<td>-0.07982</td>
<td>0.22000</td>
<td>0.01535</td>
</tr>
<tr>
<td>$Y_{4,jkl}$</td>
<td>$\tau_1$</td>
<td>UP</td>
<td>-0.00780</td>
<td>-0.00786</td>
<td>0.74359</td>
<td>0.00193</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSA</td>
<td>-0.00780</td>
<td>-0.00786</td>
<td>0.71538</td>
<td>0.00230</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRA</td>
<td>-0.00780</td>
<td>-0.00757</td>
<td>2.96538</td>
<td>0.00200</td>
</tr>
<tr>
<td>$Y_{5,jkl}$</td>
<td>$\tau_2$</td>
<td>UP</td>
<td>-0.02500</td>
<td>-0.03161</td>
<td><strong>26.44281</strong></td>
<td>0.04828</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSA</td>
<td>-0.02500</td>
<td>-0.03230</td>
<td><strong>29.19424</strong></td>
<td>0.06440</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRA</td>
<td>-0.02500</td>
<td>-0.03041</td>
<td><strong>21.62552</strong></td>
<td>0.04915</td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI= Parameter of interest; WM = Weighting methods; TV = True value; AE=Average Estimate for parameter of interest; PB=Percent bias; RMSE = Root mean square error; UP= Unweighted parameter; PS=ED selection weight; HS = Household selection weight; PHS=ED and Household selection weight; IR=Item non-response weight

SCALED METHOD A - MISSING AT RANDOM (MAR)
Table 5.8: Parameter estimates from weighted multilevel models (scaled method A) from the missing mechanism is MNAR with missing rates of 20%, 40% and 60% for each data simulated from age-only; sex-only; sex-age and multivariable covariate models

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>20%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>40%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>60%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_{1,j}</td>
<td>\beta_1</td>
<td>UP</td>
<td>-0.08000</td>
<td>-0.07992</td>
<td>0.10000</td>
<td>0.07194</td>
<td>-0.08014</td>
<td>0.17250</td>
<td>0.07217</td>
<td>-0.07989</td>
<td>0.13750</td>
<td>0.07194</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSA</td>
<td>-0.08000</td>
<td>-0.07988</td>
<td>0.15125</td>
<td>0.07191</td>
<td>-0.07998</td>
<td>0.02500</td>
<td>0.07201</td>
<td>-0.07993</td>
<td>0.09000</td>
<td>0.07199</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRA</td>
<td>-0.08000</td>
<td>-0.07999</td>
<td>0.01750</td>
<td>0.07200</td>
<td>-0.08009</td>
<td>0.11125</td>
<td>0.07212</td>
<td>-0.07986</td>
<td>0.17125</td>
<td>0.07192</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_{2,j}</td>
<td>\alpha_1</td>
<td>UP</td>
<td>5.00000</td>
<td>4.99988</td>
<td>0.00240</td>
<td>0.00591</td>
<td>4.99467</td>
<td>0.07217</td>
<td>0.07989</td>
<td>0.07993</td>
<td>0.09000</td>
<td>0.07199</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSA</td>
<td>5.00000</td>
<td>4.99901</td>
<td>0.01980</td>
<td>0.06597</td>
<td>4.98983</td>
<td>0.07201</td>
<td>0.07993</td>
<td>0.09000</td>
<td>0.07199</td>
<td>0.07989</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRA</td>
<td>5.00000</td>
<td>5.00129</td>
<td>0.01750</td>
<td>0.07200</td>
<td>4.99612</td>
<td>0.07760</td>
<td>0.07993</td>
<td>0.09000</td>
<td>0.07199</td>
<td>0.07989</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_{3,j}</td>
<td>\lambda_2</td>
<td>UP</td>
<td>-0.08000</td>
<td>-0.08034</td>
<td>0.21700</td>
<td>0.05398</td>
<td>5.00473</td>
<td>0.09460</td>
<td>0.06083</td>
<td>4.99372</td>
<td>0.12560</td>
<td>0.07657</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSA</td>
<td>-0.08000</td>
<td>-0.08042</td>
<td>0.52125</td>
<td>0.00193</td>
<td>5.00968</td>
<td>0.19360</td>
<td>0.08148</td>
<td>4.99564</td>
<td>0.08720</td>
<td>0.08539</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRA</td>
<td>-0.08000</td>
<td>-0.08037</td>
<td>0.45625</td>
<td>0.00170</td>
<td>5.00507</td>
<td>0.10140</td>
<td>0.06302</td>
<td>4.99345</td>
<td>0.13100</td>
<td>0.07459</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_{4,j}</td>
<td>\tau_3</td>
<td>UP</td>
<td>-0.02500</td>
<td>-0.02113</td>
<td>0.74359</td>
<td>0.00188</td>
<td>-0.00794</td>
<td>0.07000</td>
<td>0.00166</td>
<td>0.08004</td>
<td>0.04625</td>
<td>0.00240</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSA</td>
<td>-0.02500</td>
<td>-0.00786</td>
<td>1.12821</td>
<td>0.00214</td>
<td>-0.00794</td>
<td>0.03125</td>
<td>0.00209</td>
<td>0.07995</td>
<td>0.06500</td>
<td>0.00295</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRA</td>
<td>-0.02500</td>
<td>-0.07899</td>
<td>0.86795</td>
<td>0.00205</td>
<td>-0.00794</td>
<td>0.07750</td>
<td>0.00175</td>
<td>0.08007</td>
<td>0.09250</td>
<td>0.00238</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI = Parameter of interest; WM = Weighting methods; TV = True value; AE = Average Estimate for parameter of interest; PB = Percent bias; RMSE = Root mean square error; UP = Unweighted parameter; PHSA = ED and Household selection weight using scaled Method A; IRA = Item non-response weight using scaled Method A
Table 5.9: Random Component – estimates from weighted multilevel models from the missing mechanism is MAR with missing rates of 20%, 40% and 60% for each data simulated from age-only; sex-only; sex- age and multivariable covariate models

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>20%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>40%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>60%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{1,ijkl}$</td>
<td>E</td>
<td>UP</td>
<td>1.00000</td>
<td>0.99779</td>
<td>0.22119</td>
<td>0.01824</td>
<td>0.99976</td>
<td>0.02384</td>
<td>0.02041</td>
<td>0.99734</td>
<td>0.02660</td>
<td>0.03157</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PS</td>
<td>1.00000</td>
<td>0.99806</td>
<td>0.19430</td>
<td>0.02244</td>
<td>0.92759</td>
<td>7.24080</td>
<td>0.02499</td>
<td>0.99458</td>
<td>0.54192</td>
<td>0.03657</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HS</td>
<td>1.00000</td>
<td>0.94483</td>
<td>5.51740</td>
<td>0.05830</td>
<td>0.92749</td>
<td>7.25131</td>
<td>0.07568</td>
<td>0.88442</td>
<td>11.55830</td>
<td>0.11931</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHS</td>
<td>1.00000</td>
<td>0.99782</td>
<td>0.21833</td>
<td>0.02388</td>
<td>0.99924</td>
<td>0.07562</td>
<td>0.02755</td>
<td>0.99352</td>
<td>0.64814</td>
<td>0.03882</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IR</td>
<td>1.00000</td>
<td>0.99777</td>
<td>0.22319</td>
<td>0.01827</td>
<td>0.99977</td>
<td>0.02340</td>
<td>0.02035</td>
<td>0.99719</td>
<td>0.28091</td>
<td>0.03161</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{2,ijkl}$</td>
<td>E</td>
<td>UP</td>
<td>1.00000</td>
<td>0.99538</td>
<td>0.46173</td>
<td>0.02015</td>
<td>0.99873</td>
<td>0.12680</td>
<td>0.02205</td>
<td>1.00159</td>
<td>0.15913</td>
<td>0.02671</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PS</td>
<td>1.00000</td>
<td>0.99757</td>
<td>0.24325</td>
<td>0.02182</td>
<td>1.0042</td>
<td>0.04244</td>
<td>0.02483</td>
<td>1.00357</td>
<td>0.35730</td>
<td>0.02885</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HS</td>
<td>1.00000</td>
<td>0.94207</td>
<td>5.79348</td>
<td>0.06154</td>
<td>0.92472</td>
<td>7.52834</td>
<td>0.07827</td>
<td>0.89000</td>
<td>10.99964</td>
<td>0.11232</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHS</td>
<td>1.00000</td>
<td>0.99715</td>
<td>0.28461</td>
<td>0.02360</td>
<td>0.99945</td>
<td>0.05516</td>
<td>0.02491</td>
<td>1.00285</td>
<td>0.28543</td>
<td>0.02891</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IR</td>
<td>1.00000</td>
<td>0.99530</td>
<td>0.47006</td>
<td>0.02018</td>
<td>0.99881</td>
<td>0.11906</td>
<td>0.02218</td>
<td>1.00141</td>
<td>0.14107</td>
<td>0.02688</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{3,ijkl}$</td>
<td>E</td>
<td>UP</td>
<td>1.00000</td>
<td>0.99818</td>
<td>0.18217</td>
<td>0.01880</td>
<td>0.99955</td>
<td>0.04455</td>
<td>0.02088</td>
<td>0.99361</td>
<td>0.63923</td>
<td>0.02713</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PS</td>
<td>1.00000</td>
<td>0.99792</td>
<td>0.20794</td>
<td>0.01852</td>
<td>0.99797</td>
<td>0.20341</td>
<td>0.02430</td>
<td>0.99286</td>
<td>0.71360</td>
<td>0.03150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HS</td>
<td>1.00000</td>
<td>0.94509</td>
<td>5.49118</td>
<td>0.05854</td>
<td>0.92626</td>
<td>7.37393</td>
<td>0.07667</td>
<td>0.88194</td>
<td>11.80576</td>
<td>0.12055</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHS</td>
<td>1.00000</td>
<td>0.99809</td>
<td>0.19088</td>
<td>0.02094</td>
<td>0.96182</td>
<td>3.81821</td>
<td>0.05995</td>
<td>0.99245</td>
<td>0.75469</td>
<td>0.03258</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IR</td>
<td>1.00000</td>
<td>0.99819</td>
<td>0.18059</td>
<td>0.01882</td>
<td>0.99956</td>
<td>0.04385</td>
<td>0.02078</td>
<td>0.99363</td>
<td>0.63691</td>
<td>0.02698</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{4,ijkl}$</td>
<td>E</td>
<td>UP</td>
<td>1.00000</td>
<td>0.99818</td>
<td>0.18217</td>
<td>0.01880</td>
<td>0.99955</td>
<td>0.04455</td>
<td>0.02088</td>
<td>0.99361</td>
<td>0.63923</td>
<td>0.02713</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PS</td>
<td>1.00000</td>
<td>0.99792</td>
<td>0.20794</td>
<td>0.01852</td>
<td>0.99797</td>
<td>0.20341</td>
<td>0.02430</td>
<td>0.99286</td>
<td>0.71360</td>
<td>0.03150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HS</td>
<td>1.00000</td>
<td>0.94509</td>
<td>5.49118</td>
<td>0.05854</td>
<td>0.92626</td>
<td>7.37393</td>
<td>0.07667</td>
<td>0.88194</td>
<td>11.80576</td>
<td>0.12055</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHS</td>
<td>1.00000</td>
<td>0.99809</td>
<td>0.19088</td>
<td>0.02094</td>
<td>0.96182</td>
<td>3.81821</td>
<td>0.05995</td>
<td>0.99245</td>
<td>0.75469</td>
<td>0.03258</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI = Parameter of interest; WM = Weighting methods; TV = True value; AE = Average Estimate for parameter of interest; PB = Percent bias; RMSE = Root mean square error; UP = Unweighted parameter; PS = ED selection weight; HS = Household selection weight; PHS = ED and Household selection weight; IR = Item non-response weight
Table 5.10: Parameter estimates from weighted multilevel models using sampling selection weight adjusted for missing data (SWAM) from the Missing At Random missing mechanism with missing rates of 20%, 40% and 60% for data simulated from age-only, sex-only, sex-age, multivariable covariate models

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Weighting Methods</th>
<th>PI = Parameter of interest</th>
<th>WM = Weighting methods</th>
<th>TV = True value</th>
<th>AE = Average Estimate for parameter of interest</th>
<th>PB = Percent bias</th>
<th>RMSE = Root mean square error</th>
<th>20%</th>
<th>40%</th>
<th>60%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_{1,ijkl}</td>
<td>$\alpha_1$</td>
<td>SWAM</td>
<td>-0.08</td>
<td>-0.07993</td>
<td>0.0837</td>
<td>0.0019</td>
<td>-0.0801</td>
<td>0.05875</td>
<td>0.0736</td>
<td>-0.07984</td>
</tr>
<tr>
<td>Y_{2,ijkl}</td>
<td>$\beta_1$</td>
<td>SWAM</td>
<td>5</td>
<td>4.9953</td>
<td>0.0940</td>
<td>0.0000</td>
<td>4.9885</td>
<td>0.2296</td>
<td>0.0000</td>
<td>5.02054</td>
</tr>
<tr>
<td>Y_{3,ijkl}</td>
<td>$\lambda_2$</td>
<td>SWAM</td>
<td>-0.08</td>
<td>-0.0804</td>
<td>0.5375</td>
<td>0.0019</td>
<td>-0.0799</td>
<td>0.0125</td>
<td>0.0021</td>
<td>-0.07994</td>
</tr>
<tr>
<td>Y_{4,ijkl}</td>
<td>$\tau_1$</td>
<td>SWAM</td>
<td>-0.008</td>
<td>-0.0078</td>
<td>1.8350</td>
<td>0.0022</td>
<td>-0.0080</td>
<td>0.1644</td>
<td>0.0026</td>
<td>-0.00790</td>
</tr>
<tr>
<td></td>
<td>$\tau_3$</td>
<td>SWAM</td>
<td>-0.025</td>
<td>-0.01869</td>
<td><strong>25.2176</strong></td>
<td>0.0689</td>
<td>-0.0308</td>
<td><strong>23.3450</strong></td>
<td>0.0800</td>
<td>-0.01612</td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI = Parameter of interest; WM = Weighting methods; TV = True value; AE = Average Estimate for parameter of interest; PB = Percent bias; RMSE = Root mean square error; SWAM = Sampling Weight Adjusted for Missing Data
Table 5.11: Parameter estimates from weighted multilevel models using sampling weight adjusted for missing data (SWAM) from the Not Missing At Random missing mechanism with missing rates of 20%, 40% and 60% for data simulated from age-only, sex-only, sex-age, multivariable covariate models

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{1jkl}$</td>
<td>$\alpha_1$</td>
<td>SWAM</td>
<td>-0.08</td>
<td>-0.08008</td>
<td>0.0937</td>
<td>0.00173</td>
<td>-0.08008</td>
<td>0.0937</td>
<td>0.00221</td>
<td>-0.08003</td>
<td>0.0363</td>
<td>0.00276</td>
</tr>
<tr>
<td>$Y_{2jkl}$</td>
<td>$\beta_1$</td>
<td>SWAM</td>
<td>5</td>
<td>5.00366</td>
<td>0.0732</td>
<td>0.06502</td>
<td>5.00914</td>
<td>0.1828</td>
<td>0.0687</td>
<td>4.99833</td>
<td>0.0334</td>
<td>0.10171</td>
</tr>
<tr>
<td>$Y_{3jkl}$</td>
<td>$\lambda_1$</td>
<td>SWAM</td>
<td>5</td>
<td>5.00331</td>
<td>0.0662</td>
<td>0.06191</td>
<td>5.00607</td>
<td>0.1214</td>
<td>0.0714</td>
<td>4.99134</td>
<td>0.1732</td>
<td>0.09865</td>
</tr>
<tr>
<td>$Y_{4jkl}$</td>
<td>$\tau_1$</td>
<td>SWAM</td>
<td>-0.08</td>
<td>-0.08013</td>
<td>0.1587</td>
<td>0.00202</td>
<td>-0.07980</td>
<td>0.2550</td>
<td>0.0022</td>
<td>-0.08022</td>
<td>0.2687</td>
<td>0.00274</td>
</tr>
<tr>
<td>$Y_{4jkl}$</td>
<td>$\tau_2$</td>
<td>SWAM</td>
<td>-0.008</td>
<td>-0.01030</td>
<td>28.7500</td>
<td>0.04583</td>
<td>-0.01040</td>
<td>30.0000</td>
<td>0.0028</td>
<td>-0.00935</td>
<td>16.8750</td>
<td>0.00332</td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI = Parameter of interest; WM = Weighting methods; TV = True value; AE = Average Estimate for parameter of interest; PB = Percent bias; RMSE = Root mean square error; SWAM = Sampling Weight Adjusted for Missing Data
Table 5.12: Parameter estimates from weighted multilevel models using sampling selection weight adjusted for missing data (SWAM) from the Missing At Random missing mechanism with missing rates of 20%, 40% and 60% for data simulated from age-only, sex-only, sex-age, multivariable covariate models

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>20%</th>
<th>40%</th>
<th>60%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{1,kl}$</td>
<td>$\alpha_1$</td>
<td>SWAM</td>
<td>-0.08</td>
<td>-0.0799</td>
<td>0.1237</td>
<td>0.0020</td>
<td>-0.0800</td>
<td>0.0588</td>
<td>0.0023</td>
<td>-0.0799</td>
<td>0.1465</td>
<td>0.0030</td>
</tr>
<tr>
<td>$Y_{2,kl}$</td>
<td>$\beta_1$</td>
<td>SWAM</td>
<td>5</td>
<td>4.9952</td>
<td>0.0955</td>
<td>0.0678</td>
<td>4.9897</td>
<td>0.2056</td>
<td>0.0778</td>
<td>5.0205</td>
<td>0.4091</td>
<td>0.1053</td>
</tr>
<tr>
<td>$Y_{3,kl}$</td>
<td>$\lambda_2$</td>
<td>SWAM</td>
<td>5</td>
<td>5.0018</td>
<td>-0.0350</td>
<td>0.0863</td>
<td>5.0108</td>
<td>0.2168</td>
<td>0.0126</td>
<td>4.9960</td>
<td>0.0807</td>
<td>0.0004</td>
</tr>
<tr>
<td>$Y_{4,kl}$</td>
<td>$\tau_1$</td>
<td>SWAM</td>
<td>-0.08</td>
<td>-0.0804</td>
<td>0.5375</td>
<td>0.0019</td>
<td>-0.0800</td>
<td>0.0125</td>
<td>0.0020</td>
<td>-0.0799</td>
<td>0.0776</td>
<td>0.0030</td>
</tr>
<tr>
<td>$Y_{4,kl}$</td>
<td>$\tau_2$</td>
<td>SWAM</td>
<td>-0.008</td>
<td>-0.0080</td>
<td>0.3450</td>
<td>0.0023</td>
<td>-0.0081</td>
<td>0.7158</td>
<td>0.0028</td>
<td>-0.0075</td>
<td>6.2853</td>
<td>0.0033</td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI= Parameter of interest; WM =Weighting methods; TV = True value; AE=Average Estimate for parameter of interest; PB=Percent bias; RMSE=Root mean square error; SWAM=Sampling Weight Adjusted for Missing Data
Table 5.13: Parameter estimates from weighted multilevel models using sampling weight adjusted for missing data (SWAM) from the Missing NOT At Random missing mechanism with missing rates of 20%, 40% and 60% for data simulated from age-only, sex-only, sex-age, multivariable covariate models

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{1,ijkl}$</td>
<td>$\alpha_1$</td>
<td>SWAM</td>
<td>-0.08</td>
<td>-0.0801</td>
<td>0.1363</td>
<td>0.0801</td>
<td>-0.0794</td>
<td>0.7125</td>
<td>0.0022</td>
<td>-0.08003</td>
<td>0.0363</td>
<td>0.0801</td>
</tr>
<tr>
<td>$Y_{2,ijkl}$</td>
<td>$\beta_1$</td>
<td>SWAM</td>
<td>5</td>
<td>4.9982</td>
<td>0.0368</td>
<td>0.0644</td>
<td>5.0118</td>
<td>0.2352</td>
<td>0.0748</td>
<td>4.99833</td>
<td>0.0334</td>
<td>0.1017</td>
</tr>
<tr>
<td>$Y_{3,ijkl}$</td>
<td>$\lambda_1$</td>
<td>SWAM</td>
<td>5</td>
<td>4.9982</td>
<td>0.0368</td>
<td>0.0644</td>
<td>5.0118</td>
<td>0.2352</td>
<td>0.0748</td>
<td>4.99833</td>
<td>0.0334</td>
<td>0.1017</td>
</tr>
<tr>
<td>$Y_{4,ijkl}$</td>
<td>$\tau_1$</td>
<td>SWAM</td>
<td>-0.008</td>
<td>-0.0080</td>
<td>0.3450</td>
<td>0.0458</td>
<td>-0.0081</td>
<td>0.7158</td>
<td>0.0028</td>
<td>-0.00935</td>
<td>16.8750</td>
<td>0.0033</td>
</tr>
<tr>
<td>$Y_{4,ijkl}$</td>
<td>$\tau_3$</td>
<td>SWAM</td>
<td>-0.025</td>
<td>-0.0319</td>
<td>27.6117</td>
<td>0.0679</td>
<td>-0.0220</td>
<td>11.8142</td>
<td>0.0824</td>
<td>-0.02301</td>
<td>7.9766</td>
<td>0.1085</td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI= Parameter of interest; WM =Weighting methods; TV = True value; AE=Average Estimate for parameter of interest; PB=Percent bias; RMSE=Root mean square error; SWAFMD=Sampling Weight Adjusted for Missing Data.
Table 5.14: Parameter estimates from weighted multilevel models using sampling weight adjusted for missing data (SWAM) from the Not Missing At Random missing mechanism with missing rates of 20%, 40% and 60% for data simulated from age-only, sex-only, sex-age, multivariable covariate models.

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{1jkl}$</td>
<td>$\alpha_1$</td>
<td>SWAM</td>
<td>-0.08</td>
<td>-0.0799</td>
<td>0.1738</td>
<td>0.0019</td>
<td>-0.0800</td>
<td>0.0025</td>
<td>0.0022</td>
<td>-0.0799</td>
<td>0.1375</td>
<td>0.0029</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{2jkl}$</td>
<td>$\beta_1$</td>
<td>SWAM</td>
<td>5</td>
<td>4.9975</td>
<td>0.0508</td>
<td>0.0666</td>
<td>4.9890</td>
<td>0.2196</td>
<td>0.0713</td>
<td>5.0215</td>
<td>0.4300</td>
<td>0.1034</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{3jkl}$</td>
<td>$\lambda_2$</td>
<td>SWAM</td>
<td>5</td>
<td>5.0041</td>
<td>0.0816</td>
<td>0.0858</td>
<td>5.0077</td>
<td>0.1536</td>
<td>0.0133</td>
<td>4.9973</td>
<td>0.0550</td>
<td>0.0006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_{4jkl}$</td>
<td>$\tau_3$</td>
<td>SWAM</td>
<td>-0.08</td>
<td>-0.0805</td>
<td>0.5675</td>
<td>0.0019</td>
<td>-0.0800</td>
<td>0.0063</td>
<td>0.0021</td>
<td>-0.0799</td>
<td>0.1150</td>
<td>0.0029</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI = Parameter of interest; WM = Weighting methods; TV = True value; AE = Average Estimate for parameter of interest; PB = Percent bias; RMSE = Root mean square error; SWAFMD = Sampling Weight Adjusted for Missing Data.
Table 5.15: Parameter estimates from weighted multilevel models using sampling weight adjusted for missing data (SWAM) from the Not Missing At Random missing mechanism with missing rates of 20%, 40% and 60% for data simulated from age-only, sex-only, sex-age, multivariable covariate models.

