2016

Robust inference in poverty mapping

Sumonkanti Das

University of Wollongong

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

School of Mathematics and Applied Statistics

Robust Inference in Poverty Mapping

Sumonkanti Das

"This thesis is presented as part of the requirements for the award of the Degree of Doctor of Philosophy of the University of Wollongong"

April 2016

Thesis submitted in fulfilment of the requirements for the degree

Doctor of Philosophy

on the topic

Robust Inference in Poverty Mapping

to the

School of Mathematics and Applied Statistics

Faculty of Engineering and Information Sciences

at the University of Wollongong

submitted by

Sumonkanti Das

Supervisor:

Prof. Raymond Chambers, PhD (University of Wollongong)

Co-supervisor:

Prof. David Steel, PhD (University of Wollongong)

April 2016

Dedicated to my daughter ***Arushi Prachi Das*** *born on June 11, 2015*

ABSTRACT

Small area estimation (SAE) methods are widely used for estimating poverty indicators at finer levels of a country's geography. Three unit-level SAE techniques – the ELL method (Elbers, Lanjouw, and Lanjouw, 2003), also known as the World Bank method, the Empirical Best Prediction (EBP) method (Molina and Rao, 2010) and the M-Quantile (MQ) method (Tzavidis *et al.*, 2008) have all been used to estimate micro-level FGT poverty indicators (Foster, Greer, and Thorbecke, 1984). These methods vary in terms of their underlying model assumptions particularly differences in consideration of random effects. This thesis provides results from a numerical comparison of the statistical performance of these three methodologies in the context of a realistic simulation scenario based on a recent Bangladesh poverty study. This comparison study shows that the ELL method is the better performer in terms of relative bias but also significantly underestimates the MSEs of its small area poverty estimates when its underlying area homogeneity assumption is violated.

A modified MSE estimation method for ELL-type poverty estimates is therefore developed in this thesis. This method is robust to the presence of significant unexplained between-area variability in the income distribution. This ELL-based MSE estimation methodology is based on a separate bootstrap procedure for MSE estimation, where a correction factor is

used to generate cluster-specific random errors that capture the potential between-area variability unaccounted for by the explanatory variables in the ELL regression model.

A further issue is that in realistic applications of the ELL method, the cluster-specific (level-two) and household-specific (level-one) random errors are typically assumed to be homoskedastic and heteroskedastic respectively. The standard approach to modeling heteroskedasticity of household-level errors in the ELL method is to use a parametric logistic function (called the “alpha” model). Trying to find an adequate set of explanatory variables to include in a regression model for heteroskedastic errors is difficult, and so the alpha model is prone to misspecification. In this thesis we therefore propose a semi-parametric approach to estimating the level-one heteroskedastic error variances, based on a stratified method of moments approach that does not require explanatory variables to be specified.

We also develop a new small area poverty estimation method based on the smearing-based prediction method proposed by Chambers and Dunstan (1986) for estimating a finite population distribution function conforming to a 2-level superpopulation model with unknown heteroskedasticity at level-one as in the ELL methodology. Estimated level-specific error variances are combined with this smearing approach, which is then used in a non-parametric bootstrap procedure to obtain the small area estimates of interest along with their mean squared errors. A modified version of this smearing method that is robust to when between-area variability has been incorrectly ignored is also described.

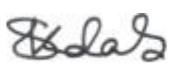
Finally, we apply these new methods for robust MSE estimation and poverty estimation under heteroskedasticity to an actual poverty mapping study of Bangladesh. This application shows that the ideas developed in this thesis have practical, as well as theoretical, impact.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my supervisor Prof. Raymond Chambers (University of Wollongong, Australia) for his encouraging guidance, thoughtful comments, great enthusiasm and invaluable patience during the long journey of my study. I wish to express my gratefulness to Prof. Chambers for the opportunity to be closely in touch with his profound understanding of theoretical and applied statistics. I would like to acknowledge University of Wollongong (Australia) for the financial support through International Postgraduate Tuition Award, University Postgraduate Award, and part-time teaching opportunity, and Shahjalal University of Science & Technology (Bangladesh) for providing the opportunity to complete the study. I would also want to send my appreciation to all academic and administrative staffs of National Institute for Applied Statistics Research Australia (NIASRA) for their cordial assistance and friendly behaviour. I would like to thank Prof. Stephen Haslett (Australian National University, Australia) for providing access to the Bangladesh datasets. I owe much more than what I can express here to my wife Kakoli Rani Bhowmik for her love, support, and understand with her grace during all these years I spent my nights and weekends with books and computers. I would like to thank my family back home for their sacrifices, commitment to my education, and unconditional love. To my friends and colleagues particularly Dr. Chandra Gulati, Dr. Payam Mokhtarian, and Mossamet Kamrun Nesa with her family, I say thank you for valuing our companionship and inspiring me to complete my study.

STATMENT OF ORIGINALITY

I, Sumonkanti Das, declare that this thesis, submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Mathematics and Applied Statistics, Faculty of Engineering and Information Sciences, University of Wollongong, is solely my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.



(Sumonkanti Das)

April 2016

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF NOTATIONS AND SYMBOLS

i, j, k	Index for area (level-three), cluster (level-two), & household (HH, level-one)
y_{ijk}	Response value for k^{th} HH of j^{th} cluster in i^{th} area
$\mathbf{x}_{ijk}$	Vector of explanatory information for k^{th} HH of j^{th} cluster in i^{th} area
m_{ijk}	Number of individuals in k^{th} HH of j^{th} cluster in i^{th} area
$\mathbf{y}$	Vector of response variable
$\mathbf{X}$	Matrix of explanatory variables
$\boldsymbol{\beta}_{(l)}$	Vector of regression coefficients under l-level model
ε_{ijk}	HH-level random error
u_{ij}	Cluster-level random effect
η_i	Area-level random effect
e_{ijk}	HH-level marginal error, $y_{ijk} - \mathbf{x}_{ijk}^T \beta_{(l)}$
$\sigma_{\varepsilon(l)}^2$