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{1_{jkl}}$</td>
<td>$\alpha_1$</td>
<td>SWAM</td>
<td>-0.08</td>
<td>-0.0801</td>
<td>0.1075</td>
<td>0.0801</td>
<td>-0.0797</td>
<td>0.3725</td>
<td>0.0797</td>
<td>-0.0799</td>
<td>0.1137</td>
<td>0.0800</td>
</tr>
<tr>
<td>$Y_{2_{jkl}}$</td>
<td>$\beta_1$</td>
<td>SWAM</td>
<td>5</td>
<td>5.0040</td>
<td>0.0794</td>
<td>0.0658</td>
<td>5.0105</td>
<td>0.2090</td>
<td>0.0694</td>
<td>5.0016</td>
<td>0.0328</td>
<td>0.1047</td>
</tr>
<tr>
<td>$Y_{3_{jkl}}$</td>
<td>$\lambda_1$</td>
<td>SWAM</td>
<td>5</td>
<td>5.0032</td>
<td>0.0638</td>
<td>0.0610</td>
<td>5.0028</td>
<td>0.0552</td>
<td>0.0746</td>
<td>4.9940</td>
<td>0.1192</td>
<td>0.0974</td>
</tr>
<tr>
<td>$Y_{4_{jkl}}$</td>
<td>$\tau_1$</td>
<td>SWAM</td>
<td>-0.08</td>
<td>-0.0800</td>
<td>0.0613</td>
<td>0.0019</td>
<td>-0.0797</td>
<td>0.3475</td>
<td>0.0022</td>
<td>-0.0801</td>
<td>0.1662</td>
<td>0.0027</td>
</tr>
<tr>
<td>$Y_{5_{jkl}}$</td>
<td>$\tau_3$</td>
<td>SWAM</td>
<td>-0.008</td>
<td>-0.0079</td>
<td>0.8087</td>
<td>0.0434</td>
<td>-0.0080</td>
<td>0.4529</td>
<td>0.0027</td>
<td>-0.0094</td>
<td>16.8750</td>
<td>0.0033</td>
</tr>
<tr>
<td>$Y_{6_{jkl}}$</td>
<td>$\tau_3$</td>
<td>SWAM</td>
<td>-0.025</td>
<td>-0.0309</td>
<td>23.7023</td>
<td>0.0613</td>
<td>-0.0195</td>
<td>22.1277</td>
<td>0.0796</td>
<td>-0.0230</td>
<td>7.9766</td>
<td>0.1060</td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI = Parameter of interest; WM = Weighting methods; TV = True value; AE = Average Estimate for parameter of interest; PB = Percent bias; RMSE = Root mean square error; SWAM = Sampling Weight Adjusted for Missing Data
Chapter 6

Comparative Analysis of Survey Data with emphasis on Item non-response Weighting Method

Chapter Outline

The aim of this thesis is to assess weighting methods from different scenarios in multilevel modelling of a continuous outcome with missing data. Sequel to Chapter 6, five weighting adjustment methods were assessed. In Chapter 6, the focus is on the application of item non-response weight to the multilevel modelling of annual household expenditure of food purchased away from home in Jamaica and also the modelling of individual reported income in China.

6.1 Aim and Scope of the Analysis

Statistical analysis and methodology involved in the analysis of complex and large survey data are a challenge to data analysts, especially as they relate to datasets from non-uniform sampling technique. The creation of models and the development of interpretations from hierarchical survey datasets continue to generate debate (Maas & Hox, 2004). Accordingly, the aim of Chapter 6 was to develop a series of weighted multilevel models for continuous outcome variables with missing data from multistage survey datasets, using the item non-response weight in the analysis. In these analyses, the lessons learnt in Chapter 5 from the simulation study regarding the construction and incorporation of item non-response weights are used. Specifically, the item non-response weight will be applied at different stages of the model to illustrate its effects on the model-selection processes. The Jamaica Survey of Living Conditions 2007 (JSLC 2007) data is used in the first analysis while the China Health and Nutrition Survey (CHNS) data for 1989 and 2011 is used in the second analysis. Using the JSLC 2007, the analysis focused on
the reported household annual expenditure of meals purchased away from home while the analysis for the CHNS datasets focused on the reported individual income.

The set of analyses in this chapter stem from the simulation study conducted in Chapter 5 on the reliability of weighting methodologies for survey data with missing data from a multistage survey design. The findings from the simulation study in Chapter 5 confirmed that the item non-response weight was the most reliable under different proportions of missing data, hence the application to the modelling of food-related expenditure and income variables.

6.2 Objectives and Rationale for the Analysis

The first objective of the analysis was the creation of two- and three-level random coefficient weighted models. The focus of these models was the annual household expenditure data for meals purchased away from home in the JSCLC 2007. The second objective of the analysis involved the creation of a four-level weighted model at the province, stratum, community, and household levels for the reported individual income data in the CHNS for the years 1989 and 2011.

The rationale for these analyses were three-fold. The first rationale stemmed from the findings of the simulation study in Chapter 5, while the second rationale was to investigate the size of the intraclass correlation and the need for multilevel model. Huang (2018) affirmed that it is a myth to state that a multilevel model would be necessary on the basis of the intraclass correlation. This misconception will be investigated by varying the analysis using different models to highlight different intraclass values and estimated parameters. The third rationale for the analyses in this Chapter is to compare weighted parameter estimates from different multilevel models when estimates are adjusted for item non-response. Furthermore, the gains of using item non-response weight as an adjustment method to compensate for missing data when developing a multilevel model will be elaborated upon.
6.3 Multilevel Models in the Analysis with 14 Parishes

The proposed study models were multilevel as they were developed to fully incorporate the hierarchical structures in the survey design, and also to achieve the chapter’s objectives. The method of parameter estimation is described in Chapter 4.

6.3.1 JSLC2007 Multilevel Models

Model II-A

Two multilevel model scenarios were investigated in this analysis. The first scenario involved a series of two-level models consisting of Parish and Household levels with the application of the item non-response weight at the Parish level. I started the analysis with the random effect as expressed in 6.1, then followed by the fixed effect version with the addition of the Parish as predictor variable in 6.1.1. These models are labelled as Model II-A in the illustrations in Figure 1 as well as in the result in Table 6.3.

\[
y_{ji} = x_{ji}^T \beta + u_j + \varepsilon_{ji} \quad (6.1)
\]

\[
y_{ji} = x_{ji}^T \beta + d_j + \varepsilon_{ji} \quad (6.1.1)
\]

In both models, \( y_{ji} \) is the dependent variable, the log of the annual household expenditure of meals purchased away from home for the \( i^{th} \) household in the \( j^{th} \) Parish. In model 6.1, the random and error terms are represented by \( u_j \sim N(0, \sigma^2_j) \), \( \varepsilon_{ji} \sim N(0, \sigma^2_{ji}) \) respectively while in the 6.1.1 represented fixed term for the Parish was represented by \( d_j \). The \( x_{ji}^T \) denoted the vector of subject level variables from the socio-demographic (Age, Household Size, Sex, Education, Employment Status, and Occupation Status). The result of the two-level model series was denoted as model II-A, and are index from A1 to A8 in Table 6.3 followed by the weighted fixed effect model with the inclusion of the parish as one of the explanatory variables. The result of this models respectively and illustrated in Table 6.3.1. denoted Mode II-A9.1 and
Model II-A9.2 respectively to highlight the building process and final models. Further collapsing of the parish into two categories resulted in results in Table 6.3.1.1.

In the parameter estimation, $w_{ji}$ is incorporated as the item non-response weight from the response probability of the household and average per Parish representing non-response for each individual in the Parish. The incorporation is done by weighting the parameter estimator by $w_{ji}$ to adjust for non-response. Figure 6.1 illustrates the two-level model denoted Model II-A.

![Figure 6.1. Model II-A](image)

Model II-B

In continuation of the two-level model, an alternative two-level model consisting of the Enumeration District (ED), the Primary Sampling Unit (PSU) and Household was also developed for comparison purposes. There are 168 PSUs which was too many to consider for a fixed effect only model. The random effect model is stated in 6.2.

$$y_{ki} = x_{ki}^T \beta + v_k + \epsilon_{ki}$$  \hspace{1cm} (6.2)
In this model $y_{ki}$ represent the log of the annual household expenditure of meals purchased away from home for the $i^{th}$ household in the $k^{th}$ PSU. In this model, $v_k \sim N(0, \sigma_k^2)$ and $\varepsilon_{ki} \sim N(0, \sigma_{ki}^2)$ represent the random components at the ED and household levels, respectively. In this model, $W_{ki}^2$ represented the item non-response weight from the response probability of the household and average per ED unlike the earlier model in 6.1 which is average per Parish. This weight is used to adjust the estimates for item non-response for each individual at the ED level. In this model II-B and illustrated in Figure 6. 2.

**Figure 6. 2. Model II-B**

![Model II-B Diagram]

**Model -III A**

The second scenario involves the development of the final model in the JSJC 2007 analysis, a three-level model to draw comparisons with the two-level models in the first scenario. This model incorporated the Parish, ED, and Household levels while using the item non-response weight previously averaged at the ED level. The three-level model represented as:

$$y_{jki} = x_{jki}^T \beta + u_j + v_{jk} + \varepsilon_{jki}$$

(5.3)
where \( u_j \sim N(0, \sigma^2_j) \), \( v_{jk} \sim N(0, \sigma^2_{jk}) \) and \( e_{jki} \sim N(0, \sigma^2_{jki}) \) represents the random component at the Parish, ED, and household level respectively. The model is illustrated in Figure 3.

**Figure 6.3. Model III-A**

**LEVEL 3-PARISH**  \( j=1, 2, 3, \ldots, 14 \)

**LEVEL 2 – ED / PSU**

\( k=1, 2, 3, 4, \ldots, 168 \)

**LEVEL 1 – HOUSEHOLD**

\( i=1, 2, 3, 4, 5, 6, \ldots, 1994 \)

### 6.3.2 Multilevel Model for the China Health and Nutrition Survey (CHNS)

For the CHNS analyses, two separate multilevel models are formulated for comparative analysis. A four – level model was developed for each Year 1989 and 2011 respectively to illustrate the hierarchical structure of the data in the unweighted and weighted models. Consequently, the process was repeated with a three – level models each year for comparative analysis to determine if the use of three level will cause a change in the predictor variables and parameter estimates.
Four Level Model IV

The four-level model is grounded on the basis that the household is nested within the Neighborhood-Village which is nested within county-city and nested within the province.

The four-level model for the CHNS 1989 and 2011 is illustrated as follows:

Let $j$ represent the provinces in each model, the highest level. In 1989, data was collected from 8 Provinces compared with the 12 in the 2011 data set. The provinces are the highest level in the model. Hence for every $j$, there are $k$ nested County/City further nested by $l$ Neighbourhood-Village and $m$ individuals at the lowest level. (Zhang et al., 2014).

In the four-level model IV, the outcome variable $y_{jil}$ was defined as the reported total income from $i^{th}$ household nested within $l^{th}$ Neighbourhood-Village then nested within $k^{th}$ County-City then nested within the $j^{th}$ Province in model IV

$$y_{jikl} = X^T_{jikl} \beta + u_j + v_{jk} + \tau_{jkl} + \epsilon_{jkl}$$  \hspace{1cm} (6.4)

The $X^T$ is the vector variables for age, gender, marital status, and completed years of formal education. The notation: $u_j \sim N(0, \sigma_j^2)$, $v_{jk} \sim N(0, \sigma_{jk}^2)$, $\tau_{jkl} \sim N(0, \sigma_{jkl}^2)$, and $\epsilon_{jkl} \sim N(0, \sigma_{jkl}^2)$ represents the random components at the Province, County-City, Neighbourhood-Village, and household, respectively as illustrated in Figure 6.4 Model IV

Figure 6.4. Model IV-A
Three -Level Model III for the Year 1989 and 2011

In the three -level model, the Neighborhood -Village is nested within the province. The three -level model for the CHNS 1989 & 2011 is illustrated as follows:

Let $j$ represent the provinces for $j = 1, 2, 3, \ldots, 12$, for the twelve Provinces for the Year 2011 survey. Within these Provinces, $l$, Neighborhood-Village is selected for $l = 1, 2, 3, \ldots, 161$, consisting of $m$ individuals with $m = 1, 2, 3, \ldots, 2109$. In the three-level model, $y_{jm}$, the outcome variable reported income from $m^{th}$ household nested within $l^{th}$ Neighborhood-Village nested within the $j^{th}$ Province is presented as follows:

$$y_{jl} = x_{jl}^T \beta + u_j + v_{jl} + \varepsilon_{jl}$$  \hspace{1cm} (6.5)

The $x^T$ is the vector variables for age, gender, marital status, and completed years of formal education. The notation: $u_j \sim N(0, \sigma_j^2)$, $v_{jl} \sim N(0, \sigma_{jl}^2)$ and $\varepsilon_{jl} \sim N(0, \sigma_{jl}^2)$ represents the random components at the Province, Neighborhood-Village, and household, respectively as illustrated in Figure 6.5: Model III

---

**Figure 6.5: Model III**

[Diagram of three-level model showing levels and nested relationships]
6.4 Estimating Response Probabilities for each Data Set and Item
Non-Response Weight

In the development of the item non-response weight, the response probability
via logistic regression was used to determine the probability of every
individual’s likelihood of responding to the outcome variable. Dehija and
Wahba (1999) and Kalton and Flores-Cervantes (2003) applied logistic
regression in the estimation of response probability and also used the inverse
of the probability to create item non-response weight. In these analyses, the
notation and context of the logistic regression is described in the section
below:

For each dataset let \( R_i \) be the missing indicator variables for every \( i \)
individual, and also \( R_i \) is a binary variable containing: 1 for missing, 0 for not
missing. Similarly, socio-demographic variables of the individuals are
represented by \( X_i \), a vector of all the independent variables under the
assumptions of full data.

The regression model used in the study and in Bethlehem et al. (2011):

\[
\log \left( \frac{R_i}{1-R_i} \right) = \logit(p(X_i)) = X_i' \beta 
\]

(6.6)

the final model is determined for only significant (p-value less than 0.05)
variables are included in the model. The response probability, \( \hat{p}_i(R) \) is
finally estimated as follows:

\[
\hat{p}_i(R) = \frac{\exp(X'_i \beta)}{1 + \exp(X'_i \beta)} 
\]

(6.7)

In the model, the inverse of \( \hat{p}_i(R) \) is used to determine the weight for all
the models.
6.5 Intraclass Correlation Estimation

The Intraclass Correlation (ICC) was denoted as $\rho$ and defined as the measure of the proportion of variation in the outcome variable that occurred between groups in relation to the total variation present (Finch et al., 2014). In these analyses, the proportion of variance at each level of the data hierarchy were estimated to investigate underlying relationship between the selected model and the intraclass estimated correlations. Below is an illustration of the estimation of ICC for a three-level model with variances at household ($\sigma^2_e$), ED ($\sigma^2_v$) and Parish ($\sigma^2_u$) levels:

ICC for the household level variation:

$$\rho_e = \frac{\sigma^2_e}{\sigma^2_u + \sigma^2_v + \sigma^2_e}$$

(6.8)

ICC for the PSU level variation:

$$\rho_v = \frac{\sigma^2_v}{\sigma^2_u + \sigma^2_v + \sigma^2_e}$$

(6.9)

ICC for the Parish level variation:

$$\rho_u = \frac{\sigma^2_u}{\sigma^2_u + \sigma^2_v + \sigma^2_e}$$

(6.10)

6.6 Model Building and Selection Approach

The models in these analyses were created using a forward selection approach while noting the level of significance as the variables were added to the model. The baseline for adding a variable was a $p$-value less than 0.05. Any variable with a $p$-value > 0.05 was excluded from the model. Additionally, the effect of the weights with regards to the Akaike Information Criterion (AIC) from the pseudo-maximum likelihoods were also observed. Hirotugu Akaike(1978) introduced an information criterion which is defined as:

$$\text{AIC} = (-2) \log(\text{maximum likelihood}) + 2(\text{number of parameters})$$
6.7 Analysis of the JSLC 2007

The JSLC 2007 was conducted from May 1 to July 31, 2007 as part of a longitudinal study on the social conditions of the Jamaican population as explained in Chapter 3. The survey design employed multistage sampling for the sample selection. The island of Jamaica is divided into 14 Parishes with 5,235 Enumeration Districts (EDs) proportional to the size of the population in each Parish. In the JSLC 2007, 168 EDs were sampled from the available 5,235 EDs. The size of the sampled EDs per parish ranged from 6 for the smaller to 36 for the larger parishes. Table 5.1 contains the sampled EDs per parish, for example, in the parish of Kingston, 6 EDs or Primary Sampling Units were sampled out of the 224 Enumeration Districts. In the survey design, 16 households were sampled in each of the EDs with the expectation that the total responses would be 2,688 households from the sampled units. However, at the end of the survey, only 1,994 households responded, of which 1,653 households responded to the outcome variable. The minimum number of households in a given ED was 3 while the maximum was 16 thus affirming the contrast between the expected and actual sample size. The Unit non-response adjustment is not considered in this study because there is no availability of information on the sampled households who did not respond during the survey and adjusted such as the post-stratification method is not the focus of this study. In addition, the focus of this Chapter is the application of the item non-response weight which was identified as the least bias in the simulation study.

Table 6.1: Distribution of the sampled EDs per Parish

<table>
<thead>
<tr>
<th>PARISH NUMBER</th>
<th>PARISH</th>
<th>SAMPLE PSU</th>
<th>TOTAL PSU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>KINGSTON</td>
<td>6</td>
<td>224</td>
</tr>
<tr>
<td>2</td>
<td>ST. ANDREW</td>
<td>36</td>
<td>973</td>
</tr>
<tr>
<td>3</td>
<td>ST. THOMAS</td>
<td>6</td>
<td>224</td>
</tr>
<tr>
<td>4</td>
<td>PORTLAND</td>
<td>6</td>
<td>187</td>
</tr>
<tr>
<td>5</td>
<td>ST. MARY</td>
<td>6</td>
<td>278</td>
</tr>
<tr>
<td>6</td>
<td>ST. ANN</td>
<td>10</td>
<td>319</td>
</tr>
<tr>
<td>7</td>
<td>TRELAWNY</td>
<td>6</td>
<td>178</td>
</tr>
<tr>
<td>8</td>
<td>ST. JAMES</td>
<td>10</td>
<td>347</td>
</tr>
<tr>
<td>9</td>
<td>HANOVER</td>
<td>6</td>
<td>159</td>
</tr>
<tr>
<td>10</td>
<td>WESTMORELAND</td>
<td>8</td>
<td>322</td>
</tr>
<tr>
<td>11</td>
<td>ST. ELIZABETH</td>
<td>10</td>
<td>308</td>
</tr>
<tr>
<td>12</td>
<td>MANCHESTER</td>
<td>12</td>
<td>343</td>
</tr>
<tr>
<td>13</td>
<td>CLARENDON</td>
<td>14</td>
<td>467</td>
</tr>
<tr>
<td>14</td>
<td>ST. CATHERINE</td>
<td>32</td>
<td>906</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>168</td>
<td>5,235</td>
</tr>
</tbody>
</table>
The motivation for the selected variable in the JSLC 2007 analysis was based on the study from Manrique and Jensen (1998). Prior exploratory analysis of the relationship between the socio-demographics and the outcome variable. The outcome variable in the JSLC 2007 had 17.35% missing of the 1994. Manrique and Jensen (1998) reported that food away from home was a function of time, wage, and vector of socio-demographic variables such as age, gender, race, and household size. Similar variables were postulated by Bai et al. (2010), Lee (2006), McCracken and Brandt (1987), and Ma et al. (2006). The covariates in this analysis were extracted from the socio-demographic variables for each head of household in the JSLC 2007. The variables selected included age, household size, sex, level of education, employment, and occupational statuses, respectively. The average age of the household heads was 48.97 years with standard deviation of 16.73 years. The youngest head of household was aged 16 while the oldest was aged 99. Household size ranged from 1 to 18 members with an average of 3.4 members.

In the JSCL 2007, there were more male-headed households (53.66%; n=1,070) than female-headed households (46.34%; n=924). For the analysis, the education variable was collapsed into two categories: educated (8.81%; n=176) and neither stated nor had a formal education (91.17%; n=1,818). Three occupation categories were used in the model development: Office Related Work; Non-Office Related Work and Non-Classified (ORW=14.74%, n=294; NORW=69.45%, n=1385; NC=15.80%, n=315) and two employment categories(unemployed and not classified =15.59%; n=311; employed= 84.40%; n=1,683).

Prior to the weighted model, the item non-response was derived as the inverse of the response probability as a function of age of the respondents. This was based on the result of the logistic regression model for the JSLC 2007. The analysis revealed that of all the socio-demographic variables, age was found to be a significant (p<0.05) predictor of the missing indicator variable.
6.7.1 Result of the JSLC2007 Multilevel Model Analysis

In the JSLC 2007 analysis, two separate weighted two-level models were investigated. Model II-A seeks to illustrate the effects of the item non-response weight on the parameter estimate when a two-level model only consists of the Parish and the household level random slope. In this model, the minimum item non-response weight for a Parish was 1.184 while the maximum was 1.252. Similarly, the sample size per parish ranges from 65 to 404. The item non-response weight in this scenario represents the weighted estimate at the Parish level which manifested in the estimated parameters in Models II-A1 to Model II-A8 in Table 6.3. A continual decrease of the AIC values was observed as more variables are added to the model, until the final and best Model II-A6 consisting of age, household size level of education and occupation as the significant predictor variables (p< 0.05).

In order to make comparison with the first-set of analysis involving the Parish as the second and highest level in ModelII-A1 to ModelII-A8. Another two - level ModelIII-B1 to ModelIII-B8 having the ED as the highest level instead of Parish were also developed. There are 168 Enumeration Districts in the model compared to the earlier model with 14 Parishes. In these Models, the item non-response weight constant for the ED as a weight to adjust for missing data from household heads in each of the selected EDs. In this scenario, the average item weight per PSU ranges from 1.109 to 1.357 while the sample size per ED also ranges from 5 to 18. These results show that for small sample size, the average weight per ED were higher than the average weight per Parish with larger sample. This phenomenon account for the estimates in Tables 6.3 and 6.4 respectively.

In ModelIII-A9.1, ModelIII-A9.2, ModelIII-A9.1.1 & ModelIII-A9.2.2 the Parish was added as a fixed effect, similar results were observed with a lower AIC value as demonstrated in Table 6.3.1. This suggested that the addition of the Parish did not affect parameter estimates significantly. However, the results revealed that the parameter estimates for the parishes in some less developed parishes especially those with higher number of rural areas have reduced annual expenditure of meals purchased away from home. Some of the parishes with lower annual expenditure on meals away from home are contiguous. The parishes with the predominantly rural areas include St. Thomas, Portland, St. Mary, St. Ann, Trelawny, Hanover, Westmoreland, St.
Elizabeth and Clarendon, while the parishes with more urban areas includes St. Andrew, St. James, Manchester and St. Catherine. These parishes have a greater annual household expenditure of meals purchased away from home. These parishes have some common qualities especially Kingston and St. Andrew which is the commercial hub of the island with the highest concentration of the tertiary institutions. The parish of St. Catherine is in close proximity to these two Parishes. St. James is the tourism hub of the island which also contributed to the development of the parish. Similar qualities exist in the parish of Manchester which has a university and opportunity for employment in small and medium scale industries as illustrated in Figure 6.

Figure 6.: Illustration of the parameter estimates from the fixed effect model

The parishes in grey while the baseline is Kingston in red below St. Andrew in Figure 6 have higher concentration of fast-food restaurants when compared to the more rural dominated parishes in the eastern and western end.. This clearly suggest that the tendency of the household members to eat away from home will be higher for persons who reside in the parishes with higher urban areas than those who reside in more rural areas of the island. Further merging of these parishes with negative parameter estimates as a group and those with positive parameter estimates clearly revealed the differences in the parameter estimates in the fixed effects in Table 6.4.

Unlike the estimates in Table 6.3, the parameter estimates for the ED scenario were found to be higher as demonstrated in Table 6.6. Similar variables such
as age, household size, level of education and occupation were the significant
predictor variable in the best model II-B8. In the three-level model involving
EDs nesting in the Parish. The lessons learnt from these analyses, is that
sample size determines the magnitude of the weight per unit (Parish or ED)
because the sizes of each cluster (Parish or ED). Also, the magnitude of the
parameter estimates is proportional to the size of the weight. Additionally,
level at which the weight is applied is more relevant in the parameter estimate
than the hierarchical levels in the model.

Table 6.2 also revealed that the majority of the proportion of the variation in
the outcome variable were found in the household level of the models.
However, as the model level increased from two to three, the ICC values
increased for the Parish and ED while decreasing for the household level.

Table 6.2: Estimated ICC Values for the Levels in the Models in the JSCLC 2007

<table>
<thead>
<tr>
<th>Model</th>
<th>Variance estimates and ICC at each level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model II-A</strong></td>
<td></td>
</tr>
<tr>
<td>( y_{jl} = x_{jl}^T \beta + u_j + \epsilon_{jl} ) , for ( \sigma_{u_j}^2 ) and ( \sigma_{\epsilon_j}^2 ) represents Parish and Household variances respectively</td>
<td>( \sigma_{u}^2 = 0.039 )</td>
</tr>
<tr>
<td></td>
<td>( \sigma_{\epsilon}^2 = 0.868 )</td>
</tr>
<tr>
<td></td>
<td>( \rho_{\text{PARISH}} = 0.043 )</td>
</tr>
<tr>
<td></td>
<td>( \rho_{\text{HOUSEHOLD}} = 0.957 )</td>
</tr>
<tr>
<td><strong>Model II-B</strong></td>
<td></td>
</tr>
<tr>
<td>( y_{kl} = x_{kl}^T \beta + v_k + \epsilon_{kl} ), for ( \sigma_{v_k}^2 ) and ( \sigma_{\epsilon_k}^2 ) represents ED and household variances respectively</td>
<td>( \sigma_{v}^2 = 0.121 )</td>
</tr>
<tr>
<td></td>
<td>( \sigma_{\epsilon}^2 = 0.795 )</td>
</tr>
<tr>
<td></td>
<td>( \rho_{\text{ED}} = 0.132 )</td>
</tr>
<tr>
<td></td>
<td>( \rho_{\text{HOUSEHOLD}} = 0.868 )</td>
</tr>
</tbody>
</table>
**Model III-A**

\[ y_{jkl} = x^T_{jkl} \beta + u_j + v_{jk} + \varepsilon_{jkl} \]

for \( \sigma^2_{u,j} \), \( \sigma^2_{v,j} \) and \( \sigma^2_{\varepsilon_{jkl}} \) represents Parish, ED and Household variances respectively.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma^2_u )</td>
<td>0.035</td>
</tr>
<tr>
<td>( \sigma^2_v )</td>
<td>0.080</td>
</tr>
<tr>
<td>( \sigma^2_\varepsilon )</td>
<td>0.795</td>
</tr>
<tr>
<td>( \rho_{\text{PARISH}} )</td>
<td>0.039</td>
</tr>
<tr>
<td>( \rho_{\text{ED}} )</td>
<td>0.088</td>
</tr>
<tr>
<td>( \rho_{\text{HOUSEHOLD}} )</td>
<td>0.873</td>
</tr>
</tbody>
</table>
Table 6.3: Parameter Estimates for Two-Level Random Coefficient Model using Item Non-response Weight Average at the Parish level

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model II-A1</th>
<th>Model II-A2</th>
<th>Model II-A3</th>
<th>Model II-A4</th>
<th>Model II-A5</th>
<th>Model II-A6</th>
<th>Model II-A7</th>
<th>Model II-A8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0609)</td>
<td>(0.07916)</td>
<td>(0.1066)</td>
<td>(0.1102)</td>
<td>(0.1066)</td>
<td>(0.1318)</td>
<td>(0.1256)</td>
<td>(0.1162)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0070</td>
<td>-0.0091</td>
<td>-0.0092</td>
<td>-0.0084</td>
<td>-0.0071</td>
<td>-0.0079</td>
<td>-0.0078</td>
<td>-0.0078</td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td>(0.0012)</td>
<td>(0.0012)</td>
<td>(0.00112)</td>
<td>(0.0014)</td>
<td>(0.0014)</td>
<td>(0.0014)</td>
<td>(0.0014)</td>
</tr>
<tr>
<td>Household Size</td>
<td>0.1221</td>
<td>0.1222</td>
<td>0.1232</td>
<td>0.1156</td>
<td>0.1153</td>
<td>0.1154</td>
<td>0.1154</td>
<td>0.1154</td>
</tr>
<tr>
<td></td>
<td>(0.0134)</td>
<td>(0.0133)</td>
<td>(0.1344)</td>
<td>(0.0144)</td>
<td>(0.0148)</td>
<td>(0.0149)</td>
<td>(0.0149)</td>
<td>(0.0149)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-0.0183</td>
<td>-0.0172</td>
<td>-0.0249</td>
<td>-0.0249</td>
<td>-0.0187</td>
<td>-0.0187</td>
<td>-0.0187</td>
<td>-0.0187</td>
</tr>
<tr>
<td>Male -Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educated</td>
<td>0.3732</td>
<td>0.3215</td>
<td>0.2486</td>
<td>0.2375</td>
<td>0.2475</td>
<td>0.2475</td>
<td>0.2475</td>
<td>0.2475</td>
</tr>
<tr>
<td></td>
<td>(0.0473)</td>
<td>(0.0426)</td>
<td>(0.0353)</td>
<td>(0.0361)</td>
<td>(0.0361)</td>
<td>(0.0361)</td>
<td>(0.0361)</td>
<td>(0.0361)</td>
</tr>
<tr>
<td>Not Stated – Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>0.4169</td>
<td>0.4169</td>
<td>0.2020</td>
<td>0.2020</td>
<td>0.2020</td>
<td>0.2020</td>
<td>0.2020</td>
<td>0.2020</td>
</tr>
<tr>
<td></td>
<td>(0.0869)</td>
<td>(0.0869)</td>
<td>(0.1137)</td>
<td>(0.1137)</td>
<td>(0.1137)</td>
<td>(0.1137)</td>
<td>(0.1137)</td>
<td>(0.1137)</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Office Related Work (ORW)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Office Related Work (NORW)</td>
<td>0.84020</td>
<td>0.6512</td>
<td>0.4195</td>
<td>0.4195</td>
<td>0.4195</td>
<td>0.4195</td>
<td>0.4195</td>
<td>0.4195</td>
</tr>
<tr>
<td></td>
<td>(0.0829)</td>
<td>(0.0787)</td>
<td>(0.0929)</td>
<td>(0.0929)</td>
<td>(0.0929)</td>
<td>(0.0929)</td>
<td>(0.0929)</td>
<td>(0.0929)</td>
</tr>
<tr>
<td>Not -Classified -Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{\text{PARISH}}$</td>
<td>0.038730</td>
<td>0.035872</td>
<td>0.034857</td>
<td>0.034782</td>
<td>0.029207</td>
<td>0.02955</td>
<td>0.02739</td>
<td>0.027889</td>
</tr>
<tr>
<td>$\sigma^2_{\text{HOUSEHOLD}}$</td>
<td>0.868997</td>
<td>0.857106</td>
<td>0.786592</td>
<td>0.786592</td>
<td>0.775809</td>
<td>0.758467</td>
<td>0.747706</td>
<td>0.748052</td>
</tr>
<tr>
<td>AIC</td>
<td>5416.97</td>
<td>5390.78</td>
<td>5222.07</td>
<td>5223.86</td>
<td>5196.44</td>
<td>5153.99</td>
<td>5128.59</td>
<td>5125.97</td>
</tr>
</tbody>
</table>

PE=Parameter Estimates, SE = Standard Error, AIC = Akaike Information Criterion
Table 6.4: Parameter Estimates for Two-Level Random Coefficient Model using Item Non-response Weight Average at the Parish level incorporating Parish as a fixed effect variable

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>(S.E)</th>
<th>p-value</th>
<th>Parameter Estimate</th>
<th>(S.E)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>10.5646</td>
<td>(0.1082)</td>
<td>0.0000</td>
<td>10.5469</td>
<td>(0.0988)</td>
<td>0.0000</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0078</td>
<td>(0.0014)</td>
<td>0.0000</td>
<td>-0.0077</td>
<td>(0.0014)</td>
<td>0.0000</td>
</tr>
<tr>
<td>Household Size</td>
<td>0.1154</td>
<td>(0.0148)</td>
<td>0.0000</td>
<td>0.1153</td>
<td>(0.0149)</td>
<td>0.0000</td>
</tr>
<tr>
<td>Sex</td>
<td>Male-Ref</td>
<td>-0.0169</td>
<td>(0.0339)</td>
<td>0.6180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>Educated</td>
<td>0.2342</td>
<td>(0.0367)</td>
<td>0.0000</td>
<td>0.2333</td>
<td>(0.0374)</td>
</tr>
<tr>
<td>Employment Status</td>
<td>Employed</td>
<td>-0.1832</td>
<td>(0.1175)</td>
<td>0.1190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td>ORW</td>
<td>0.8196</td>
<td>(0.0848)</td>
<td>0.0000</td>
<td>0.6484</td>
<td>(0.0794)</td>
</tr>
<tr>
<td></td>
<td>NORW</td>
<td>0.5919</td>
<td>(0.1232)</td>
<td>0.0000</td>
<td>0.4191</td>
<td>(0.0948)</td>
</tr>
<tr>
<td>Parish</td>
<td>St. Andrew -2</td>
<td>0.2232</td>
<td>(0.0087)</td>
<td>0.0000</td>
<td>0.2201</td>
<td>(0.0082)</td>
</tr>
<tr>
<td></td>
<td>St. Thomas -3</td>
<td>-0.3789</td>
<td>(0.0134)</td>
<td>0.0000</td>
<td>-0.3879</td>
<td>(0.0132)</td>
</tr>
<tr>
<td></td>
<td>Portland – 4</td>
<td>-0.3703</td>
<td>(0.0133)</td>
<td>0.0000</td>
<td>-0.3758</td>
<td>(0.0113)</td>
</tr>
<tr>
<td></td>
<td>St. Mary -5</td>
<td>-0.2459</td>
<td>(0.0116)</td>
<td>0.0000</td>
<td>-0.2493</td>
<td>(0.0106)</td>
</tr>
<tr>
<td></td>
<td>St. Ann – 6</td>
<td>-0.2278</td>
<td>(0.0201)</td>
<td>0.0000</td>
<td>-0.2389</td>
<td>(0.0152)</td>
</tr>
<tr>
<td></td>
<td>Trewlany -7</td>
<td>-0.2990</td>
<td>(0.0149)</td>
<td>0.0000</td>
<td>-0.3068</td>
<td>(0.0118)</td>
</tr>
<tr>
<td></td>
<td>St. James -8</td>
<td>0.1026</td>
<td>(0.0116)</td>
<td>0.0000</td>
<td>0.0978</td>
<td>(0.0094)</td>
</tr>
<tr>
<td></td>
<td>Hanover – 9</td>
<td>-0.1458</td>
<td>(0.0153)</td>
<td>0.0000</td>
<td>-0.1500</td>
<td>(0.0137)</td>
</tr>
<tr>
<td></td>
<td>Westmoreland -10</td>
<td>-0.1345</td>
<td>(0.0120)</td>
<td>0.0000</td>
<td>-0.1371</td>
<td>(0.0110)</td>
</tr>
<tr>
<td></td>
<td>St. Elizabeth -11</td>
<td>-0.1326</td>
<td>(0.0137)</td>
<td>0.0000</td>
<td>-0.1366</td>
<td>(0.0122)</td>
</tr>
<tr>
<td></td>
<td>Manchester – 12</td>
<td>0.0632</td>
<td>0.0174</td>
<td>0.0000</td>
<td>0.0592</td>
<td>(0.0160)</td>
</tr>
<tr>
<td></td>
<td>Clarendon -13</td>
<td>-0.2957</td>
<td>(0.0142)</td>
<td>0.0000</td>
<td>-0.3015</td>
<td>(0.0116)</td>
</tr>
<tr>
<td></td>
<td>St. Catherine -14</td>
<td>0.0625</td>
<td>0.0152</td>
<td>0.0000</td>
<td>0.0572</td>
<td>(0.0131)</td>
</tr>
<tr>
<td>Parish -1 -Ref</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>σ²_PARISH</td>
<td>0.0000</td>
<td></td>
<td></td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ²_HOUSEHOLD</td>
<td>0.7421</td>
<td></td>
<td></td>
<td>0.7417</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>5084.1</td>
<td></td>
<td></td>
<td>5081.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6.5: Parameter Estimates for Two-Level Random Coefficient Model using Item Non-response Weight Average at the Parish level incorporating Parish as a fixed effect variable

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model II-A9.1.1</th>
<th></th>
<th></th>
<th>Model II-A9.2.2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter</td>
<td>(S.E)</td>
<td>p-value</td>
<td>Parameter</td>
<td>(S.E)</td>
<td>p-value</td>
</tr>
<tr>
<td>Constant</td>
<td>10.5744</td>
<td>0.1124</td>
<td>0.0000</td>
<td>10.5543</td>
<td>0.1032</td>
<td>0.0000</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0080</td>
<td>0.0015</td>
<td>0.0000</td>
<td>-0.0078</td>
<td>0.0015</td>
<td>0.0000</td>
</tr>
<tr>
<td>Household Size</td>
<td>0.1145</td>
<td>0.0336</td>
<td>0.0000</td>
<td>0.1144</td>
<td>0.0155</td>
<td>0.0000</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-0.0181</td>
<td>0.0336</td>
<td>0.5880</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educated</td>
<td>0.2453</td>
<td>0.0282</td>
<td>0.0000</td>
<td>0.2443</td>
<td>0.0289</td>
<td>0.0000</td>
</tr>
<tr>
<td>Not Stated-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>-0.2143</td>
<td>0.1191</td>
<td>0.0720</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORW</td>
<td>0.8479</td>
<td>0.0849</td>
<td>0.0000</td>
<td>0.6472</td>
<td>0.0796</td>
<td>0.0000</td>
</tr>
<tr>
<td>NORW</td>
<td>0.6222</td>
<td>0.1260</td>
<td>0.0000</td>
<td>0.4199</td>
<td>0.0964</td>
<td>0.0000</td>
</tr>
<tr>
<td>Not-Classified-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d281214</td>
<td>0.1224</td>
<td>0.0464</td>
<td>0.0011</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d567910111334</td>
<td>-0.2419</td>
<td>0.0344</td>
<td>0.0012</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parish -1-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{\text{PARISH}}$</td>
<td>0.0011</td>
<td></td>
<td></td>
<td>0.0012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{\text{HOUSEHOLD}}$</td>
<td>0.7466</td>
<td></td>
<td></td>
<td>0.7470</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>5105.94</td>
<td></td>
<td></td>
<td>5103.45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 6.6: Parameter Estimates for Two-Level Random Coefficient Model using Item Nonresponse Weight Average at the ED Level

<table>
<thead>
<tr>
<th>Variable</th>
<th>PE (SE)</th>
<th>p-value</th>
<th>PE (SE)</th>
<th>p-value</th>
<th>PE (SE)</th>
<th>p-value</th>
<th>PE (SE)</th>
<th>p-value</th>
<th>PE (SE)</th>
<th>p-value</th>
<th>PE (SE)</th>
<th>p-value</th>
<th>PE (SE)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>10.9860</td>
<td>0.0000</td>
<td>11.3542</td>
<td>0.0000</td>
<td>11.0039</td>
<td>0.0000</td>
<td>11.0119</td>
<td>0.0000</td>
<td>10.9307</td>
<td>0.0000</td>
<td>10.5561</td>
<td>0.0000</td>
<td>10.5669</td>
<td>0.0000</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0078</td>
<td>0.0014</td>
<td>-0.0099</td>
<td>0.0011</td>
<td>-0.0099</td>
<td>0.0013</td>
<td>-0.0099</td>
<td>0.0013</td>
<td>-0.0078</td>
<td>0.0014</td>
<td>-0.0078</td>
<td>0.0014</td>
<td>-0.0085</td>
<td>0.0014</td>
</tr>
<tr>
<td>Household Size</td>
<td>0.1246</td>
<td>0.0105</td>
<td>0.1247</td>
<td>0.0106</td>
<td>0.1249</td>
<td>0.0106</td>
<td>0.1176</td>
<td>0.0107</td>
<td>0.1169</td>
<td>0.0108</td>
<td>0.1169</td>
<td>0.0108</td>
<td>0.1169</td>
<td>0.0108</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-0.0201</td>
<td>0.0435</td>
<td>0.6440</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0291</td>
<td>0.0436</td>
<td>-0.0291</td>
<td>0.0436</td>
<td>-0.0259</td>
<td>0.0434</td>
</tr>
<tr>
<td>Male -Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0.3585</td>
<td>0.0598</td>
<td>0.3163</td>
<td>0.0809</td>
<td>0.2513</td>
<td>0.0621</td>
<td>0.2507</td>
<td>0.0624</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Stated – Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>0.3976</td>
<td>0.0809</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.2283</td>
<td>0.1545</td>
<td>0.1390</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Office Related Work (ORW)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Office Related Work (NORW)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not -Classified -Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Random Component</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{PSU}$</td>
<td>0.121243</td>
<td></td>
<td>0.117924</td>
<td></td>
<td>0.126309</td>
<td></td>
<td>0.126238</td>
<td></td>
<td>0.110822</td>
<td></td>
<td>0.109561</td>
<td></td>
<td>0.104006</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{HOUSEHOLD}$</td>
<td>0.796199</td>
<td></td>
<td>0.782340</td>
<td></td>
<td>0.706104</td>
<td></td>
<td>0.705936</td>
<td></td>
<td>0.702076</td>
<td></td>
<td>0.686744</td>
<td></td>
<td>0.678646</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>5392.82</td>
<td></td>
<td>5358.74</td>
<td></td>
<td>5176.98</td>
<td></td>
<td>5178.72</td>
<td></td>
<td>5154.05</td>
<td></td>
<td>5113.38</td>
<td></td>
<td>5088.93</td>
<td></td>
</tr>
</tbody>
</table>

PE = Parameter Estimates, SE = Standard Error, AIC = Akaike Information Criterion
Table 6.7: Parameter Estimates for Three-Level Random Coefficient Model at Using Item Nonresponse Weight Average at the ED Level

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model III-C1</th>
<th>Model III-C2</th>
<th>Model III-C3</th>
<th>Model III-C4</th>
<th>Model III-C5</th>
<th>Model III-C6</th>
<th>Model III-C7</th>
<th>Model III-C8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PE (SE)</td>
<td>p-value</td>
<td>PE (SE)</td>
<td>p-value</td>
<td>PE (SE)</td>
<td>p-value</td>
<td>PE (SE)</td>
<td>p-value</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>10.8969</td>
<td>0.0000</td>
<td>11.2468</td>
<td>0.0000</td>
<td>10.8938</td>
<td>0.0000</td>
<td>10.8437</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0636)</td>
<td></td>
<td>(0.0802)</td>
<td></td>
<td>(0.1023)</td>
<td></td>
<td>(0.1049)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.0073</td>
<td>0.0000</td>
<td>-0.0095</td>
<td>0.0000</td>
<td>-0.0095</td>
<td>0.0000</td>
<td>-0.0087</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td></td>
<td>(0.0013)</td>
<td></td>
<td>(0.0011)</td>
<td></td>
<td>(0.0011)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household Size</td>
<td>0.1248</td>
<td>0.0000</td>
<td>0.1249</td>
<td>0.0000</td>
<td>0.1254</td>
<td>0.0000</td>
<td>0.1178</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0134)</td>
<td></td>
<td>(0.0133)</td>
<td></td>
<td>(0.0134)</td>
<td></td>
<td>(0.0144)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-0.0191</td>
<td>0.0000</td>
<td>-0.0189</td>
<td>0.0000</td>
<td>-0.0276</td>
<td>0.0000</td>
<td>-0.0259</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0364)</td>
<td></td>
<td>(0.0359)</td>
<td></td>
<td>(0.0355)</td>
<td></td>
<td>(0.0345)</td>
<td></td>
</tr>
<tr>
<td>Male -Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educated</td>
<td>0.3189</td>
<td>0.0000</td>
<td>0.2734</td>
<td>0.0000</td>
<td>0.2103</td>
<td>0.0000</td>
<td>0.2097</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0435)</td>
<td></td>
<td>(0.0399)</td>
<td></td>
<td>(0.0348)</td>
<td></td>
<td>(0.0354)</td>
<td></td>
</tr>
<tr>
<td>Not Stated – Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>0.4029</td>
<td>0.0000</td>
<td>-0.2008</td>
<td>0.0000</td>
<td>-0.2008</td>
<td>0.0000</td>
<td>-0.2008</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0898)</td>
<td></td>
<td>(0.1123)</td>
<td></td>
<td>(0.1123)</td>
<td></td>
<td>(0.1123)</td>
<td></td>
</tr>
<tr>
<td>Unemployed-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Office Related Work (ORW)</td>
<td>0.8114</td>
<td>0.0000</td>
<td>0.6227</td>
<td>0.0000</td>
<td>0.4073</td>
<td>0.0000</td>
<td>0.4073</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0672)</td>
<td></td>
<td>(0.0779)</td>
<td></td>
<td>(0.0951)</td>
<td></td>
<td>(0.0951)</td>
<td></td>
</tr>
<tr>
<td>Non-Office Related Work (NORW)</td>
<td>0.5974</td>
<td>0.0000</td>
<td>0.4073</td>
<td>0.0000</td>
<td>0.4073</td>
<td>0.0000</td>
<td>0.4073</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.1071)</td>
<td></td>
<td>(0.0951)</td>
<td></td>
<td>(0.0951)</td>
<td></td>
<td>(0.0951)</td>
<td></td>
</tr>
<tr>
<td>Not -Classified -Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma^2_{\text{PARISH}} )</td>
<td>0.037947</td>
<td>0.034782</td>
<td>0.033452</td>
<td>0.033416</td>
<td>0.028690</td>
<td>0.028527</td>
<td>0.026634</td>
<td>0.027093</td>
</tr>
<tr>
<td>( \sigma^2_{\text{ED}} )</td>
<td>0.077674</td>
<td>0.07958</td>
<td>0.086084</td>
<td>0.088084</td>
<td>0.079468</td>
<td>0.077395</td>
<td>0.074365</td>
<td>0.074365</td>
</tr>
<tr>
<td>( \sigma^2_{\text{HOUSEHOLD}} )</td>
<td>0.795842</td>
<td>0.781946</td>
<td>0.705768</td>
<td>0.705768</td>
<td>0.705768</td>
<td>0.686247</td>
<td>0.678152</td>
<td>0.678646</td>
</tr>
<tr>
<td>AIC</td>
<td>5364.57</td>
<td>5334.80</td>
<td>5151.05</td>
<td>5152.81</td>
<td>5133.63</td>
<td>5091.66</td>
<td>5068.51</td>
<td>5066.083</td>
</tr>
</tbody>
</table>

PE=Parameter Estimates, SE = Standard Error, AIC = Akaike Information Criterion
6.8 Analysis of the CHNS -1989 &2011 Survey Data

The China Health and Nutrition Survey is an ongoing international study between the Carolina Population Centre at the University of North Carolina at Chapel Hill and the National Institute for Nutrition and Health (NINH, formerly the National Institute of Nutrition and Food Safety) at the Chinese Centre for Disease Control and Prevention (CCDC). This survey covered a range of Chinese provinces which varied substantially in geography, economic development, public resources, and health indicators.

The survey data was downloaded from the Institute’s website at: 
http://cpc.unc.unnc.edu/projects/china/about/proj_desc/survey.

Based on the objectives of this study, two separate data set were extracted for the years 1989 and 2011, which focused on reported income and socio-demographic variables of the survey respondents. This was to enable the modelling of Chinese population income dynamics for the years 1989 and 2011. Three thousand and fifty-two (3,052) subjects’ information in the age range 16 to 65 years were extracted from 1,760 households in 161 communities spread across four strata in eight provinces in the 1989 data set. In 2011, a total of 2,109 subjects’ information in the age range 16 to 65 years were extracted from 1,188 households in 167 communities spread across four strata in 12 provinces.

Table 6.6 contains the summary of the variables extracted for this analysis. The list included gender, marital status, years of schooling, reported income with imputation, and without imputation. In the summary, the data extracted for 1989 showed that the majority of individuals for the age group 16 to 65 years were male (52.6%), married (87.2%), and rural folks (69.3%) with an average of 15.31 years of schooling and average age of 35.03 years. However, in the 2011 data, the majority of individuals for the age group 16 to 65 years were female (53.3%), married (92.7%), city dwellers (35.7%) with an average of 24.01 years of schooling and average age of 45.95 years.

These two datasets showed sharp contrast in the years of schooling of city dwellers and rural folks, and also the spread of the survey in 1989 and 2011 across the provinces in China. The natural logarithm of the reported income,
denoted as $y_{ilm}$ was used in the two analyses for the years 1989 and 2011, respectively.

Table 6.8: Profile of Individuals in the Extracted Data from China Health and Nutrition Survey

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CATEGORICAL VARIABLE</th>
<th>1989</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FREQUENCY</td>
<td>PERCENTAGE</td>
<td>FREQUENCY</td>
</tr>
<tr>
<td>GENDER</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALE</td>
<td>1606</td>
<td>52.6</td>
<td>984</td>
</tr>
<tr>
<td>FEMALE</td>
<td>1446</td>
<td>47.4</td>
<td>1125</td>
</tr>
<tr>
<td>TOTAL</td>
<td>3052</td>
<td>100</td>
<td>2109</td>
</tr>
<tr>
<td>MARITAL STATUS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MARRIED</td>
<td>2662</td>
<td>87.2</td>
<td>1955</td>
</tr>
<tr>
<td>NOT MARRIED</td>
<td>390</td>
<td>12.8</td>
<td>154</td>
</tr>
<tr>
<td>TOTAL</td>
<td>3052</td>
<td>100</td>
<td>2109</td>
</tr>
<tr>
<td>PROVINCE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEIJING</td>
<td>598</td>
<td>28.4</td>
<td>7</td>
</tr>
<tr>
<td>LIAONING</td>
<td>303</td>
<td>9.9</td>
<td>7</td>
</tr>
<tr>
<td>HEILONGJIANG</td>
<td>355</td>
<td>11.6</td>
<td>89</td>
</tr>
<tr>
<td>SHANGHAI</td>
<td>335</td>
<td>11</td>
<td>57</td>
</tr>
<tr>
<td>JIANGSU</td>
<td>333</td>
<td>10.9</td>
<td>22</td>
</tr>
<tr>
<td>SHANDONG</td>
<td>470</td>
<td>15.4</td>
<td>41</td>
</tr>
<tr>
<td>HENAN</td>
<td>383</td>
<td>12.5</td>
<td>38</td>
</tr>
<tr>
<td>HUBEI</td>
<td>478</td>
<td>15.7</td>
<td>29</td>
</tr>
<tr>
<td>HUNAN</td>
<td>395</td>
<td>12.9</td>
<td>40</td>
</tr>
<tr>
<td>GUANGXI</td>
<td>528</td>
<td>25.0</td>
<td>29</td>
</tr>
<tr>
<td>GUANGXI</td>
<td>528</td>
<td>25.0</td>
<td>29</td>
</tr>
<tr>
<td>GUANGXI</td>
<td>528</td>
<td>25.0</td>
<td>29</td>
</tr>
<tr>
<td>CHONGQING</td>
<td>528</td>
<td>25.0</td>
<td>29</td>
</tr>
<tr>
<td>TOTAL</td>
<td>3052</td>
<td>100</td>
<td>2109</td>
</tr>
<tr>
<td>STRATUM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CITY</td>
<td>173</td>
<td>5.7</td>
<td>753</td>
</tr>
<tr>
<td>SUBURBAN</td>
<td>449</td>
<td>14.7</td>
<td>663</td>
</tr>
<tr>
<td>TOWN OR COUNTY</td>
<td>316</td>
<td>10.4</td>
<td>433</td>
</tr>
<tr>
<td>CAPITAL</td>
<td>2114</td>
<td>69.3</td>
<td>260</td>
</tr>
<tr>
<td>RURAL</td>
<td>3052</td>
<td>100</td>
<td>2109</td>
</tr>
<tr>
<td>VILLAGE</td>
<td>3052</td>
<td>100</td>
<td>2109</td>
</tr>
<tr>
<td>MISSING INDICATOR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MISSING</td>
<td>49</td>
<td>1.6</td>
<td>164</td>
</tr>
<tr>
<td>NOT MISSING</td>
<td>3003</td>
<td>98.4</td>
<td>1945</td>
</tr>
<tr>
<td>TOTAL</td>
<td>3052</td>
<td>100</td>
<td>2109</td>
</tr>
<tr>
<td>CONTINUOUS VARIABLE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>35.03</td>
<td>10.14</td>
<td>45.95</td>
</tr>
<tr>
<td>YEARS OF SCHOOLING</td>
<td>15.31</td>
<td>8.81</td>
<td>24.01</td>
</tr>
<tr>
<td>REPORTED INCOME</td>
<td>1371.93</td>
<td>1652.96</td>
<td>28733.88</td>
</tr>
<tr>
<td>WITHOUT IMPUTATION</td>
<td>1376.01</td>
<td>1655.09</td>
<td>29383.51</td>
</tr>
<tr>
<td>REPORTED INCOME</td>
<td>1371.93</td>
<td>1652.96</td>
<td>28733.88</td>
</tr>
<tr>
<td>WITH IMPUTATION</td>
<td>1376.01</td>
<td>1655.09</td>
<td>29383.51</td>
</tr>
</tbody>
</table>
6.9 Result of the CHNS 1989 Analysis

There were 3,052 reported income cases in the 1989 dataset. Of the 3,052 cases, 49 were imputed representing 1.6% of the data. This led to two separate models: Model -IV D1 to Model -IV D8 (Table 5.7) using reported income without missing data and another denoted by Model -IVE1 to Model -IVE6 using reported income with missing data (Table 5.8). Both analyses were under the presumption of four-level multilevel model involving, Provinces at the highest level, followed by County-City level, then followed by Community and lastly the individual level. The four-level model enable estimates of the variation at each level in the overall variation in the outcome variable.

In Table 6.7, the estimated parameters for the Models: Model-IVD1 to Model-IVD6 for reported income without missing data were found to have a lower standard errors of estimate and AIC value when compared to the estimated parameters for the weighted Model-IVE1 to Model-IVE8 in Table 6.8 based on the reported income with missing data. In essence, the weight had contributed to the higher standard errors and AIC values. However, the final predictor variables are similar for the final Models in Tables 6.7 and 6.8 respectively. The significant (p-value less than 0.05) predictor variables for the reported income in the Year 1989 include age, gender and years of schooling. The noticeable difference in the Models is the higher standard errors and higher AIC values from the weighted Models arising from the adjustment for the missing data and the non-response of the individuals in the data. The ICC values showed that the majority (85.2%) of the variation in the reported income without missing data and 84.3% for the reported income with missing data existed at the individual level (see Table 6.11). This compared to the percentage of variation in the reported income at the Province, County-City and Neighbourhood -Village levels respectively.
6.10 Result of the CHNS 2011 Analysis

The CHNS 2011 extracted data had 2,109 individuals of which 164 or 7.8% did not provide the income. A series of multilevel models were investigated to identify variations in the data structure. Of all the multilevel models investigated, variations were found to exist in the three-level model starting with individual nested within the Community and nested within the province. Table 6.9 contains the parameter estimates for Model-IIIF1 to Model-IIIF6, involving income variable without missing data while Table 6.10 contains parameter estimates for Model -IIIG1 to Model -IIIG6 involving income variable with missing data. A comparable analysis revealed that standard errors in Table 6.9 are lower than the errors in Table 6.10 which due to the incorporation of the item non-response weight. Similar results were observed for the AIC values. Of all the investigated predictor variables in the formulated Models in Tables 6.9 and 6.10, gender and years of schooling were found to be significant (p-value less than 0.05). The use of the item non-response weight did not affect the final variables in either Model (Model - IIIF6, Model-IIIG6)

The ICC estimates for the 2011 dataset in Table 6.11 showed that the majority (73.97%) of the variation in the reported income without missing data and 74.86% of the variation in the reported income with missing data existed at the individual level.
### Table 6.9: Parameter Estimates for Four-Level Model of Reported Income without Missing Data - Year 1989 for Subjects 16 to 65 Years

<table>
<thead>
<tr>
<th>Variable</th>
<th>MODEL IV-D1</th>
<th>MODEL IV-D2</th>
<th>MODEL IV-D3</th>
<th>MODEL IV-D4</th>
<th>MODEL IV-D5</th>
<th>MODEL IV-D6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PE (SE) p-value</td>
<td>PE (SE) p-value</td>
<td>PE (SE) p-value</td>
<td>PE (SE) p-value</td>
<td>PE (SE) p-value</td>
<td>PE (SE) p-value</td>
</tr>
<tr>
<td><strong>Fixed Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>6.7190 (0.0736)</td>
<td>6.4072 (0.1001)</td>
<td>6.3134 (0.1038)</td>
<td>6.0161 (0.1217)</td>
<td>5.9052 (0.1362)</td>
<td>6.0191 (0.1217)</td>
</tr>
<tr>
<td>Age</td>
<td>0.0089 (0.0019)</td>
<td>0.0096 (0.0019)</td>
<td>0.0135 (0.0021)</td>
<td>0.0143 (0.0021)</td>
<td>0.0137 (0.0021)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.1324 (0.0381)</td>
<td>0.0639 (0.0407)</td>
<td>0.1170 (0.0409)</td>
<td>0.0572 (0.0409)</td>
<td>0.1620 (0.0409)</td>
<td></td>
</tr>
<tr>
<td>Female-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of Schooling</td>
<td></td>
<td>0.0127 (0.0027)</td>
<td>0.0122 (0.0027)</td>
<td>0.0142 (0.0025)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1087 (0.0603)</td>
<td>0.0720 (0.0603)</td>
</tr>
<tr>
<td>Never married-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Random Component</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{Province}$</td>
<td>0.0187 (0.0194)</td>
<td>0.0194 (0.0194)</td>
<td>0.0229 (0.0227)</td>
<td>0.0227 (0.0233)</td>
<td>0.0233</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{County-City}$</td>
<td>0.0529 (0.0522)</td>
<td>0.0537 (0.0537)</td>
<td>0.0436 (0.0435)</td>
<td>0.0435 (0.0419)</td>
<td>0.0419</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{Neighbourhood-Village}$</td>
<td>0.1190 (0.1182)</td>
<td>0.1172 (0.1072)</td>
<td>0.1072 (0.1068)</td>
<td>0.1068 (0.1065)</td>
<td>0.1065</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{Individual}$</td>
<td>1.0943 (1.0866)</td>
<td>1.0824 (1.0777)</td>
<td>1.0777 (1.0766)</td>
<td>1.0766 (1.0789)</td>
<td>1.0789</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>9128 (9108)</td>
<td>9098 (9079)</td>
<td>9079 (9078)</td>
<td>9078 (9079)</td>
<td>9079 (9079)</td>
<td></td>
</tr>
</tbody>
</table>

PE = Parameter Estimates, SE = Standard Error, AIC = Akaike Information Criterion
Table 6.10: Parameter Estimates for Weighted Four-Level Model of Reported Income with Missing Data - Year 1989 for Subjects 16 to 65 Years

<table>
<thead>
<tr>
<th>Variable</th>
<th>MODEL IV-E1</th>
<th>MODEL IV-E2</th>
<th>MODEL IV-E3</th>
<th>MODEL IV-E4</th>
<th>MODEL IV-E5</th>
<th>MODEL IV-E6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PE (SE)</td>
<td>p-value</td>
<td>PE (SE)</td>
<td>p-value</td>
<td>PE (SE)</td>
<td>p-value</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>6.7061</td>
<td>0.0000</td>
<td>6.3836</td>
<td>0.0000</td>
<td>6.2882</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0815)</td>
<td></td>
<td>(0.1175)</td>
<td></td>
<td>(0.1154)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.0092</td>
<td>0.0000</td>
<td>0.0099</td>
<td>0.0000</td>
<td>0.0137</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0028)</td>
<td></td>
<td>(0.0024)</td>
<td></td>
<td>(0.0030)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.1417</td>
<td>0.2520</td>
<td>0.0754</td>
<td>0.5760</td>
<td>0.0702</td>
<td>0.6110</td>
</tr>
<tr>
<td></td>
<td>(0.1237)</td>
<td></td>
<td>(0.1348)</td>
<td></td>
<td>(0.1334)</td>
<td></td>
</tr>
<tr>
<td>Female-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of Schooling</td>
<td>0.0123</td>
<td>0.0000</td>
<td>0.0120</td>
<td>0.0000</td>
<td>0.0141</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0031)</td>
<td></td>
<td>(0.0031)</td>
<td></td>
<td>(0.0018)</td>
<td></td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0856</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0593)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{\text{Province}}$</td>
<td>0.0210</td>
<td>0.0214</td>
<td>0.0216</td>
<td>0.0255</td>
<td>0.0255</td>
<td>0.0261</td>
</tr>
<tr>
<td>$\sigma^2_{\text{County-City}}$</td>
<td>0.1016</td>
<td>0.1001</td>
<td>0.1015</td>
<td>0.0884</td>
<td>0.0887</td>
<td>0.0864</td>
</tr>
<tr>
<td>$\sigma^2_{\text{Neighbourhood/Village}}$</td>
<td>0.0794</td>
<td>0.0789</td>
<td>0.0782</td>
<td>0.0714</td>
<td>0.0711</td>
<td>0.0708</td>
</tr>
<tr>
<td>$\sigma^2_{\text{Individual}}$</td>
<td>1.0897</td>
<td>1.0889</td>
<td>1.0841</td>
<td>1.0791</td>
<td>1.0785</td>
<td>1.0806</td>
</tr>
<tr>
<td>AIC</td>
<td>556372</td>
<td>554971</td>
<td>554131</td>
<td>552879</td>
<td>552759</td>
<td>553087</td>
</tr>
</tbody>
</table>

PE = Parameter Estimate, SE = Standard Error, AIC = Akaike Information Criterion
Table 6.11: Parameter Estimates for Three-Level Model of Reported Income without missing data - Year 1989 for Subjects 16 to 65 Years (n=3052)

<table>
<thead>
<tr>
<th>Variable</th>
<th>MODEL IV-A1</th>
<th>MODEL IV-B1</th>
<th>MODEL IV-C1</th>
<th>MODEL IV-D1</th>
<th>MODEL IV-E1</th>
<th>MODEL IV-F1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PE (SE)</td>
<td>p-value</td>
<td>PE (SE)</td>
<td>p-value</td>
<td>PE (SE)</td>
<td>p-value</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>6.760389</td>
<td>0.0000</td>
<td>6.4368</td>
<td>0.0000</td>
<td>6.346416</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0710017)</td>
<td></td>
<td>(0.099857)</td>
<td></td>
<td>(0.1037023)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.00917</td>
<td>0.0000</td>
<td>0.0097959</td>
<td>0.0000</td>
<td>0.015186</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.019705)</td>
<td></td>
<td>(0.0019762)</td>
<td></td>
<td>(0.0021357)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.1296162</td>
<td>0.0010</td>
<td>0.0373579</td>
<td>0.3730</td>
<td>0.0301265</td>
<td>0.4740</td>
</tr>
<tr>
<td></td>
<td>(0.0395742)</td>
<td></td>
<td>(0.0419098)</td>
<td></td>
<td>(0.0420557)</td>
<td></td>
</tr>
<tr>
<td>Female-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of Schooling</td>
<td>0.0171333</td>
<td>0.0000</td>
<td>0.0166797</td>
<td>0.0000</td>
<td>0.0179623</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0026803)</td>
<td></td>
<td>(0.0026892)</td>
<td></td>
<td>(0.025161)</td>
<td></td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>0.1182221</td>
<td>0.0560</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0618875)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{Province}$</td>
<td>0.0209</td>
<td>0.0222</td>
<td>0.0227</td>
<td>0.0251</td>
<td>0.1571</td>
<td>0.0252</td>
</tr>
<tr>
<td>$\sigma^2_{Neighbourhood/Village}$</td>
<td>0.0627</td>
<td>0.0605</td>
<td>0.0603</td>
<td>0.0482</td>
<td>0.2193</td>
<td>0.0478</td>
</tr>
<tr>
<td>$\sigma^2_{Individual}$</td>
<td>1.1885</td>
<td>1.1803</td>
<td>1.1761</td>
<td>1.1622</td>
<td>1.0774</td>
<td>1.1626</td>
</tr>
<tr>
<td>AIC</td>
<td>9257.8</td>
<td>9238.2</td>
<td>9229.5</td>
<td>9191.2</td>
<td>9189.57</td>
<td>9190</td>
</tr>
</tbody>
</table>

PE = Parameter Estimate, SE = Standard Error, AIC = Akaike Information Criterion
Table 6.12: Parameter Estimates for Weighted Three-Level Model of Reported Income without Imputed data - Year 1989 for Subjects 16 to 65 Years

<table>
<thead>
<tr>
<th>Variable</th>
<th>MODEL IV-A2</th>
<th>MODEL IV-B2</th>
<th>MODEL IV-C2</th>
<th>MODEL IV-D2</th>
<th>MODEL IV-E2</th>
<th>MODEL IV-F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>6.756339</td>
<td>6.39597</td>
<td>6.301246</td>
<td>5.869976</td>
<td>5.7665</td>
<td>5.9039</td>
</tr>
<tr>
<td></td>
<td>(0.077009)</td>
<td>(0.13054)</td>
<td>(0.129399)</td>
<td>(0.1516843)</td>
<td>(0.1645051)</td>
<td>(0.144765)</td>
</tr>
<tr>
<td>Age</td>
<td>0.0102184</td>
<td>0.0108658</td>
<td>0.0163964</td>
<td>0.0171027</td>
<td>0.0158784</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.00313)</td>
<td>(0.0028902)</td>
<td>(0.0032)</td>
<td>(0.0031)</td>
<td>(0.0030157)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.134377</td>
<td>0.0420</td>
<td>0.0420</td>
<td>0.0359656</td>
<td>0.7650</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1178)</td>
<td>(0.12203)</td>
<td>(0.1202)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of Schooling</td>
<td>0.0181041</td>
<td>0.017725</td>
<td>0.017725</td>
<td>0.0185</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0026086)</td>
<td>(0.00266)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>0.1005249</td>
<td>0.0980</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{\text{Province}}$</td>
<td>0.0414</td>
<td>0.0420</td>
<td>0.0425</td>
<td>0.0423</td>
<td>0.0419</td>
<td>0.0423</td>
</tr>
<tr>
<td>$\sigma^2_{\text{Neighbourhood/Village}}$</td>
<td>0.0412</td>
<td>0.0398</td>
<td>0.0396</td>
<td>0.0310</td>
<td>0.0308</td>
<td>0.0297</td>
</tr>
<tr>
<td>$\sigma^2_{\text{Individual}}$</td>
<td>1.1961</td>
<td>1.1860</td>
<td>1.1816</td>
<td>1.1667</td>
<td>1.1657</td>
<td>1.1694</td>
</tr>
<tr>
<td>$\text{AIC}$</td>
<td>574339.9</td>
<td>572696.5</td>
<td>571989</td>
<td>569257.3</td>
<td>569097.7</td>
<td>563146.6</td>
</tr>
</tbody>
</table>

PE =Parameter Estimate, SE=Standard Error, AIC = Akaike Information Criterion
Table 6.13: Parameter Estimates for Three-Level Model of Reported Income without Missing Data - Year 2011 for Subjects 16 to 65 Years

<table>
<thead>
<tr>
<th>Variable</th>
<th>MODEL III-F1</th>
<th>MODEL III-F2</th>
<th>MODEL III-F3</th>
<th>MODEL III-F4</th>
<th>MODEL III-F5</th>
<th>MODEL III-F6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>9.6737</td>
<td>0.0000</td>
<td>10.1311</td>
<td>0.0000</td>
<td>10.0296</td>
<td>0.0000</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0108</td>
<td>0.0000</td>
<td>-0.0113</td>
<td>0.6680</td>
<td>0.0007</td>
<td>0.6680</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.2750</td>
<td>0.0000</td>
<td>0.1802</td>
<td>0.0000</td>
<td>0.1765</td>
<td>0.0000</td>
</tr>
<tr>
<td>Female-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of Schooling</td>
<td>0.0469</td>
<td>0.0000</td>
<td>0.0470</td>
<td>0.0000</td>
<td>0.0463</td>
<td>0.0000</td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{\text{Province}}$</td>
<td>0.1779</td>
<td>0.2124</td>
<td>0.2016</td>
<td>0.1224</td>
<td>0.1235</td>
<td>0.1249</td>
</tr>
<tr>
<td>$\sigma^2_{\text{Neighbourhood/Village}}$</td>
<td>0.7397</td>
<td>0.7238</td>
<td>0.7051</td>
<td>0.6367</td>
<td>0.6363</td>
<td>0.6366</td>
</tr>
</tbody>
</table>

PE = Parameter Estimates, SE = Standard Error, AIC = Akaike Information Criterion
Table 6.14: Parameter Estimates for Weighted Three-Level Model of Reported Income with Missing Data - Year 2011 for Subjects 16 to 65 Years

<table>
<thead>
<tr>
<th>Variable</th>
<th>MODEL III-G1</th>
<th>MODEL III-G2</th>
<th>MODEL III-G3</th>
<th>MODEL III-G4</th>
<th>MODEL III-G5</th>
<th>MODEL III-G6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>9.6557 0.0000</td>
<td>10.09018 0.0000</td>
<td>9.98304 0.0000</td>
<td>8.45406 0.0000</td>
<td>8.394639 0.0000</td>
<td>8.556097 0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.1437)</td>
<td>(0.243043)</td>
<td>(0.23154)</td>
<td>(0.2239)</td>
<td>(0.240086)</td>
<td>(0.138493)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.01019 0.0030</td>
<td>-0.0109 0.0010</td>
<td>0.00096 0.7590</td>
<td>0.000904 0.7790</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003384)</td>
<td>(0.00321)</td>
<td>(0.00313)</td>
<td>(0.003221)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.2903 0.19322</td>
<td>0.190192 0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02499)</td>
<td>(0.03829)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of Schooling</td>
<td>0.04555 0.0000</td>
<td>0.045628 0.0000</td>
<td>0.046806 0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00515)</td>
<td>(0.00525)</td>
<td>(0.00554)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.065247 0.466</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.08957)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{\text{Province}}$</td>
<td>0.2241</td>
<td>0.2595</td>
<td>0.2470</td>
<td>0.1619</td>
<td>0.1626</td>
<td>0.1719</td>
</tr>
<tr>
<td>$\sigma^2_{\text{Neighbourhood/Village}}$</td>
<td>0.0434</td>
<td>0.0441</td>
<td>0.0443</td>
<td>0.0201</td>
<td>0.0201</td>
<td>0.0196</td>
</tr>
<tr>
<td>$\sigma^2_{\text{Individual}}$</td>
<td>0.7599</td>
<td>0.7463</td>
<td>0.7253</td>
<td>0.6579</td>
<td>0.6576</td>
<td>0.6675</td>
</tr>
<tr>
<td>AIC</td>
<td>65776.46</td>
<td>65329.7</td>
<td>64615.51</td>
<td>61960.92</td>
<td>61952.49</td>
<td>62316.8</td>
</tr>
</tbody>
</table>

PE = Parameter Estimate, SE = Standard Error, AIC = Akaike Information Criterion
Table 6.15: Parameter Estimates for Four-Level Model of Reported Income without Missing Data - Year 2011 for Subjects 16 to 65 Years

<table>
<thead>
<tr>
<th>Variable</th>
<th>MODEL IV-A3</th>
<th>MODEL IV-B3</th>
<th>MODEL IV-C3</th>
<th>MODEL IV-D3</th>
<th>MODEL IV-E3</th>
<th>MODEL IV-F3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PE (SE)</td>
<td>p-value</td>
<td>PE (SE)</td>
<td>p-value</td>
<td>PE (SE)</td>
<td>p-value</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>9.6629</td>
<td>0.0000</td>
<td>10.079</td>
<td>0.0000</td>
<td>8.5576</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>0.1390</td>
<td>0.1646</td>
<td>0.1619</td>
<td>0.1674</td>
<td>0.1777</td>
<td>0.1319</td>
</tr>
<tr>
<td>Age</td>
<td>0.0099</td>
<td>0.0000</td>
<td>0.0105</td>
<td>0.0004</td>
<td>0.7970</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.0017</td>
<td>0.0017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2699</td>
<td>0.0000</td>
<td>0.1846</td>
<td>0.0000</td>
<td>0.1814</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>0.0356</td>
<td>0.0348</td>
<td>0.0349</td>
<td></td>
<td>0.0346</td>
<td></td>
</tr>
<tr>
<td>Female-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of Schooling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0429</td>
<td>0.0000</td>
<td>0.0429</td>
<td>0.0000</td>
<td>0.0429</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>0.0348</td>
<td>0.031</td>
<td></td>
<td></td>
<td>0.0027</td>
<td></td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0683</td>
<td>0.3070</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0667</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{\text{Province}}$</td>
<td>0.1981</td>
<td>0.2330</td>
<td>0.2223</td>
<td>0.1321</td>
<td>0.1332</td>
<td>0.1337</td>
</tr>
<tr>
<td>$\sigma^2_{\text{County-City}}$</td>
<td>0.0000</td>
<td>0.0066</td>
<td>0.0079</td>
<td>0.0056</td>
<td>0.0055</td>
<td>0.0058</td>
</tr>
<tr>
<td>$\sigma^2_{\text{Neighbourhood/Village}}$</td>
<td>0.1142</td>
<td>0.1110</td>
<td>0.1103</td>
<td>0.0591</td>
<td>0.0590</td>
<td>0.0593</td>
</tr>
<tr>
<td>$\sigma^2_{\text{Individual}}$</td>
<td>0.6808</td>
<td>0.6677</td>
<td>0.6495</td>
<td>0.6058</td>
<td>0.6054</td>
<td>0.6056</td>
</tr>
<tr>
<td>AIC</td>
<td>5334</td>
<td>5299</td>
<td>5244</td>
<td>5064</td>
<td>5065</td>
<td>5062</td>
</tr>
</tbody>
</table>
Table 6.16: Parameter Estimates for Weighted Four-Level Model of Reported Income with Missing Data - Year 2011 for Subjects 16 to 65 Years

<table>
<thead>
<tr>
<th>Variable</th>
<th>MODEL IV-A4</th>
<th>MODEL IV-B4</th>
<th>MODEL IV-C4</th>
<th>MODEL IV-D4</th>
<th>MODEL IV-E4</th>
<th>MODEL IV-F4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PE (SE)</td>
<td>p-value</td>
<td>PE (SE)</td>
<td>p-value</td>
<td>PE (SE)</td>
<td>p-value</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>9.5715</td>
<td>0.0000</td>
<td>9.9768</td>
<td>0.0000</td>
<td>8.4479</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>0.1669</td>
<td>0.25959</td>
<td>0.2474</td>
<td>0.2215</td>
<td>0.2309</td>
<td>0.1526</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0099</td>
<td>0.0070</td>
<td>-0.0111</td>
<td>0.0020</td>
<td>-0.0000</td>
<td>0.9930</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.9930</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>0.0037</td>
<td>0.0036</td>
<td>0.0034</td>
<td>0.0035</td>
<td>0.0356</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of Schooling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0433</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0047</td>
<td>0.0049</td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0743</td>
<td>0.3300</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0763</td>
<td></td>
</tr>
<tr>
<td>Never married-Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{province}$</td>
<td>0.2296</td>
<td>0.2330</td>
<td>0.2223</td>
<td>0.1321</td>
<td>0.1332</td>
<td>0.1337</td>
</tr>
<tr>
<td>$\sigma^2_{County-City}$</td>
<td>0.2604</td>
<td>0.0066</td>
<td>0.0079</td>
<td>0.0056</td>
<td>0.0055</td>
<td>0.0058</td>
</tr>
<tr>
<td>$\sigma^2_{Neighbourhood/Village}$</td>
<td>0.0738</td>
<td>0.1110</td>
<td>0.1103</td>
<td>0.0591</td>
<td>0.0590</td>
<td>0.0593</td>
</tr>
<tr>
<td>$\sigma^2_{Individual}$</td>
<td>0.6808</td>
<td>0.6677</td>
<td>0.6495</td>
<td>0.6058</td>
<td>0.6054</td>
<td>0.6056</td>
</tr>
<tr>
<td>AIC</td>
<td>5334</td>
<td>5299</td>
<td>5244</td>
<td>5064</td>
<td>5065</td>
<td>5062</td>
</tr>
</tbody>
</table>
Table 6.17: Variance and Intraclass Estimates for the CHNS Data-1989 and 2011

<table>
<thead>
<tr>
<th>Four Level Model</th>
<th>1989</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_{jklm} = x_{jklm}^T \beta + u_j + v_{jk} + \tau_{kld} + \varepsilon_{jklm} )</td>
<td></td>
</tr>
<tr>
<td>Where ( y_{jklm} ) represents reported individual income without missing data while ( \sigma_u^2, \sigma_v^2, \sigma_{\tau}^2, \sigma_{\varepsilon}^2 ) represents the variances at the Province, County-City, Neighbourhood / Village and individual levels respectively.</td>
<td></td>
</tr>
<tr>
<td>( \sigma_u^2 = 0.0187 \quad \rho_u = 0.015 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma_v^2 = 0.0529 \quad \rho_v = 0.041 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma_{\tau}^2 = 0.1190 \quad \rho_{\tau} = 0.093 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma_{\varepsilon}^2 = 1.0943 \quad \rho_{\varepsilon} = 0.852 )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Three Level Model</th>
<th>1989</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_{jkl} = x_{jkl}^T \beta + u_j + v_{jk} + \varepsilon_{jkl} )</td>
<td></td>
</tr>
<tr>
<td>Where ( y_{jkl} ) represents reported individual income without imputed data while ( \sigma_u^2, \sigma_v^2, \sigma_{\varepsilon}^2 ) represents the variances at the Province, Neighbourhood/Village and individual levels respectively.</td>
<td></td>
</tr>
<tr>
<td>( \sigma_u^2 = 0.0210 \quad \rho_u = 0.016 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma_v^2 = 0.1016 \quad \rho_v = 0.078 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma_{\varepsilon}^2 = 0.0974 \quad \rho_{\varepsilon} = 0.061 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma_{\varepsilon}^2 = 1.0971 \quad \rho_{\varepsilon} = 0.845 )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Three Level Model</th>
<th>1989</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_{jkl} = x_{jkl}^T \beta + u_j + v_{jk} + \varepsilon_{jkl} )</td>
<td></td>
</tr>
<tr>
<td>Where ( y_{jkl} ) represents reported individual income with missing data while ( \sigma_u^2, \sigma_v^2, \sigma_{\varepsilon}^2 ) represents the variances at the Province, Neighbourhood-Village and individual levels respectively.</td>
<td></td>
</tr>
<tr>
<td>( \sigma_u^2 = 0.0214 \quad \rho_u = 0.0164 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma_v^2 = 0.063 \quad \rho_v = 0.04923 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma_{\varepsilon}^2 = 1.189 \quad \rho_{\varepsilon} = 0.93429 )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Three Level Model</th>
<th>1989</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_{jkl} = x_{jkl}^T \beta + u_j + v_{jk} + \varepsilon_{jkl} )</td>
<td></td>
</tr>
<tr>
<td>Where ( y_{jkl} ) represents reported individual income without missing data while ( \sigma_u^2, \sigma_v^2, \sigma_{\varepsilon}^2 ) represents the variances at the Province, Neighbourhood / Village and individual levels respectively.</td>
<td></td>
</tr>
<tr>
<td>( \sigma_u^2 = 0.0414 \quad \rho_u = 0.0323 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma_v^2 = 0.0412 \quad \rho_v = 0.0322 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma_{\varepsilon}^2 = 1.1961 \quad \rho_{\varepsilon} = 0.9354 )</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.18: Variance and Intraclass Estimates for the CHNS Data-2011

<table>
<thead>
<tr>
<th>Four Level Model</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_{jklm} = x_{jklm}^T \beta + u_j + v_{jk} + \tau_{jkl} + \epsilon_{jklm} )</td>
<td></td>
</tr>
<tr>
<td>Where ( y_{jklm} ) represents reported individual income without missing data while ( \sigma_u^2, \sigma_v^2, \sigma_{\tau jkl}^2, \sigma_{\epsilon jklm}^2 ) represents the variances at the Province, County-City, Neighbourhood/Village and individual levels respectively.</td>
<td>( \sigma_u^2 = 0.1981 \quad \rho_u = 0.199 )</td>
</tr>
<tr>
<td>( \sigma_v^2 = 0.0000 \quad \rho_v = 0.000 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma_{\tau jkl}^2 = 0.1142 \quad \rho_{\tau} = 0.115 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma_{\epsilon jklm}^2 = 0.6808 \quad \rho_{\epsilon} = 0.686 )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Three Level Model</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_{jkl} = x_{jkl}^T \beta + u_j + v_{jk} + \epsilon_{jkl} )</td>
<td></td>
</tr>
<tr>
<td>Where ( y_{jkl} ) represents reported individual income without missing data while ( \sigma_u^2, \sigma_v^2, \sigma_{\epsilon jkl}^2 ) represents the variances at the Province, Neighbourhood/Village and individual levels respectively.</td>
<td>( \sigma_u^2 = 0.2296 \quad \rho_u = 0.184 )</td>
</tr>
<tr>
<td>( \sigma_v^2 = 0.2604 \quad \rho_v = 0.209 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma_{\epsilon}^2 = 0.0738 \quad \rho_{\epsilon} = 0.059 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma_{\epsilon jkl}^2 = 0.6808 \quad \rho_{\epsilon} = 0.547 )</td>
<td></td>
</tr>
</tbody>
</table>

|      |      |
|      |      |

|      |      |
|      |      |
6.11 Summary of the Analyses

Two separate datasets from the Jamaican and Chinese population were analysed with emphasis of household expenditure of meals purchased away from home and income dynamics, respectively. In the JSCL 2007 analysis, the result provided the evidence that type of weighted models was associated with the proportion of variation in the data hierarchy. The best model had the least AIC and also incorporates all the hierarchy in the survey design.

The findings from the analysis of the China Health Survey Data showed that standard error of estimates is usually lower for parameter estimates from the weighted model than unweighted model especially in this scenario where some of the data values are imputed. Similarly, depending on the profile of the socio-demographics of the survey respondents, there will be varying predictors for probability of response and ultimately the item non-response weight, the inverse of the response probability. This has led to different predictors in the logistics models for years 1989 and 2011. The result also showed that depending on the level of significance, several models can be created.

In summary, the advantage of the item non-response weighted multilevel model was reduced standard error of estimate. In addition, conducting a multilevel model was a right approach to analyse survey data from non-uniform sampling method.
Chapter 7

Discussion, Conclusion and Limitation

7.1 Discussion

The first and second objectives of the study were developed to identify the weighing adjustment method that produces the least bias parameter estimate when addressing the unequal probability of selection and missing data problems in multilevel models. In addition to the least bias and variance, the objectives also included the conditions which are favourable for this least bias estimate in a multilevel modelling.

The findings of the simulation study on the parameter estimates from the multilevel model suggested that reliable estimates are possible when weights are applied to address the unequal probability of selection at each level of the sampling design. This is comparable to the suggestions from DuMouchel and Duncan (1983) as well as Pfefferman (1993) regarding the use of weighting adjustment methods to address the bias from unequal selection probability in regression models involving stratified samples.

Reliable estimates were also observed when adjusting for item non response because the missing data in the sample confirmed the usefulness of item non response weighting adjustments as expressed in Skinner and Arrigo (2011); Bethlehem et.al (2011); Barbara and Stephen(2001). However, sampling weights adjustments for unequal probability of selection of sample units with further
adjustments for item non response produced the least bias estimate from the simulation study. This suggests that better estimates can be produced when both unequal probability of sample selection at each level of the stratification along while also addressing the missing data in the sample. This concept is not common in the literature but serves as a way forward for researchers having this type of problems.

As expected, the findings revealed that the less missing data, the better the estimates. The weighting adjustment methods perform better when the missing data is 20% or less under the Missing at Random (MAR) and also Missing not at Random mechanisms. Performance of the weighting adjustment methods decreases linearly as the percentage of missing increases from 40% to 60%. In addition, the study findings also revealed that adjusting for missing data and unequal sample selection in survey sampling design simultaneously during the estimation of model parameters produces better estimate and performance of the weight adjustment methods.

The earlier findings were only in relation to a fixed intraclass correlation coefficient (ICC). The inclusion of varying ICC values was to address the third study objective of the reliability of estimates from weighted multilevel model under varying ICC values. The findings revealed that at higher ICC values the parameter estimates from the weighting adjustment methods produces smaller root mean square error (RMSE) while at lower ICC values higher root mean square error were produced. This simply implies that the weighting adjustments methods performs better at higher ICC. The ICC values in an indication variation within and between the cluster based on the sample size. This finding is synonymous to the issues in the literature regarding cluster size in multilevel models. The varying sizes of clusters thus influence the variation between and withing clusters and by extension the intraclass correlation coefficient (Hox, 1998, 2010; Snijders and Bosker, 2012).
For the fourth objective of the thesis, which seeks to identify any possible effects of the cluster sizes at level 1 or 2 in a weighted or unweighted multilevel model on the predictor variables in the model, the findings clearly shows that there is no effect.

**7.2 Conclusions**

From the study, the simulations revealed adjusting for unequal probability of selection of units in a multilevel model is necessary more importantly simultaneous adjustments for missing data when the information for such adjustment are available. Researchers conducting multilevel model should be free to apply the method to account for clustering so far the value of the ICC is not zero and the clustering was part of the data design, that was integral in the design information of the survey. Reliable parameter estimates are generally guaranteed when the proportion of missing data in the study variable is 20% or less and weights in the parameter estimation should be incorporated when the information is available.

**7.3 Limitation**

The simulation findings are only limited to one hundred repetitions due to computer resources and time factor to complete the study. Lack of information on the nonrespondents who constitute the unit-non response make it impossible to explore adjustment for unit non-response. This is the reality which many secondary data users are usually faced around the world and did not provide the opportunity for comparison with the other methods.

**7.4 Future Research**

The future research from this study would be comparative analysis of multiple imputation with item non response to determine the effects of both on parameters from weighted multilevel linear model. This will provide guidance on which conditions support weighting adjustment and which conditions different the use of imputation from weighting adjustments.
Bibliography


Appendix A

Simulation Codes

setwd("C:/Users/KOCI 2/Desktop/Y20AGE_SEPT_5_19")

X1=read.csv(file="exampleM0Y20.5002AGE.csv")
X1.2=read.csv(file="exampleM0Y20.1002AGE.csv")
M1=matrix(0,ncol=5,nrow=100)
M1=data.frame(M1)

for(i in 1:50)
{
    M1[i,]=X1[c(3,5,8,11,14),(i+1)]
    M1[(50+i),]=X1.2[c(3,5,8,11,14),(i+1)]
}
colnames(M1)=c("Age","Constant","log U","log V","log E")

write.csv(M1,file="M0Y20AGE.csv")

Y0=read.csv("M0Y20AGE.csv")
#
#
X1=read.csv(file="exampleM1Y20.5002AGE.csv")
X1.2=read.csv(file="exampleM1Y20.1002AGE.csv")
M1=matrix(0,ncol=5,nrow=100)
M1=data.frame(M1)

for(i in 1:50)
{
    M1[i,]=X1[c(3,5,8,11,14),(i+1)]
    M1[(50+i),]=X1.2[c(3,5,8,11,14),(i+1)]
}
colnames(M1)=c("Age","Constant","log U","log V","log E")

write.csv(M1,file="M1Y20AGE.csv")
Y140DA=read.csv("M1Y20AGE.csv")

#
#
#

X2=read.csv(file="exampleM2Y20.5002AGE.csv")
X2.2=read.csv(file="exampleM2Y20.1002AGE.csv")
M2=matrix(0,ncol=5,nrow=100)
M2=data.frame(M2)
for(i in 1:50)
{
    M2[i,]=X2[c(3,5,8,11,14),(i+1)]
    M2[(50+i),]=X2.2[c(3,5,8,11,14),(i+1)]
}
colnames(M2)=c("Age","Constant","log U","log V","log E")
write.csv(M2,file="M2Y20AGE.csv")

Y240DA=read.csv("M2Y20AGE.csv")
#
#
#
setwd("C:/Users/KOCI 2/Desktop/Y60AGE_SEPT_5_19")
X3=read.csv(file="exampleM3Y60.5002AGE.csv")
X3.2=read.csv(file="exampleM3Y60.1002AGE.csv")
M3=matrix(0,ncol=5,nrow=100)
M3=data.frame(M3)
for(i in 1:50)
{
    M3[i,]=X3[c(3,5,8,11,14),(i+1)]
    M3[(50+i),]=X3.2[c(3,5,8,11,14),(i+1)]
}
colnames(M3)=c("Age","Constant","log U","log V","log E")
write.csv(M3,file="M3Y60AGE.csv")
Y340DA=read.csv("M3Y20AGE.csv")
#
#
#
X4=read.csv(file="exampleM4Y20.5002AGE.csv")
X4.2=read.csv(file="exampleM4Y20.1002AGE.csv")
M4=matrix(0,ncol=5,nrow=100)
M4=data.frame(M4)
for(i in 1:50)
{
  M4[i,]=X4[c(3,5,8,11,14),(i+1)]
  M4[(50+i),]=X4.2[c(3,5,8,11,14),(i+1)]
}
colnames(M4)=c("Age","Constant","log U","log V","log E")
write.csv(M4,file="M4Y20AGE.csv")

Y440DA=read.csv("M4DAY40AGE.csv")
#
#
#
X1=read.csv(file="exampleM5Y20.5002AGE.csv")
X1.2=read.csv(file="exampleM5Y20.1002AGE.csv")
M1=matrix(0,ncol=5,nrow=100)
M1=data.frame(M1)
for(i in 1:50)
{
  M1[i,]=X1[c(3,5,8,11,14),(i+1)]
  M1[(50+i),]=X1.2[c(3,5,8,11,14),(i+1)]
}
colnames(M1)=c("Age","Constant","log U","log V","log E")
write.csv(M1,file="M5Y20AGE.csv")
Y540DA=read.csv("M5DAY40AGE.csv")

# # #
setwd("C:/Users/OLUSEGUNISMAIL/Desktop/Combined_files_MDA_MNDA_JULY_17_19")

X6=read.csv(file="exampleM6DAY40.50AGE.csv")
X6.2=read.csv(file="exampleM6DAY40.100AGE.csv")
M6=matrix(0,ncol=5,nrow=100)
M6=data.frame(M6)
for(i in 1:50)
{
    M6[i,]=X6[c(3,5,8,11,14),(i+1)]
    M6[(50+i),]=X6.2[c(3,5,8,11,14),(i+1)]
}
colnames(M6)=c("Age","Constant","log U","log V","log E")
write.csv(M6,file="M6DAY40AGE.csv")

Y640DA=read.csv("M6DAY40AGE.csv")

# # # # #
setwd("C:/Users/OLUSEGUNISMAIL/Desktop/Combined_files_MDA_MNDA_JULY_17_19")

X7=read.csv(file="exampleM7DAY40.50AGE.csv")
X7.2=read.csv(file="exampleM7DAY40.100AGE.csv")
M7=matrix(0,ncol=5,nrow=100)
M7=data.frame(M7)
for(i in 1:50)
{
    M7[i,]=X7[c(3,5,8,11,14),(i+1)]
    M7[(50+i),]=X7.2[c(3,5,8,11,14),(i+1)]
}
colnames(M7)=c("Age","Constant","log U","log V","log E")
write.csv(M7,file="M7DAY40AGE.csv")
Y740DA=read.csv("M7DAY40AGE.csv")

setwd("C:/Users/OLUSEGUNISMAIL/Desktop/Combined_files_MDA_MNDA_JULY_17_19")

X8=read.csv(file="exampleM8DAY40.50AGE.csv")
X8.2=read.csv(file="exampleM8DAY40.100AGE.csv")
M8=matrix(0,ncol=5,nrow=100)
M8=data.frame(M8)
for(i in 1:50)
{
  M8[i,]=X8[c(3,5,8,11,14),(i+1)]
  M8[(50+i),]=X8.2[c(3,5,8,11,14),(i+1)]
}
colnames(M8)=c("Age","Constant","log U","log V","log E")
write.csv(M8,file="M8DAY40AGE.csv")

Y840DA=read.csv("M8DAY40AGE.csv")

setwd("C:/Users/OLUSEGUNISMAIL/Desktop/Combined_files_MDA_MNDA_JULY_17_19")

X9=read.csv(file="exampleM9DAY20.50AGE.csv")
X9.2=read.csv(file="exampleM9DAY20.100AGE.csv")
M9=matrix(0,ncol=5,nrow=100)
M9=data.frame(M9)
for(i in 1:50)
{
  M9[i,]=X9[c(3,5,8,11,14),(i+1)]
  M9[(50+i),]=X9.2[c(3,5,8,11,14),(i+1)]
}
colnames(M9)=c("Age","Constant","log U","log V","log E")
write.csv(M9,file="M9DAY40AGE.csv")
Y940DA=read.csv("M9DAY40AGE.csv")
setwd("C:/Users/OLUSEGUNISMAIL/Desktop/Combined_files_MDA_MNDA_JULY_17_19")

X10=read.csv(file="exampleM10MDAY40.50AGE.csv")
X10.2=read.csv(file="exampleM10MDAY40.100AGE.csv")
M10=matrix(0,ncol=5,nrow=100)
M10=data.frame(M10)
for(i in 1:50)
{
  M10[i,]=X10[c(3,5,8,11,14),(i+1)]
  M10[(50+i),]=X10.2[c(3,5,8,11,14),(i+1)]
}
colnames(M10)=c("Age","Constant","log U","log V","log E")
write.csv(M10,file="M10DAY40AGE.csv")

Y1040DA=read.csv("M10DAY40AGE.csv")
#
#
setwd("C:/Users/OLUSEGUNISMAIL/Desktop/Combined_files_MDA_MNDA_JULY_17_19")

X11=read.csv(file="exampleM11DAY40.50AGE.csv")
X11.2=read.csv(file="exampleM11DAY40.100AGE.csv")
M11=matrix(0,ncol=5,nrow=100)
M11=data.frame(M11)
for(i in 1:50)
{
  M11[i,]=X11[c(3,5,8,11,14),(i+1)]
  M11[(50+i),]=X11.2[c(3,5,8,11,14),(i+1)]
}
colnames(M11)=c("Age","Constant","log U","log V","log E")
write.csv(M11,file="M11DAY40AGE.csv")
Y1140DA = read.csv("M11DAY40AGE.csv")
#
#
setwd("C:/Users/OLUSEGUNISMAIL/Desktop/Combined_files_MDA_MNDA_JULY_17_19")

X12 = read.csv(file="exampleM12DAY40.50AGE.csv")
X12.2 = read.csv(file="exampleM12DAY40.100AGE.csv")
M12 = matrix(0, ncol=5, nrow=100)
M12 = data.frame(M12)
for(i in 1:50)
{
  M12[i,] = X12[c(3,5,8,11,14),(i+1)]
  M12[(50+i),] = X12.2[c(3,5,8,11,14),(i+1)]
}
colnames(M12) = c("Age","Constant","log U","log V","log E")
write.csv(M12, file="M12DAY40AGE.csv")

Y1240DA = read.csv("M12DAY40AGE.csv")
#
#

setwd("C:/Users/OLUSEGUNISMAIL/Desktop/Combined_files_MDA_MNDA_JULY_17_19")

X13 = read.csv(file="exampleM13DAY40.50AGE.csv")
X13.2 = read.csv(file="exampleM13DAY40.100AGE.csv")
M13 = matrix(0, ncol=5, nrow=100)
M13 = data.frame(M13)
for(i in 1:50)
{
  M13[i,] = X13[c(3,5,8,11,14),(i+1)]
  M13[(50+i),] = X13.2[c(3,5,8,11,14),(i+1)]
}
colnames(M13) = c("Age","Constant","log U","log V","log E")
write.csv(M13, file="M13DAY40AGE.csv")
Y1340DA=read.csv("M13DAY40AGE.csv")
#
#
setwd("C:/Users/OLUSEGUNISMAIL/Desktop/Combined_files_MDA_MNDA_ JULY_17_19")

X14=read.csv(file="exampleM14DAY40.50AGE.csv")
X14.2=read.csv(file="exampleM14DAY40.100AGE.csv")
M14=matrix(0,ncol=5,nrow=100)
M14=data.frame(M14)
for(i in 1:50)
{
  M14[i,]=X14[c(3,5,8,11,14),(i+1)]
  M14[(50+i),]=X14.2[c(3,5,8,11,14),(i+1)]
}
colnames(M14)=c("Age","Constant","log U","log V","log E")
write.csv(M14,file="M14DAY40AGE.csv")

Y1440DA=read.csv("M14DAY40AGE.csv")
#
#
setwd("C:/Users/OLUSEGUNISMAIL/Desktop/Combined_files_MDA_MNDA_ JULY_17_19")

X15=read.csv(file="exampleM15DAY40.50AGE.csv")
X15.2=read.csv(file="exampleM15MDAY40.100AGE.csv")
M15=matrix(0,ncol=5,nrow=100)
M15=data.frame(M15)
for(i in 1:50)
{
  M15[i,]=X15[c(3,5,8,11,14),(i+1)]
  M15[(50+i),]=X15.2[c(3,5,8,11,14),(i+1)]
}
colnames(M15)=c("Age","Constant","log U","log V","log E")
write.csv(M15,file="M15DAY40AGE.csv")

Y1540DA=read.csv("M15DAY40AGE.csv")
#
#
setwd("C:/Users/OLUSEGUNISMAIL/Desktop/Combined_files_MDA_MNDA_JULY_17_19")

X16=read.csv(file="exampleM16DAY40.50AGE.csv")
X16.2=read.csv(file="exampleM16DAY40.100AGE.csv")
M16=matrix(0,ncol=5,nrow=100)
M16=data.frame(M16)

for(i in 1:50)
{
    M16[i,]=X16[c(3,5,8,11,14),(i+1)]
    M16[(50+i),]=X16.2[c(3,5,8,11,14),(i+1)]
}

colnames(M16)=c("Age","Constant","log U","log V","log E")
write.csv(M16,file="M16DAY40AGE.csv")

Y1640DA=read.csv("M16DAY40AGE.csv")

for(i in 1:5)
{
    boxplot(Y040DA[,i],Y140DA[,i],Y240DA[,i],Y340DA[,i],Y440DA[,i],Y540DA[,i],Y640DA[,i],
            Y740DA[,i],Y840DA[,i],Y940DA[,i],Y1040DA[,i],Y1140DA[,i],Y1240DA[,i],Y1340DA[,i],Y1440DA[,i],Y1540DA[,i],Y1640DA[,i],main=CM[i])
    abline(h=truth[i],col=2)
}
Appendix B

Weighted Analysis of the JSLC2007

*M0-Unweighted
  gen M0=1

*M1-Dwelling weight
  gen M1=1/pr_dwell_ed

*M2-Enumeration district weight
  gen M2=1/pr_ed_const

*M4-sampling weight from dwelling to ED to constituency
  gen M4= (1/pr_dwell_ed)*(1/pr_ed_const)

logistic tmissing age
  predict pi,
  bysort psu_1:egen avgpi=mean(pi)
  gen M6 =1/avgpi
  table M6
  encode educ1, generate(educ2)
  encode es1, generate(es2)
  encode ocg1, generate(ocg2)

  xtmixed logtmeal|| parish:,|| psu_1:,pweight(M0)
  estat ic
  estat icc

  xtmixed logtmeal|| parish:,pweight(M0)
  estat ic
  estat icc

  xtmixed logtmeal|| psu_1:,pweight(M0)
  estat ic
  estat icc

  xtmixed logtmeal|| parish:,|| psu_1:,pweight(M4)
  estat ic
  estat icc

  xtmixed logtmeal|| psu_1:,pweight(M4)
  estat ic
  estat icc
xtmixed logtmeal|| parish:|| psu_1:,pweight(M6)
estat ic
estat icc

xtmixed logtmeal age|| parish:|| psu_1:,pweight(M0)
estat ic
estat icc

xtmixed logtmeal age|| parish:|| psu_1:,pweight(M4)
estat ic
estat icc

xtmixed logtmeal age|| parish:|| psu_1:,pweight(M6)
estat ic
estat icc

xtmixed logtmeal hhsizel|| parish:|| psu_1:,pweight(M0)
estat ic
estat icc

xtmixed logtmeal hhsizel|| parish:|| psu_1:,pweight(M4)
estat ic
estat icc

xtmixed logtmeal hhsizel|| parish:|| psu_1:,pweight(M6)
estat ic
estat icc

xtmixed logtmeal sex_female || parish:|| psu_1:,pweight(M0)
estat ic
estat icc

xtmixed logtmeal sex_female || parish:|| psu_1:,pweight(M4)
estat ic
estat icc

xtmixed logtmeal sex_female || parish:|| psu_1:,pweight(M6)
estat ic
estat icc

xtmixed logtmeal sex_female || parish:|| psu_1:,pweight(M0)
estat ic
estat icc

xtmixed logtmeal sex_female || parish:|| psu_1:,pweight(M4)
estat ic
estat icc

xtmixed logtmeal sex_female || parish:|| psu_1:,pweight(M6)
estat ic
estat icc
xtmixed logtmeal i.educ2 || parish:|| psu_1:,pweight(M0)
estat ic
estat icc

xtmixed logtmeal i.educ2 || parish:|| psu_1:,pweight(M6)
estat ic
estat icc

xtmixed logtmeal educ1_cxcgce educ1_degreeother || parish:||
psu_1:,pweight(M0)
estat ic
estat icc

xtmixed logtmeal educ1_cxcgce educ1_degreeother || parish:||
psu_1:,pweight(M4)
estat ic
estat icc

xtmixed logtmeal educ1_cxcgce educ1_degreeother || parish:||
psu_1:,pweight(M6)
estat ic
estat icc

xtmixed logtmeal es1_out es1_unemployed|| parish:|| psu_1:,pweight(M4)
estat ic
estat icc

xtmixed logtmeal es1_out es1_unemployed || parish:|| psu_1:,pweight(M6)
estat ic
estat icc

xtmixed logtmeal ocg1_clerk ocg1_crtw ocg1_eo ocg1_lsom ocg1_nc
ocg1_pmoa ocg1_professionals ocg1_swsm ocg1_tap || parish:||
psu_1:,pweight(M0)
estat ic
estat icc

xtmixed logtmeal ocg1_clerk ocg1_crtw ocg1_eo ocg1_lsom ocg1_nc
ocg1_pmoa ocg1_professionals ocg1_swsm ocg1_tap || parish:||
psu_1:,pweight(M4)
estat ic
estat icc
xtmixed logtmeal ocg1_clerk ocg1_crtw ocg1_eo ocg1_lsom ocg1_nc ocg1_pmoa ocg1_professionals ocg1_swsm ocg1_tap || parish:|| psu_1:,pweight(M6)
estat ic
estat icc

xtmixed logtmeal age hhsize1 educ1_cxcgce educ1_degreeother ocg1_clerk ocg1_crtw ocg1_eo ocg1_lsom ocg1_pmoa ocg1_professionals ocg1_swsm ocg1_tap || parish:|| psu_1:,pweight(M6)
estat ic
estat icc

xtmixed logtmeal age hhsize1 educ1_cxcgce educ1_degreeother ocg1_clerk ocg1_crtw ocg1_eo ocg1_lsom ocg1_pmoa ocg1_professionals ocg1_swsm ocg1_tap || parish:|| psu_1:,pweight(M4)
estat ic
estat icc

xtmixed logtmeal age hhsize1 i.sex2 i.educ2 i.es2 i.ocg2 || parish:|| psu_1:,pweight(M0)
estat ic
estat icc

xtmixed logtmeal age hhsize1 i.sex2 i.educ2 i.es2 i.ocg2 || parish:|| psu_1:,pweight(M6)
estat ic
estat icc
Appendix C

CHNS 1989

**MODEL IV**-D1-D2-D3-D4-D5-D6 & **MODEL IV**-E1-E2-E3-E4-E5-E6

**income** has missing income
**indinc** without missing
destring indcome, replace
gen loginA=ln(indcome)
destring indinc, replace
gen loginB=ln(indinc)

gen male = (gender==1)
gen female = (gender==2)
table male
table female

gen Nevermarried =(a8==1)
gen Married =a8==2)
gen Divorced =a8==3)
gen Widowed =a8==4)
gen Separated =a8==5)
gen Unknown =a8==9)
gen Notmarried= Nevermarried + Divorced + Widowed + Separated + Unknown
table Married
table Notmarried

****Table 5.7

***MODEL IV**-D1
xtmixed loginB||t1:||t3:||t4:
estat ic

***MODEL IV**-D2
xtmixed loginB age || t1:||t3:||t4:
estat ic

***MODEL IV**-D3
xtmixed loginB age male || t1:||t3:||t4:
estat ic

***MODEL IV**-D4
xtmixed loginB age male a11 || t1:||t3:||t4:
estat ic
***MODEL IV-D5
xtmixed loginB age male a11 Married || t1:||t3:||t4:
estat ic

***MODEL IV-D6
xtmixed loginB age a11 || t1:||t3:||t4:
estat ic

****Weighted Multilevel Model

logistic miss age a11 Married male
logistic miss age a11 Married female
logistic miss female
predict pi,

bysort t1:egen avgpi=mean(pi)
gen W_jkli =1/avgpi

***Table 5.8
***MODEL IV-E1
xtmixed loginA|| t1:||t3:||t4:,pweight(W_jkli)
estat ic

***MODEL IV-E2
xtmixed loginA age || t1:|| t3:|| t4:,pweight(W_jkli)
estat ic

***MODEL IV-E3
xtmixed loginA age male|| t1:|| t3:||t4:,pweight(W_jkli)
estat ic

***MODEL IV-E4
xtmixed loginA age male a11 || t1:|| t3:||t4:,pweight(W_jkli)
estat ic

***MODEL IV-E5
xtmixed loginA age male a11 Married || t1:|| t3:|| t4:,pweight(W_jkli)
estat ic

***MODEL IV-E6
xtmixed loginA age a11 || t1:|| t3:|| t4:,pweight(W_jkli)
estat ic
Appendix D

CHNS 2011

***MODEL III-F2
xtmixed loginB age || t1:||t4:
estat ic

***MODEL III-F3
xtmixed loginB age male || t1:||t4:
estat ic

***MODEL III-F4
xtmixed loginB age male a11 || t1:||t4:
estat ic

***MODEL III-F5
xtmixed loginB age male a11 Married || t1:||t4:
estat ic

***MODEL III-F6
xtmixed loginB male a11 || t1:||t4:
estat ic

****Weighted Multilevel Model
logistic miss age a11 Married male
logistic miss age a11 Married female
logistic miss female
predict pi,

bysort t1:egen avgpi=mean(pi)
gen W_jkli =1/avgpi
***Table 5.10

***MODEL III-G1
xtmixed login A|| t1:|| t4:, pweight(W_jkli)
estat ic

***MODEL III-G2
xtmixed login A age || t1:|| t4:, pweight(W_jkli)
estat ic

***MODEL III-G3
xtmixed login A age male || t1:|| t3:|| t4:, pweight(W_jkli)
estat ic

***MODEL III-G4
xtmixed login A age male a11|| t1:|| t4:, pweight(W_jkli)
estat ic

***MODEL III-G5
xtmixed login A age male a11 Married || t1:|| t4:, pweight(W_jkli)
estat ic

***MODEL III-G6
xtmixed login A male a11 || t1:|| t4:, pweight(W_jkli)
estat ic
R-CODE FOR DATA GENERATION SIMULATION

setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSVTODAY")

20 PERCENT PROPORTION OF MISSING IN THE DEPENDENT VARIABLE
MISSINGNESS DEPENDS ON AGE

DATA GENERATION BY AGE OF RESPONDENTS

X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")

ser=X$serial
psu=X$psu_1
parish=X$parish
age=X$age

N=length(ser)
Npsu=max(psu)
Npar=max(parish)

# Parameters
M=100

# Creates missingness based on age
Mdelta0=1.9
Mdelta1=-0.03

loMiss=exp(Mdelta0+Mdelta1*age)/(1+exp(Mdelta0+Mdelta1*age))

mean(loMiss)

# mean(loMiss) gives the mean amount of missingness.

Current setup missingness decreases with age.

sigU=3
sigV=2
sigE=1
beta0=6
beta1=-0.08

# Y=matrix(0,ncol=M,nrow=N)
M=matrix(0,nrow=N,ncol=M)

for(i in 1:M)
{
    Upar=rnorm(Npar,0,sigU)
    Vpsu=rnorm(Npsu,0,sigV)
    E=rnorm(N,0,sigE)

    U=Upar[parish]
    V=Vpsu[psu]

    Y[,i]=beta0+beta1*age+U+V+E
    Mi[,i]=rbinom(N,1,loMiss)
    # Mi records which observations are observed/missing
}

YM=Y
YM[Mi==1]=NA

W=matrix(0,ncol=(M+4),nrow=N)
W[,1]=ser
W[,2]=psu
W[,3]=parish
W[,4]=age
W[,5:(M+4)]=YM

Ynam=0
for(i in 1:M) Ynam[i]=paste("y",i,sep="")

YD=data.frame(W)
colnames(YD)=c("serial","psu_1","parish","age",Ynam)
write.csv(YD,"sim_y76011age.csv")
20 PERCENT PROPORTION OF MISSING IN THE DEPENDENT VARIABLE 
MISSINGNESS NOT DEPENDING ON AGE

X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")
ser=X$serial
psu=X$psu_1
parish=X$parish
age=X$age

N=length(ser)
Npsu=max(psu)
Npar=max(parish)

# Parameters
M=100
Miss=0.2

sigU=3
sigV=2
sigE=1

beta0=6
beta1=-0.08

Y=matrix(0,ncol=M,nrow=N)
Mi=matrix(0,nrow=N,ncol=M)

for(i in 1:M)
{
  Upar=rnorm(Npar,0,sigU)
  Vpsu=rnorm(Npsu,0,sigV)
  E=runif(N,0,sigE)

  U=Upar[parish]
  V=Vpsu[psu]
\[
Y[i] = \beta_0 + \beta_1 \cdot \text{age} + U + V + E
\]

\[
M_i = \text{matrix}(\text{rbinom}(N \times M, 1, \text{Miss}), \text{ncol}=M)
\]
#Mi records which observations are observed/missing
}
#

YM = Y
YM[Mi==1] = NA

W = matrix(0, ncol=(M+4), nrow=N)
W[,1] = ser
W[,2] = psu
W[,3] = parish
W[,4] = age
W[,5:(M+4)] = YM

Ynam = 0
for(i in 1:M) Ynam[i] = paste("y", i, sep="")

YD = data.frame(W)
colnames(YD) = c("serial", "psu_1", "parish", "age", Ynam)
write.csv(YD, "sim_y76011age.csv")
MISSINGNESS NOT DEPENDING ON AGE
DATA GENERATED BY SEX OF RESPONDENTS

#setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSVTODAY/Simulation")
X=read.csv("JAM1994TODAY.csv")
File_Number=3
filename=paste("sim_y",File_Number,".csv",sep="")
Name="sex" # covariate dependent
ser=X$serial
psu=X$psu_1
parish=X$parish
variname=as.numeric(X$sex)-1 # covariate dependent
N=length(ser)
Npsu=max(psu)
Npar=max(parish)

# Parameters
M=100 # Total number of simulations
Miss=0.2
sigU=3
sigV=2
sigE=1
beta0=6
beta1=5

Y=matrix(0,ncol=M,nrow=N)
for(i in 1:M)
{
  Upar=rnorm(Npar,0,sigU)
  Vpsu=rnorm(Npsu,0,sigV)
  E=rnorm(N,0,sigE)
  U=Upar[parish]
  V=Vpsu[psu]
Y[,i]=beta0+beta1*variname+U+V+E
}
#

Mi=matrix(rbinom(N*M,1,Miss),ncol=M)
#Mi

YM=Y
YM[Mi==1]=NA

W=matrix(0,ncol=(M+4),nrow=N)
W[,1]=ser
W[,2]=psu
W[,3]=parish
W[,4]=variname
W[,5:(M+4)]=YM

Ynam=0
for(i in 1:M) Ynam[i]=paste("y",i,sep="")

YD=data.frame(YM)
colnames(YD)=c("serial","psu_1","parish",Name,Ynam)
write.csv(YD,filename)
MISSINGNESS NOT DEPENDING ON AGE
DATA GENERATED BY AGE AND SEX OF RESPONDENTS

#setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSVTODAY")
X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")
File_Number=45000
filename=paste("sim_y",File_Number,".csv",sep="")
Name="sex" # covariate dependent
ser=X$serial
psu=X$psu_1
parish=X$parish
variname=as.numeric(X$sex)-1 # covariate dependent
age=X$age
N=length(ser)
Npsu=max(psu)
Npar=max(parish)
# Parameters
M=100 # Total number of simulations
Miss=0.2
sigU=3
sigV=2
sigE=1
beta0=6
beta1=5
betta2=3

Y=matrix(0,ncol=M,nrow=N)
for(i in 1:M){

Upar=rnorm(Npar,0,sigU)
Vpsu=rnorm(Npsu,0,sigV)
E=rnorm(N,0,sigE)

U=Upar[parish]
V=Vpsu[psu]

Y[,i]=beta0+beta1*variname+betta2*age+U+V+E
}
#
Mi = matrix(rbinom(N*M,1,Miss),ncol=M)
#Mi

YM = Y
YM[Mi==1] = NA

W = matrix(0,ncol=(M+5),nrow=N)
W[,1] = ser
W[,2] = psu
W[,3] = parish
W[,4] = variname
W[,5] = age
W[,6:(M+5)] = YM

Ynam = 0
for(i in 1:M) Ynam[i] = paste("y",i,sep="")

YD = data.frame(W)
colnames(YD) = c("serial","psu_1","parish","variname","age","Ynam")
write.csv(YD,filename)
R-CODE TO INVESTIGATE THE EFFECT INTRACLASS CORRELATION
BASED ON DATA SIMULATION -MISSING DEPENDS ON AGE

```r
setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSVOCT192021")

X = read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")
ser = X$serial
psu = X$psu_1
parish = X$parish
age = X$age

N = length(ser)
Npsu = max(psu)
Npar = max(parish)

# Parameters
M = 100
# Creates missingness based on age
Mdelta0 = 1.9
Mdelta1 = -0.03
loMiss = exp(Mdelta0 + Mdelta1 * age) / (1 + exp(Mdelta0 + Mdelta1 * age))
mean(loMiss)
# mean(loMiss) gives the mean amount of missingness. Current setup missingness decreases with age.
sigU = 3
sigV = 2
sigE = 1

beta0 = 6
beta1 = -0.08

#
Y = matrix(0, ncol = M, nrow = N)
Mi = matrix(0, nrow = N, ncol = M)
for (i in 1:M)
{
  Upar = rnorm(Npar, 0, sigU)
}
```

Appendix F
Vpsu=rnorm(Npsu,0,sigV)
E=rnorm(N,0,sigE)

U=Upar[parish]
V=Vpsu[psu]
Y[,i]=beta0+beta1*age+U+V+E
Mi[,i]=rbinom(N,1,loMiss)
# Mi records which observations are observed/missing
}
#
YM=Y
YM[Mi==1]=NA
W=matrix(0,ncol=(M+4),nrow=N)
W[,1]=ser
W[,2]=psu
W[,3]=parish
W[,4]=age
W[,5:(M+4)]=YM

Ynam=0
for(i in 1:M) Ynam[i]=paste("y",i,sep=""")
YD=data.frame(W)
colnames(YD)=c("serial","psu_1","parish","age",Ynam)
write.csv(YD,"sim_y2021101911age.csv")
setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSVOCT192021")

X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")
ser=X$serial
psu=X$psu_1
parish=X$parish
age=X$age

N=length(ser)
Npsu=max(psu)
Npar=max(parish)

# Parameters
M=100
# Creates missingness based on age

Mdelta0=1.9
Mdelta1=-0.03

loMiss=exp(Mdelta0+Mdelta1*age)/(1+exp(Mdelta0+Mdelta1*age))
mean(loMiss)

# mean(loMiss) gives the mean amount of missingness. Current setup missingness decreases with age.

sigU=1
sigV=2
sigE=3

beta0=6
beta1=-0.08

#
Y=matrix(0,ncol=M,nrow=N)
Mi=matrix(0,nrow=N,ncol=M)
for(i in 1:M)
{
    Upar=rnorm(Npar,0,sigU)
    Vpsu=rnorm(Npsu,0,sigV)
    E=rnorm(N,0,sigE)
U = Upar[parish]
V = Vpsu[psu]

Y[,i] = beta0 + beta1 * age + U + V + E
Mi[,i] = rbinom(N, 1, loMiss)
# Mi records which observations are observed/missing
}
#
#
YM = Y
YM[Mi == 1] = NA

W = matrix(0, ncol = (M + 4), nrow = N)
W[, 1] = ser
W[, 2] = psu
W[, 3] = parish
W[, 4] = age
W[, 5:(M + 4)] = YM

Ynam = 0
for (i in 1:M) Ynam[i] = paste("y", i, sep = "")

YD = data.frame(W)
colnames(YD) = c("serial", "psu_1", "parish", "age", Ynam)
write.csv(YD, "sim_y2021101912age.csv")
setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSV/OCT192021")

X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")
ser=X$serial
psu=X$psu_1
parish=X$parish
age=X$age

N=length(ser)
Npsu=max(psu)
Npar=max(parish)

# Parameters
M=100

# Creates missingness based on age
Mdelta0=1.9
Mdelta1=-0.03
loMiss=exp(Mdelta0+Mdelta1*age)/(1+exp(Mdelta0+Mdelta1*age))

mean(loMiss)
# mean(loMiss) gives the mean amount of missingness. Current setup missingness decreases with age.
sigU=5
sigV=4
sigE=3

beta0=6
beta1=-0.08

#
Y=matrix(0,ncol=M,nrow=N)
Mi=matrix(0,nrow=N,ncol=M)

for(i in 1:M)
{
    Upar=rnorm(Npar,0,sigU)
    Vpsu=rnorm(Npsu,0,sigV)
    E=rrnorm(N,0,sigE)

    U=Upar[parish]
V=Vpsu[psu]

Y[,i]=beta0+beta1*age+U+V+E
Mi[,i]=rbinom(N,1,loMiss)
    # Mi records which observations are observed/missing
    
YM=Y
YM[Mi==1]=NA

W=matrix(0,ncol=(M+4),nrow=N)
W[,1]=ser
W[,2]=psu
W[,3]=parish
W[,4]=age
W[,5:(M+4)]=YM

Ynam=0
for(i in 1:M) Ynam[i]=paste("y",i,sep="")

YD=data.frame(W)
colnames(YD)=c("serial","psu_1","parish","age",Ynam)
write.csv(YD,"sim_y2021101913age.csv")
X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")

File_Number=45034

filename=paste("sim_y",File_Number,".csv",sep="")

Name="age" # covariate dependent
ser=X$serial
psu=X$psu_1
parish=X$parish
age=X$age
N=length(ser)
Npsu=max(psu)
Npar=max(parish)

# Parameters
M=100 # Total number of simulations
Miss=0.202412
sigU=3
sigV=2
sigE=1
beta0=6
beta1=-0.08

#
Y=matrix(0,ncol=M,nrow=N)
for(i in 1:M)
{

Upar=rnorm(Npar,0,sigU)
Vpsu=rnorm(Npsu,0,sigV)
E=rnorm(N,0,sigE)
U=Upar[parish]
V=Vpsu[psu]
Y[,i]=beta0+beta1*age+U+V+E
}
#
Mi=matrix(rbinom(N*M,1,Miss),ncol=M)
#Mi
YM=Y
YM[Mi==1]=NA
W=matrix(0,ncol=(M+4),nrow=N)
W[,1]=ser
W[,2]=psu
W[,3]=parish
W[,4]=age
W[,5:(M+4)]=YM
Ynam=0
for(i in 1:M) Ynam[i]=paste("y",i,sep="")
YD=data.frame(W)
colnames(YD)=c("serial","psu_1","parish","age",Ynam)
write.csv(YD, filename)

#setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSVOCT192021")

X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")

File_Number=102020211
filename=paste("sim_y",File_Number,".csv",sep="")

Name="age" # covariate dependent

ser=X$serial
psu=X$psu_1
parish=X$parish
age=X$age
N=length(ser)
Npsu=max(psu)
Npar=max(parish)

# Parameters
M=100 # Total number of simulations
Miss=0.202412
sigU=3
sigV=2
sigE=1
beta0=6
beta1=-0.08

#
Y=matrix(0,ncol=M,nrow=N)
for(i in 1:M)
{
  Upar=rnorm(Npar,0,sigU)
  Vpsu=rnorm(Npsu,0,sigV)
  E=rnorm(N,0,sigE)
  U=Upar[parish]
V=Vpsu[psu]
Y[,i]=beta0+b*age+U+V+E
filename=paste("sim_y",File_Number,".csv",sep="")

Name="age" # covariate dependent

ser=X$serial
psu=X$psu_1
parish=X$parish
age=X$age
N=length(ser)
Npsu=max(psu)
Npar=max(parish)

# Parameters
M=100 # Total number of simulations
Miss=0.202412
sigU=1
sigV=2
sigE=3
beta0=6
beta1=-0.08

#
Y=matrix(0,ncol=M,nrow=N)
for(i in 1:M)
{
  Upar=rnorm(Npar,0,sigU)
  Vpsu=rnorm(Npsu,0,sigV)
  E=rnorm(N,0,sigE)
  U=Upar[parish]
V=Vpsu[psu]
Y[,]=beta0+beta1*age+U+V+E
}
#
Mi=matrix(rbinom(N*M,1,Miss),ncol=M)
#Mi
YM=Y
YM[Mi==1]=NA
W=matrix(0,ncol=(M+4),nrow=N)
W[,]=ser
W[,]=psu
W[,]=parish
W[,]=age
W[,]=YM
Ynam=0
for(i in 1:M) Ynam[i]=paste("y",i,sep="")
YD=data.frame(W)
colnames(YD)=c("serial","psu_1","parish","age",Ynam)
write.csv(YD, filename)
#setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSVOCT192021")

X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")

File_Number=10202021
filename=paste("sim_y",File_Number,".csv",sep="")
Name="age" # covariate dependent
ser=X$serial
psu=X$psu_1
parish=X$parish
age=X$age
N=length(ser)
Npsu=max(psu)
Npar=max(parish)

# Parameters
M=100 # Total number of simulations
Miss=0.202412
sigU=3
sigV=2
sigE=1
beta0=6
beta1=-0.08

#
Y=matrix(0,ncol=M,nrow=N)
for(i in 1:M)
{
    Upar=rnorm(Npar,0,sigU)
    Vpsu=rnorm(Npsu,0,sigV)
E=rnorm(N,0,sigE)
U=Upar[parish]
V=Vpsu[psu]
Y[,i]=beta0+beta1*age+U+V+E
}
#
Mi=matrix(rbinom(N*M,1,Miss),ncol=M)
#Mi
YM=Y
YM[Mi==1]=NA
W=matrix(0,ncol=(M+4),nrow=N)
W[,1]=ser
W[,2]=psu
W[,3]=parish
W[,4]=age
W[,5:(M+4)]=YM
Ynam=0
for(i in 1:M) Ynam[i]=paste("y",i,sep="")
YD=data.frame(W)
colnames(YD)=c("serial","psu_1","parish","age",Ynam)
write.csv(YD, filename)
#setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSVTODAY")

X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")

File_Number=45034

filename=paste("sim_y",File_Number,".csv",sep="")

Name="age" # covariate dependent

ser=X$serial

psu=X$psu_1

parish=X$parish

age=X$age

N=length(ser)

Npsu=max(psu)

Npar=max(parish)

# Parameters

M=100 # Total number of simulations

Miss=0.202412

sigU=3

sigV=2

sigE=1

beta0=6

beta1=-0.08

#

Y=matrix(0,ncol=M,nrow=N)

for(i in 1:M)
{ Upar=rnorm(Npar,0,sigU) Vpsu=rnorm(Npsu,0,sigV) E=rnorm(N,0,sigE) U=Upar[parish] V=Vpsu[psu] Y[,i]=beta0+beta1*age+U+V+E }

Mi=matrix(rbinom(N*M,1,Miss),ncol=M) #Mi YM=Y YM[Mi==1]=NA W=matrix(0,ncol=(M+4),nrow=N) W[,1]=ser W[,2]=psu W[,3]=parish W[,4]=age W[,5:(M+4)]=YM Ynam=0 for(i in 1:M) Ynam[i]=paste("y",i,sep="") YD=data.frame(W) colnames(YD)=c("serial","psu_1","parish","age",Ynam) write.csv(YD, filename)
X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")

File_Number=102020211

filename=paste("sim_y",File_Number,".csv",sep="")

Name="age" # covariate dependent

ser=X$serial
psu=X$psu_1
parish=X$parish
age=X$age
N=length(ser)
Npsu=max(psu)
Npar=max(parish)

# Parameters
M=100 # Total number of simulations
Miss=0.202412
sigU=3
sigV=2
sigE=1
beta0=6
beta1=-0.08

Y=matrix(0,ncol=M,nrow=N)
for(i in 1:M)
{
    Upar=rnorm(Npar,0,sigU)
    Vpsu=rnorm(Npsu,0,sigV)
    E=rnorm(N,0,sigE)
    U=Upar[[parish]]
    V=Vpsu[[psu]]
    Y [,i]=beta0+beta1*age+U+V+E
}
#
Mi=matrix(rbinom(N*M,1,Miss),ncol=M)
#Mi
YM=Y
YM[Mi==1]=NA
W=matrix(0,ncol=(M+4),nrow=N)
W[,1]=ser
W[,2]=psu
W[,3]=parish
W[,4]=age
W[,5:(M+4)]=YM
Ynam=0
for(i in 1:M) Ynam[i]=paste("y",i,sep="")
YD=data.frame(W)
colnames(YD)=c("serial","psu_1","parish","age",Ynam)
write.csv(YD, filename)
#setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSV/OCT192021")

X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")

File_Number=102020212

filename=paste("sim_y",File_Number,".csv",sep="")

Name="age" # covariate dependent

ser=X$serial
psu=X$psu_1
parish=X$parish
age=X$age
N=length(ser)
Npsu=max(psu)
Npar=max(parish)

# Parameters
M=100 # Total number of simulations
Miss=0.202412
sigU=1
sigV=2
sigE=3
beta0=6
beta1=-0.08

Y=matrix(0,ncol=M,nrow=N)
for(i in 1:M)
{
    Upar=rnorm(Npar,0,sigU)
    Vpsu=rnorm(Npsu,0,sigV)
    E=rnorm(N,0,sigE)
    U=Upar[parish]
    V=Vpsu[psu]
    Y[,i]=beta0+beta1*age+U+V+E
}

Mi=matrix(rbinom(N*M,1,Miss),ncol=M)
#Mi
YM=Y
YM[Mi==1]=NA
W=matrix(0,ncol=(M+4),nrow=N)
W[,1]=ser
W[,2]=psu
W[,3]=parish
W[,4]=age
W[,5:(M+4)]=YM
Ynam=0
for(i in 1:M) Ynam[i]=paste("y",i,sep="")
YD=data.frame(W)
colnames(YD)=c("serial","psu_1","parish","age",Ynam)
write.csv(YD, filename)

#setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSVOCT192021")

X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")
ser=X$serial
psu=X$psu_1
parish=X$parish
age=X$age

N=length(ser)
Npsu=max(psu)
Npar=max(parish)

# Parameters

M=100

# Creates missingness based on age

Mdelta0=0
Mdelta1=-0.03

loMiss=exp(Mdelta0+Mdelta1*age)/(1+exp(Mdelta0+Mdelta1*age))
mean(loMiss)

# mean(loMiss) gives the mean amount of missingness. Current setup missingness decreases with age.

sigU=3
sigV=2
sigE=1

beta0=6
beta1=-0.08
#

Y = matrix(0, ncol=M, nrow=N)
Mi = matrix(0, nrow=N, ncol=M)

for(i in 1:M)
{

Upar = rnorm(Npar, 0, sigU)
Vpsu = rnorm(Npsu, 0, sigV)
E = rnorm(N, 0, sigE)

U = Upar[parish]
V = Vpsu[psu]

Y[,i] = beta0 + beta1*age + U + V + E
Mi[,i] = rbinom(N, 1, IoMiss)
# Mi records which observations are observed/missing
}
#

YM = Y
YM[Mi==1] = NA

W = matrix(0, ncol=(M+4), nrow=N)
W[,1] = ser
W[,2] = psu
W[,3]=parish
W[,4]=age
W[,5:(M+4)]=YM

Ynam=0
for(i in 1:M) Ynam[i]=paste("y",i,sep="")

YD=data.frame(W)
colnames(YD)=c("serial","psu_1","parish","age",Ynam)
write.csv(YD,"sim_y102020213age.csv")

#setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSVOCT192021")

X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")
ser=X$serial
psu=X$psu_1
parish=X$parish
age=X$age

N=length(ser)
Npsu=max(psu)
Npar=max(parish)

# Parameters
M=100
# Creates missingness based on age
Mdelta0=0
Mdelta1=-0.03
loMiss=exp(Mdelta0+Mdelta1*age)/(1+exp(Mdelta0+Mdelta1*age))
mean(loMiss)
# mean(loMiss) gives the mean amount of missingness. Current setup missingness decreases with age.

\[
\text{sigU}=1
\]
\[
\text{sigV}=2
\]
\[
\text{sigE}=3
\]
\[
\text{beta0}=6
\]
\[
\text{beta1}=-0.08
\]

\[
\text{Y=matrix}(0,ncol=M,nrow=N)
\]
\[
\text{Mi=matrix}(0,nrow=N,ncol=M)
\]

\[
\text{for(i in 1:M)}
\]
\[
\{
\]
\[
\text{Upar=rnorm(Npar,0,sigU)}
\]
\[
\text{Vpsu=rnorm(Npsu,0,sigV)}
\]
\[
\text{E=rnorm(N,0,sigE)}
\]
\[
\text{U=Upar[parish]}
\]
\[
\text{V=Vpsu[psu]}
\]
\[
\text{Y[,i]=beta0+beta1*age+U+V+E}
\]
\[
\text{Mi[,i]=rbinom(N,1,loMiss)}
\]
\[
\text{# Mi records which observations are observed/missing}
\]
\[
\}
\]
\[
\]
YM=Y
YM[Mi==1]=NA

W=matrix(0,ncol=(M+4),nrow=N)
W[,1]=ser
W[,2]=psu
W[,3]=parish
W[,4]=age
W[,5:(M+4)]=YM

Ynam=0
for(i in 1:M) Ynam[i]=paste("y",i,sep="")

YD=data.frame(W)
colnames(YD)=c("serial","psu_1","parish","age",Ynam)
write.csv(YD,"sim_y102020214age.csv")

#setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSVOCT192021")
X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")
File_Number=1020202117
filename=paste("sim_y",File_Number,".csv",sep="")

Name="age"  # covariate dependent

ser=X$serial
psu=X$psu_1
parish=X$parish
age=X$age
N=length(ser)
Npsu=max(psu)
Npar=max(parish)

# Parameters
M=100 # Total number of simulations
Miss=0.202412
sigU=3
sigV=2
sigE=1
beta0=6
beta1=-0.08

#
Y=matrix(0,ncol=M,nrow=N)
for(i in 1:M)
{
  Upar=rnorm(Npar,0,sigU)
  Vpsu=rnorm(Npsu,0,sigV)
  E=rnorm(N,0,sigE)
  U=Upar[parish]
  V=Vpsu[psu]
  Y[,i]=beta0+beta1*age+U+V+E
}
#
Mi=matrix(rbinom(N*M,1,Miss),ncol=M)
#Mi
YM=Y
YM[Mi==1]=NA
W=matrix(0,ncol=(M+4),nrow=N)
W[,1]=ser
W[,2]=psu
W[,3]=parish
W[,4]=age
W[,5:(M+4)]=YM
Ynam=0
for(i in 1:M) Ynam[i]=paste("y",i,sep="")
YD=data.frame(W)
colnames(YD)=c("serial","psu_1","parish","age",Ynam)
write.csv(YD, filename)

#setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSVOC12021")
X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")
File_Number=1020202128
filename=paste("sim_y",File_Number,".csv",sep="")
Name="age" # covariate dependent

ser=X$serial
psu=X$psu_1
parish=X$parish
age=X$age
N=length(ser)
Npsu = max(psu)
Npar = max(parish)

# Parameters
M = 100  # Total number of simulations
Miss = 0.202412
sigU = 1
sigV = 2
sigE = 3
beta0 = 6
beta1 = -0.08

#
Y = matrix(0, ncol = M, nrow = N)
for (i in 1:M)
{
    Upar = rnorm(Npar, 0, sigU)
    Vpsu = rnorm(Npsu, 0, sigV)
    E = rnorm(N, 0, sigE)
    U = Upar[parish]
    V = Vpsu[psu]
    Y[, i] = beta0 + beta1 * age + U + V + E
}

#
Mi = matrix(rbinom(N*M, 1, Miss), ncol = M)
YM = Y
YM[Mi == 1] = NA
W = matrix(0, ncol = (M + 4), nrow = N)
W[, 1] = ser
W[, 2] = psu
W[, 3] = parish
W[, 4] = age
W[, 5:(M + 4)] = YM
Ynam = 0
for (i in 1:M) Ynam[i] = paste("y", i, sep = "")
YD = data.frame(W)
colnames(YD) = c("serial", "psu_1", "parish", "age", Ynam)
write.csv(YD, filename)
#setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSVOCT192021")

X=read.csv("JAM1994_JULY_5_2019_WITHOUT_WEIGHT.csv")
ser=X$serial
psu=X$psu_1
parish=X$parish
age=X$age

N=length(ser)
Npsu=max(psu)
Npar=max(parish)

# Parameters

M=100

# Creates missingness based on age

Mdelta0=0
Mdelta1=-0.03
loMiss=exp(Mdelta0+Mdelta1*age)/(1+exp(Mdelta0+Mdelta1*age))

mean(loMiss)

# mean(loMiss) gives the mean amount of missingness. Current setup missingness decreases with age.

sigU=3
sigV=2
sigE=1

beta0=6
beta1=-0.08

#

Y=matrix(0,ncol=M,nrow=N)
Mi=matrix(0,nrow=N,ncol=M)

for(i in 1:M)
{

Upar=rnorm(Npar,0,sigU)
Vpsu=rnorm(Npsu,0,sigV)
E=rnorm(N,0,sigE)

U=Upar[parish]
V=Vpsu[psu]
\[ Y_{[,i]} = \beta_0 + \beta_1 \times \text{age} + U + V + E \]
\[ M_{[,i]} = \text{rbinom}(N, 1, \text{loMiss}) \]

# Mi records which observations are observed/missing

YM = Y
YM[Mi==1] = NA

W = matrix(0, ncol=(M+4), nrow=N)
W[,1] = ser
W[,2] = psu
W[,3] = parish
W[,4] = age
W[,5:(M+4)] = YM

Ynam = 0
for (i in 1:M) Ynam[i] = paste("y", i, sep="")

YD = data.frame(W)
colnames(YD) = c("serial", "psu_1", "parish", "age", Ynam)
write.csv(YD, "sim_y1020202139age.csv")

#setwd("C:/users/OLUSEGUNISMAIL/DESKTOP/CSVOCT192021")

X = read.csv("JAM1994_JULY_5_2019WITHOUT_WEIGHT.csv")
ser = X$serial
psu = X$psu_1
parish = X$parish
age = X$age

N = length(ser)
Npsu = max(psu)
Npar = max(parish)

# Parameters

M = 100

# Creates missingness based on age

Mdelta0 = 0
Mdelta1 = -0.03
loMiss = exp(Mdelta0 + Mdelta1 * age) / (1 + exp(Mdelta0 + Mdelta1 * age))

mean(loMiss)

# mean(loMiss) gives the mean amount of missingness. Current setup missingness decreases with age.
sigU=1
sigV=2
sigE=3

beta0=6
beta1=-0.08

#

Y=matrix(0,ncol=M,nrow=N)
Mi=matrix(0,nrow=N,ncol=M)

for(i in 1:M)
{
    Upar=rnorm(Npar,0,sigU)
    Vpsu=rnorm(Npsu,0,sigV)
    E=rnorm(N,0,sigE)
    
    U=Upar[parish]
    V=Vpsu[psu]

    Y[,i]=beta0+beta1*age+U+V+E
    Mi[,i]=rbinom(N,1,loMiss)
    # Mi records which observations are observed/missing
}

#

YM=Y
YM[Mi==1]=NA

W=matrix(0,ncol=(M+4),nrow=N)
W[,1]=ser
W[,2]=psu
W[,3]=parish
W[,4]=age
W[,5:(M+4)]=YM

Ynam=0
for(i in 1:M) Ynam[i]=paste("y",i,sep="")

YD=data.frame(W)
colnames(YD)=c("serial","psu_1","parish","age",Ynam)
write.csv(YD,"sim_y1020202149age.csv")
Appendix G

Table 5.16: Parameter estimates from unweighted multilevel models from the missing mechanism is MAR with missing rates of 20%, 40% and 60% for data simulated from age-only and sex-only covariate models

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_{1,ijkl}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β_{i} Age</td>
<td>UP</td>
<td>-0.08000</td>
<td>-0.08016</td>
<td>0.20000</td>
<td>0.07217</td>
<td>-0.08018</td>
<td>0.22500</td>
<td>0.07221</td>
<td>-0.07997</td>
<td>0.03750</td>
<td>0.07201</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PS</td>
<td>-0.08000</td>
<td>-0.08010</td>
<td>0.12500</td>
<td>0.07212</td>
<td>-0.08015</td>
<td>0.18750</td>
<td>0.07219</td>
<td>-0.07999</td>
<td>0.01250</td>
<td>0.07201</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HS</td>
<td>-0.08000</td>
<td>-0.08027</td>
<td>0.33750</td>
<td>0.07228</td>
<td>-0.08020</td>
<td>0.25000</td>
<td>0.07224</td>
<td>-0.08010</td>
<td>0.12500</td>
<td>0.07216</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PHS</td>
<td>-0.08000</td>
<td>-0.08023</td>
<td>0.28125</td>
<td>0.07225</td>
<td>-0.07860</td>
<td>1.74538</td>
<td>0.07242</td>
<td>-0.08013</td>
<td>0.16375</td>
<td>0.07219</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IR</td>
<td>-0.08000</td>
<td>-0.08016</td>
<td>0.20000</td>
<td>0.07217</td>
<td>-0.08019</td>
<td>0.24250</td>
<td>0.07222</td>
<td>-0.07996</td>
<td>0.04500</td>
<td>0.07201</td>
<td></td>
</tr>
<tr>
<td>Y_{2,ijkl}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α_{i} Sex</td>
<td>UP</td>
<td>5.00000</td>
<td>5.00522</td>
<td>0.10440</td>
<td>0.05698</td>
<td>4.98987</td>
<td>0.20260</td>
<td>0.07230</td>
<td>4.99610</td>
<td>0.07800</td>
<td>0.07970</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PS</td>
<td>5.00000</td>
<td>5.00653</td>
<td>0.13060</td>
<td>0.06266</td>
<td>4.98731</td>
<td>0.25380</td>
<td>0.08272</td>
<td>4.99800</td>
<td>0.04000</td>
<td>0.09846</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HS</td>
<td>5.00000</td>
<td>5.00652</td>
<td>0.13040</td>
<td>0.06609</td>
<td>4.99182</td>
<td>0.16360</td>
<td>0.08399</td>
<td>4.99480</td>
<td>0.10400</td>
<td>0.09295</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PHS</td>
<td>5.00000</td>
<td>5.00870</td>
<td>0.17400</td>
<td>0.07176</td>
<td>4.99069</td>
<td>0.18620</td>
<td>0.09282</td>
<td>4.99770</td>
<td>0.04600</td>
<td>0.11387</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IR</td>
<td>5.00000</td>
<td>5.00503</td>
<td>0.10060</td>
<td>0.05692</td>
<td>4.99020</td>
<td>0.19600</td>
<td>0.07206</td>
<td>4.99571</td>
<td>0.08580</td>
<td>0.08009</td>
<td></td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI = Parameter of interest; WM = Weighting methods; TV = True value; AE = Average Estimate for parameter of interest; PB = Percent bias; RMSE = Root mean square error; UP = Unweighted parameter; PS = ED selection weight; HS = Household selection weight; PHS = ED and Household selection weight; IR = Item non-response weight
Table 5.17: Parameter estimates from unweighted multilevel models from the missing mechanism is MNAR with missing rates of 20%, 40% and 60% for data simulated from age-only and sex-only covariate models

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>20%</th>
<th>40%</th>
<th>60%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y₁β₁</td>
<td>Age</td>
<td>UP</td>
<td>-0.08000</td>
<td>-0.07992</td>
<td>0.10000</td>
<td>0.07194</td>
<td>-0.08014</td>
<td>0.17250</td>
<td>0.07217</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PS</td>
<td>-0.08000</td>
<td>-0.07990</td>
<td>0.12250</td>
<td>0.07192</td>
<td>-0.08000</td>
<td>0.00000</td>
<td>0.07203</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HS</td>
<td>-0.08000</td>
<td>-0.07993</td>
<td>0.08625</td>
<td>0.07195</td>
<td>-0.08013</td>
<td>0.16750</td>
<td>0.07217</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHS</td>
<td>-0.08000</td>
<td>-0.07990</td>
<td>0.12250</td>
<td>0.07193</td>
<td>-0.08002</td>
<td>0.03000</td>
<td>0.07206</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IR</td>
<td>-0.08000</td>
<td>-0.07990</td>
<td>0.12125</td>
<td>0.07194</td>
<td>-0.08014</td>
<td>0.17125</td>
<td>0.07216</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>UP</td>
<td>5.00000</td>
<td>4.99988</td>
<td>0.00240</td>
<td>0.09223</td>
<td>4.99467</td>
<td>0.10660</td>
<td>0.06040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PS</td>
<td>5.00000</td>
<td>4.99877</td>
<td>0.02460</td>
<td>0.09154</td>
<td>4.99460</td>
<td>0.10800</td>
<td>0.06821</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HS</td>
<td>5.00000</td>
<td>4.99672</td>
<td>0.06560</td>
<td>0.10060</td>
<td>4.99104</td>
<td>0.17920</td>
<td>0.06954</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHS</td>
<td>5.00000</td>
<td>4.99559</td>
<td>0.08820</td>
<td>0.10267</td>
<td>4.99074</td>
<td>0.18520</td>
<td>0.06926</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IR</td>
<td>5.00000</td>
<td>4.99968</td>
<td>0.00640</td>
<td>0.09240</td>
<td>4.99469</td>
<td>0.10620</td>
<td>0.06002</td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI = Parameter of interest; WM = Weighting methods; TV = True value; AE = Average Estimate for parameter of interest; PB = Percent bias; RMSE = Root mean square error; UP = Unweighted parameter; PS = ED selection weight; HS = Household selection weight; PHS = ED and Household selection weight; IR = Item non-response weight
Table 5.18 Parameter estimates from weighted multilevel models (scaled method B) from the missing mechanism is MAR with missing rates of 20%, 40% and 60% for each data simulated from sex- age and multivariable covariate models

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>20%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>40%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>60%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_{1,j,k,l}</td>
<td>β_{1}</td>
<td>UP</td>
<td>-0.08000</td>
<td>-0.08016</td>
<td>0.20000</td>
<td>0.07217</td>
<td>-0.08018</td>
<td>0.22500</td>
<td>0.07221</td>
<td>-0.07997</td>
<td>0.03750</td>
<td>0.07201</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSB</td>
<td>-0.08000</td>
<td>-0.08022</td>
<td>0.27875</td>
<td>0.07225</td>
<td>-0.08025</td>
<td>0.30875</td>
<td>0.07228</td>
<td>-0.08010</td>
<td>0.12750</td>
<td>0.072013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRB</td>
<td>-0.08000</td>
<td>-0.08016</td>
<td>0.20500</td>
<td>0.07218</td>
<td>-0.08020</td>
<td>0.24625</td>
<td>0.072257</td>
<td>-0.07996</td>
<td>0.05125</td>
<td>0.072003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UP</td>
<td>5.00000</td>
<td>5.00522</td>
<td>0.10440</td>
<td>0.05698</td>
<td>4.98987</td>
<td>0.20260</td>
<td>0.07230</td>
<td>4.99605</td>
<td>0.07900</td>
<td>0.07970</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSB</td>
<td>5.00000</td>
<td>5.00587</td>
<td>0.11740</td>
<td>0.06978</td>
<td>5.00067</td>
<td>0.01340</td>
<td>0.08015</td>
<td>4.99288</td>
<td>0.14240</td>
<td>0.10404</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRB</td>
<td>5.00000</td>
<td>5.00507</td>
<td>0.10140</td>
<td>0.05678</td>
<td>4.99013</td>
<td>0.19740</td>
<td>0.07227</td>
<td>4.99580</td>
<td>0.08400</td>
<td>0.08009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_{2,j,k,l}</td>
<td>α_{1}</td>
<td>UP</td>
<td>5.00000</td>
<td>4.99780</td>
<td>0.26200</td>
<td>0.06111</td>
<td>4.98966</td>
<td>0.20680</td>
<td>0.05964</td>
<td>4.99225</td>
<td>0.15500</td>
<td>0.08342</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSB</td>
<td>5.00000</td>
<td>4.99392</td>
<td>0.12160</td>
<td>0.06772</td>
<td>4.98847</td>
<td>0.23060</td>
<td>0.07016</td>
<td>4.99220</td>
<td>0.15600</td>
<td>0.09757</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRB</td>
<td>5.00000</td>
<td>4.99871</td>
<td>0.25800</td>
<td>0.06134</td>
<td>4.99999</td>
<td>0.20200</td>
<td>0.0597214</td>
<td>4.99229</td>
<td>0.15420</td>
<td>0.08338</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_{3,j,k,l}</td>
<td>λ_{1}</td>
<td>UP</td>
<td>-0.08000</td>
<td>-0.07982</td>
<td>0.22200</td>
<td>0.00153</td>
<td>-0.08019</td>
<td>0.23750</td>
<td>0.00155</td>
<td>-0.07932</td>
<td>0.84125</td>
<td>0.00250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSB</td>
<td>-0.08000</td>
<td>-0.07996</td>
<td>0.04500</td>
<td>0.00180</td>
<td>-0.08029</td>
<td>0.35625</td>
<td>0.00193</td>
<td>-0.07981</td>
<td>0.21125</td>
<td>0.00282</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRB</td>
<td>-0.08000</td>
<td>-0.07983</td>
<td>0.21625</td>
<td>0.00154</td>
<td>-0.08020</td>
<td>0.24875</td>
<td>0.00155</td>
<td>-0.07934</td>
<td>0.82500</td>
<td>0.00250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UP</td>
<td>5.00000</td>
<td>3.10769</td>
<td>0.00193</td>
<td>-0.00791</td>
<td>1.39615</td>
<td>0.00231</td>
<td>-0.00765</td>
<td>1.93282</td>
<td>0.00285</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSB</td>
<td>-0.00780</td>
<td>1.94744</td>
<td>0.00230</td>
<td>-0.00793</td>
<td>1.70513</td>
<td>0.00266</td>
<td>-0.00757</td>
<td>2.89436</td>
<td>0.00342</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRB</td>
<td>-0.00780</td>
<td>1.60769</td>
<td>0.00200</td>
<td>-0.00773</td>
<td>0.86795</td>
<td>0.00229</td>
<td>-0.00765</td>
<td>1.88423</td>
<td>0.00304</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_{4,j,k,l}</td>
<td>τ_{1}</td>
<td>UP</td>
<td>-0.02500</td>
<td>-0.03161</td>
<td>26.44281</td>
<td>0.04828</td>
<td>-0.02472</td>
<td>1.12492</td>
<td>0.06602</td>
<td>-0.01943</td>
<td>22.27556</td>
<td>0.08408</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSB</td>
<td>-0.02500</td>
<td>-0.03123</td>
<td>24.90088</td>
<td>0.05913</td>
<td>-0.01762</td>
<td>29.50520</td>
<td>0.07778</td>
<td>-0.01761</td>
<td>29.55320</td>
<td>0.10627</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRB</td>
<td>-0.02500</td>
<td>-0.02768</td>
<td>10.70384</td>
<td>0.04671</td>
<td>-0.02724</td>
<td>9.84680</td>
<td>0.06560</td>
<td>-0.01977</td>
<td>20.92624</td>
<td>0.10420</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI = Parameter of interest; WM = Weighting methods; TV = True value; AE = Average Estimate for parameter of interest; PB = Percent bias; RMSE = Root mean square error; UP = Unweighted parameter; PHSB = ED and Household selection weight using scaled Method B; IRB = Item non-response weight using scaled Method B
Table 5.19: Parameter estimates from weighted multilevel models (scaled method B) from the missing mechanism is MNAR with missing rates of 20%, 40% and 60% for each data simulated from sex- age and multivariable covariate models

<table>
<thead>
<tr>
<th></th>
<th>Scaled Method B - Missing Not At Random (MNAR)</th>
<th>20%</th>
<th>40%</th>
<th>60%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DGM</td>
<td>PI</td>
<td>WI</td>
<td>TV</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
</tr>
<tr>
<td></td>
<td>DGM</td>
<td>PI</td>
<td>WI</td>
<td>TV</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Y_{1,jkl} )</td>
<td><strong>( \beta_1 )</strong></td>
<td>UP</td>
<td>-0.08000</td>
<td>-0.08016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSB</td>
<td>-0.08000</td>
<td>-0.07987</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRB</td>
<td>-0.08000</td>
<td>-0.07993</td>
</tr>
<tr>
<td>( Y_{2,jkl} )</td>
<td><strong>( \alpha_1 )</strong></td>
<td>UP</td>
<td>0.00000</td>
<td>4.99988</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSB</td>
<td>0.00000</td>
<td>4.99646</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRB</td>
<td>0.00000</td>
<td>4.99963</td>
</tr>
<tr>
<td>( Y_{3,jkl} )</td>
<td><strong>( \lambda_1 )</strong></td>
<td>UP</td>
<td>0.00000</td>
<td>5.01085</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSB</td>
<td>0.00000</td>
<td>5.00451</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRB</td>
<td>0.00000</td>
<td>5.01092</td>
</tr>
<tr>
<td>( Y_{4,jkl} )</td>
<td><strong>( \tau_1 )</strong></td>
<td>UP</td>
<td>-0.00780</td>
<td>-0.08034</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHSB</td>
<td>-0.00780</td>
<td>-0.08040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRB</td>
<td>-0.00780</td>
<td>-0.08035</td>
</tr>
</tbody>
</table>

**DGM** = Data generation model; **PI**= Parameter of interest; **WI** = Weighting methods; **TV** = True value; **AE**=Average Estimate for parameter of interest; **PB**=Percent bias; **RMSE**=Root mean square error; **UP**= Unweighted parameter; **PHSB**=ED and Household selection weight using scaled Method B; **IRB**=Item non-response weight using scaled Method B
Table 5.20: Random Component – (U) estimates from weighted multilevel models from the missing mechanism is MAR with missing rates of 20%, 40% and 60% for each data simulated from age-only; sex-only; sex- age and multivariable covariate models

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20%</td>
<td></td>
<td></td>
<td>40%</td>
<td></td>
<td></td>
<td>60%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_{1jk}</td>
<td>U</td>
<td>3.00000</td>
<td>2.79059</td>
<td>6.98043</td>
<td>0.62431</td>
<td>2.77502</td>
<td>7.49919</td>
<td>0.59267</td>
<td>2.75840</td>
<td>8.05339</td>
<td>0.62480</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>40%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_{2jk}</td>
<td>U</td>
<td>3.00000</td>
<td>2.89286</td>
<td>3.57140</td>
<td>0.68486</td>
<td>2.83333</td>
<td>5.5564</td>
<td>0.54833</td>
<td>2.82946</td>
<td>5.68467</td>
<td>0.61433</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>40%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_{3jk}</td>
<td>U</td>
<td>3.00000</td>
<td>2.91979</td>
<td>2.67352</td>
<td>0.67522</td>
<td>2.63838</td>
<td>12.0537</td>
<td>0.69843</td>
<td>2.87170</td>
<td>4.27656</td>
<td>0.62435</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>40%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_{4jk}</td>
<td>U</td>
<td>3.00000</td>
<td>2.98337</td>
<td>0.55439</td>
<td>0.64646</td>
<td>2.83840</td>
<td>5.38662</td>
<td>0.52276</td>
<td>2.83979</td>
<td>5.34023</td>
<td>0.52276</td>
<td></td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI = Parameter of interest; WM = Weighting methods; TV = True value; AE = Average Estimate for parameter of interest; PB = Percent bias; RMSE = Root mean square error; UP = Unweighted parameter; PS = ED selection weight; HS = Household selection weight; PHS = ED and Household selection weight; IR = Item non-response weight
Table 5.21: Random Component – (U)estimates from weighted multilevel models from the missing mechanism is MNAR with missing rates of 20%, 40% and 60% for each data simulated from age-only; sex-only; sex-age and multivariable covariate models

<table>
<thead>
<tr>
<th></th>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>20% AE</th>
<th>20% PB</th>
<th>20% RMSE</th>
<th>40% AE</th>
<th>40% PB</th>
<th>40% RMSE</th>
<th>60% AE</th>
<th>60% PB</th>
<th>60% RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_{1,kl}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>UP</td>
<td>3.00000</td>
<td>2.73964</td>
<td>6.87881</td>
<td>0.61629</td>
<td>2.77502</td>
<td>7.49919</td>
<td>0.59267</td>
<td>2.75840</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PS</td>
<td>3.00000</td>
<td>2.88882</td>
<td>3.07085</td>
<td>0.57631</td>
<td>2.86711</td>
<td>4.42962</td>
<td>0.57631</td>
<td>2.85078</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HS</td>
<td>3.00000</td>
<td>2.79362</td>
<td>6.87940</td>
<td>0.61636</td>
<td>2.77547</td>
<td>7.48446</td>
<td>0.61636</td>
<td>2.75823</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PHS</td>
<td>3.00000</td>
<td>2.90971</td>
<td>3.00970</td>
<td>0.56892</td>
<td>2.90971</td>
<td>3.00970</td>
<td>0.56892</td>
<td>2.88340</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IR</td>
<td>3.00000</td>
<td>2.81065</td>
<td>6.31162</td>
<td>0.60665</td>
<td>2.79295</td>
<td>6.90175</td>
<td>0.60665</td>
<td>2.77519</td>
</tr>
<tr>
<td>Y_{2,kl}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>UP</td>
<td>3.00000</td>
<td>2.74367</td>
<td>8.54445</td>
<td>0.61951</td>
<td>2.91563</td>
<td>2.81226</td>
<td>0.56826</td>
<td>2.82772</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PS</td>
<td>3.00000</td>
<td>2.83938</td>
<td>5.35402</td>
<td>0.56669</td>
<td>3.01499</td>
<td>0.49967</td>
<td>0.54326</td>
<td>2.92767</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HS</td>
<td>3.00000</td>
<td>2.74339</td>
<td>8.55373</td>
<td>0.61953</td>
<td>2.91528</td>
<td>2.82415</td>
<td>0.56816</td>
<td>2.82757</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PHS</td>
<td>3.00000</td>
<td>2.85233</td>
<td>4.92232</td>
<td>0.55221</td>
<td>3.00803</td>
<td>0.26772</td>
<td>0.56826</td>
<td>2.82767</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IR</td>
<td>3.00000</td>
<td>2.75983</td>
<td>8.00566</td>
<td>0.60982</td>
<td>2.93357</td>
<td>2.21430</td>
<td>0.56106</td>
<td>2.84941</td>
</tr>
<tr>
<td>Y_{3,kl}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IR</td>
<td>3.00000</td>
<td>2.76906</td>
<td>7.69814</td>
<td>0.62002</td>
<td>2.81763</td>
<td>6.07915</td>
<td>0.62977</td>
<td>2.86691</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PS</td>
<td>3.00000</td>
<td>2.85435</td>
<td>4.85486</td>
<td>0.59124</td>
<td>2.91242</td>
<td>2.91925</td>
<td>0.58849</td>
<td>2.96299</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HS</td>
<td>3.00000</td>
<td>2.75175</td>
<td>8.27549</td>
<td>0.61953</td>
<td>2.91528</td>
<td>2.82415</td>
<td>0.56816</td>
<td>2.82757</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PHS</td>
<td>3.00000</td>
<td>2.87759</td>
<td>4.08024</td>
<td>0.55221</td>
<td>3.00803</td>
<td>0.26772</td>
<td>0.55514</td>
<td>2.95245</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IR</td>
<td>3.00000</td>
<td>2.76906</td>
<td>7.69814</td>
<td>0.62002</td>
<td>2.81763</td>
<td>6.07915</td>
<td>0.62977</td>
<td>2.86691</td>
</tr>
<tr>
<td>Y_{4,kl}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IR</td>
<td>3.00000</td>
<td>2.72323</td>
<td>9.25580</td>
<td>0.57939</td>
<td>2.88236</td>
<td>3.92133</td>
<td>0.52686</td>
<td>2.68868</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PS</td>
<td>3.00000</td>
<td>2.81028</td>
<td>6.32394</td>
<td>0.53045</td>
<td>2.96001</td>
<td>1.33288</td>
<td>0.50942</td>
<td>3.02418</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HS</td>
<td>3.00000</td>
<td>2.72218</td>
<td>9.26060</td>
<td>0.57938</td>
<td>2.92609</td>
<td>2.64651</td>
<td>0.52087</td>
<td>2.91029</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PHS</td>
<td>3.00000</td>
<td>2.83986</td>
<td>5.33798</td>
<td>0.52232</td>
<td>3.01210</td>
<td>0.40318</td>
<td>0.50176</td>
<td>3.02238</td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI= Parameter of interest; WM =Weighting methods; TV = True value; AE=Average Estimate for parameter of interest; PB=Percent bias; RMSE=Root mean square error; UP= Unweighted parameter; PS=ED selection weight; HS =Household selection weight ; PHS=ED and Household selection weight; IR=Item non-response weight
Table 5.22: Random Component – (V) estimates from weighted multilevel models from the missing mechanism is MAR with missing rates of 20%, 40% and 60% for each data simulated from age-only; sex-only; sex-age and multivariable covariate models

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>20% AE</th>
<th>20% PB</th>
<th>20% RMSE</th>
<th>40% AE</th>
<th>40% PB</th>
<th>40% RMSE</th>
<th>60% AE</th>
<th>60% PB</th>
<th>60% RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_{1,jkl}</td>
<td>V</td>
<td>UP</td>
<td>2.00000</td>
<td>2.00857</td>
<td>0.42850</td>
<td>0.10514</td>
<td>1.98272</td>
<td>0.86377</td>
<td>0.12605</td>
<td>1.97657</td>
<td>1.17128</td>
<td>0.12260</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PS</td>
<td>2.00000</td>
<td>1.89841</td>
<td>5.07936</td>
<td>0.15227</td>
<td>1.88268</td>
<td>5.86606</td>
<td>0.17510</td>
<td>1.87206</td>
<td>6.39685</td>
<td>0.18152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HS</td>
<td>2.00000</td>
<td>2.03399</td>
<td>1.69963</td>
<td>0.10871</td>
<td>2.02013</td>
<td>1.00631</td>
<td>0.12396</td>
<td>2.03909</td>
<td>1.95468</td>
<td>0.12255</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHS</td>
<td>2.00000</td>
<td>1.87424</td>
<td>6.28776</td>
<td>0.17484</td>
<td>1.86405</td>
<td>6.79764</td>
<td>0.19266</td>
<td>1.85858</td>
<td>7.07098</td>
<td>0.19735</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IR</td>
<td>2.00000</td>
<td>1.99196</td>
<td>0.40187</td>
<td>0.10455</td>
<td>1.96644</td>
<td>1.67798</td>
<td>0.12779</td>
<td>1.95963</td>
<td>2.01840</td>
<td>0.12629</td>
</tr>
<tr>
<td>Y_{2,jkl}</td>
<td>V</td>
<td>UP</td>
<td>2.00000</td>
<td>1.99666</td>
<td>0.16699</td>
<td>0.11751</td>
<td>2.00100</td>
<td>0.05004</td>
<td>0.11747</td>
<td>1.98896</td>
<td>0.55202</td>
<td>0.12971</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PS</td>
<td>2.00000</td>
<td>1.89629</td>
<td>5.18554</td>
<td>0.15779</td>
<td>1.89786</td>
<td>5.10675</td>
<td>0.16646</td>
<td>1.87630</td>
<td>6.18491</td>
<td>0.17620</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HS</td>
<td>2.00000</td>
<td>2.02293</td>
<td>1.14635</td>
<td>0.11818</td>
<td>2.03906</td>
<td>1.95286</td>
<td>0.12223</td>
<td>2.05005</td>
<td>2.50248</td>
<td>0.13621</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHS</td>
<td>2.00000</td>
<td>1.87958</td>
<td>6.02123</td>
<td>0.17530</td>
<td>1.88542</td>
<td>5.72896</td>
<td>0.18355</td>
<td>1.85910</td>
<td>7.04523</td>
<td>0.19698</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IR</td>
<td>2.00000</td>
<td>1.97935</td>
<td>1.03275</td>
<td>0.11825</td>
<td>1.98350</td>
<td>0.82477</td>
<td>0.11755</td>
<td>1.97180</td>
<td>1.41001</td>
<td>0.13212</td>
</tr>
<tr>
<td>Y_{3,jkl}</td>
<td>V</td>
<td>UP</td>
<td>2.00000</td>
<td>1.99770</td>
<td>0.11524</td>
<td>0.10063</td>
<td>1.99792</td>
<td>0.10387</td>
<td>0.12271</td>
<td>2.02233</td>
<td>1.11633</td>
<td>0.11250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PS</td>
<td>2.00000</td>
<td>1.88470</td>
<td>5.76513</td>
<td>0.15587</td>
<td>1.93038</td>
<td>3.48115</td>
<td>0.17136</td>
<td>1.91467</td>
<td>4.26627</td>
<td>0.14411</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HS</td>
<td>2.00000</td>
<td>2.02419</td>
<td>1.20958</td>
<td>0.10187</td>
<td>2.03491</td>
<td>1.74545</td>
<td>0.12562</td>
<td>2.08334</td>
<td>4.16713</td>
<td>0.13766</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHS</td>
<td>2.00000</td>
<td>1.86629</td>
<td>6.68553</td>
<td>0.17228</td>
<td>1.90298</td>
<td>4.85076</td>
<td>0.17355</td>
<td>1.89319</td>
<td>5.34038</td>
<td>0.16578</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IR</td>
<td>2.00000</td>
<td>1.98157</td>
<td>0.92170</td>
<td>0.10204</td>
<td>1.98168</td>
<td>0.91598</td>
<td>0.12297</td>
<td>2.00505</td>
<td>0.25264</td>
<td>0.10988</td>
</tr>
<tr>
<td>Y_{4,jkl}</td>
<td>V</td>
<td>UP</td>
<td>2.00000</td>
<td>1.99531</td>
<td>0.23431</td>
<td>0.11620</td>
<td>2.01754</td>
<td>0.87711</td>
<td>0.11024</td>
<td>2.01054</td>
<td>0.52703</td>
<td>0.12009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PS</td>
<td>2.00000</td>
<td>1.88854</td>
<td>5.57323</td>
<td>0.16214</td>
<td>1.91407</td>
<td>4.29626</td>
<td>0.13747</td>
<td>1.89035</td>
<td>5.48268</td>
<td>0.15736</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HS</td>
<td>2.00000</td>
<td>2.02195</td>
<td>1.09772</td>
<td>0.11705</td>
<td>2.05584</td>
<td>2.79176</td>
<td>0.12079</td>
<td>2.06568</td>
<td>3.28403</td>
<td>0.13208</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHS</td>
<td>2.00000</td>
<td>1.87292</td>
<td>6.35387</td>
<td>0.17729</td>
<td>1.90079</td>
<td>4.96067</td>
<td>0.15193</td>
<td>1.87705</td>
<td>6.14769</td>
<td>0.17235</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IR</td>
<td>2.00000</td>
<td>1.97865</td>
<td>1.06764</td>
<td>0.11717</td>
<td>2.00085</td>
<td>0.04253</td>
<td>0.10876</td>
<td>1.99383</td>
<td>0.30869</td>
<td>0.11883</td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI= Parameter of interest; WM = Weighting methods; TV = True value; AE=Average Estimate for parameter of interest; PB=Percent bias; RMSE=Root mean square error; UP= Unweighted parameter; PS= ED selection weight; HS = Household selection weight; PHS = ED and Household selection weight; IR= Item non-response weight
Table 5.23: Random Component – (V) estimates from weighted multilevel models from the missing mechanism is MNAR with missing rates of 20%, 40% and 60% for each data simulated from age-only; sex-only; sex-age and multivariable covariate models

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>20%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>40%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>60%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_{1,jkl}</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DGM</td>
<td>PI</td>
<td>WM</td>
<td>TV</td>
<td>20%</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
<td>40%</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
<td>60%</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
</tr>
<tr>
<td>Y_{2,jkl}</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DGM</td>
<td>PI</td>
<td>WM</td>
<td>TV</td>
<td>20%</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
<td>40%</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
<td>60%</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
</tr>
<tr>
<td>Y_{3,jkl}</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DGM</td>
<td>PI</td>
<td>WM</td>
<td>TV</td>
<td>20%</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
<td>40%</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
<td>60%</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
</tr>
<tr>
<td>Y_{4,jkl}</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DGM = Data generation model; PI= Parameter of interest; WM =Weighting methods; TV = True value; AE=Average Estimate for parameter of interest; PB=Percent bias; RMSE=Root mean square error; UP= Unweighted parameter; PS=ED selection weight; HS =Household selection weight ; PHS=ED and Household selection weight; IR=Item non-response weight

191
Table 5.24: Random Component – (E) estimates from weighted multilevel models from the missing mechanism is MNAR with missing rates of 20%, 40% and 60% for each data simulated from age-only; sex-only; sex-age and multivariable covariate models

<table>
<thead>
<tr>
<th>DGM</th>
<th>PI</th>
<th>WM</th>
<th>TV</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>20%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>40%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
<th>60%</th>
<th>AE</th>
<th>PB</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y_{1.jkl}</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DGM</td>
<td>PI</td>
<td>WM</td>
<td>TV</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
<td>20%</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
<td>40%</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
<td>60%</td>
<td>AE</td>
<td>PB</td>
</tr>
<tr>
<td></td>
<td>UP</td>
<td>1.0000</td>
<td>0.99771</td>
<td>0.22861</td>
<td>0.01928</td>
<td>1.00121</td>
<td>0.12051</td>
<td>0.02401</td>
<td>0.99948</td>
<td>0.05192</td>
<td>0.02905</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PS</td>
<td>1.0000</td>
<td>0.99777</td>
<td>0.22334</td>
<td>0.02083</td>
<td>1.00237</td>
<td>0.23671</td>
<td>0.02597</td>
<td>0.99873</td>
<td>0.12691</td>
<td>0.03158</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HS</td>
<td>1.0000</td>
<td>0.94512</td>
<td>5.48835</td>
<td>0.05852</td>
<td>0.92731</td>
<td>7.26903</td>
<td>0.07696</td>
<td>0.88745</td>
<td>11.25451</td>
<td>0.11642</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PHS</td>
<td>1.0000</td>
<td>0.99806</td>
<td>0.19441</td>
<td>0.02329</td>
<td>0.99913</td>
<td>0.08733</td>
<td>0.02354</td>
<td>0.99791</td>
<td>0.20925</td>
<td>0.03467</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IR</td>
<td>1.0000</td>
<td>0.99758</td>
<td>0.24180</td>
<td>0.01935</td>
<td>1.00149</td>
<td>0.14909</td>
<td>0.02412</td>
<td>0.99929</td>
<td>0.07129</td>
<td>0.02906</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y_{2.jkl}</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DGM</td>
<td>PI</td>
<td>WM</td>
<td>TV</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
<td>20%</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
<td>40%</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
<td>60%</td>
<td>AE</td>
<td>PB</td>
</tr>
<tr>
<td></td>
<td>UP</td>
<td>1.0000</td>
<td>0.99771</td>
<td>0.22861</td>
<td>0.01928</td>
<td>1.00121</td>
<td>0.12051</td>
<td>0.02401</td>
<td>0.99948</td>
<td>0.05192</td>
<td>0.02905</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PS</td>
<td>1.0000</td>
<td>0.99777</td>
<td>0.22334</td>
<td>0.02083</td>
<td>1.00237</td>
<td>0.23671</td>
<td>0.02597</td>
<td>0.99873</td>
<td>0.12691</td>
<td>0.03158</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HS</td>
<td>1.0000</td>
<td>0.94512</td>
<td>5.48835</td>
<td>0.05852</td>
<td>0.92731</td>
<td>7.26903</td>
<td>0.07696</td>
<td>0.88745</td>
<td>11.25451</td>
<td>0.11642</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PHS</td>
<td>1.0000</td>
<td>0.99806</td>
<td>0.19441</td>
<td>0.02329</td>
<td>0.99913</td>
<td>0.08733</td>
<td>0.02354</td>
<td>0.99791</td>
<td>0.20925</td>
<td>0.03467</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IR</td>
<td>1.0000</td>
<td>0.99758</td>
<td>0.24180</td>
<td>0.01935</td>
<td>1.00149</td>
<td>0.14909</td>
<td>0.02412</td>
<td>0.99929</td>
<td>0.07129</td>
<td>0.02906</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y_{3.jkl}</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DGM</td>
<td>PI</td>
<td>WM</td>
<td>TV</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
<td>20%</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
<td>40%</td>
<td>AE</td>
<td>PB</td>
<td>RMSE</td>
<td>60%</td>
<td>AE</td>
<td>PB</td>
</tr>
<tr>
<td></td>
<td>UP</td>
<td>1.0000</td>
<td>0.99548</td>
<td>0.45197</td>
<td>0.01847</td>
<td>0.99171</td>
<td>0.82910</td>
<td>0.02359</td>
<td>0.98645</td>
<td>1.35494</td>
<td>0.03403</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PS</td>
<td>1.0000</td>
<td>0.99389</td>
<td>0.61110</td>
<td>0.01906</td>
<td>0.99010</td>
<td>0.99041</td>
<td>0.02542</td>
<td>0.97871</td>
<td>2.12910</td>
<td>0.03531</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HS</td>
<td>1.0000</td>
<td>0.94112</td>
<td>5.88814</td>
<td>0.06193</td>
<td>0.91883</td>
<td>8.11740</td>
<td>0.08436</td>
<td>0.87113</td>
<td>12.88666</td>
<td>0.13098</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PHS</td>
<td>1.0000</td>
<td>0.99273</td>
<td>0.72710</td>
<td>0.02256</td>
<td>0.98893</td>
<td>1.10673</td>
<td>0.02881</td>
<td>0.97540</td>
<td>2.45967</td>
<td>0.03809</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IR</td>
<td>1.0000</td>
<td>0.99549</td>
<td>0.45127</td>
<td>0.01856</td>
<td>0.99177</td>
<td>0.82305</td>
<td>0.02369</td>
<td>0.98282</td>
<td>1.71769</td>
<td>0.02963</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**DGM =** Data generation model; **PI =** Parameter of interest; **WM =** Weighting methods; **TV =** True value; **AE =** Average Estimate for parameter of interest; **PB =** Percent bias; **RMSE =** Root mean square error; **UP =** Unweighted parameter; **PS =** ED selection weight; **HS =** Household selection weight; **PHS =** ED and Household selection weight; **IR =** Item non-response weight.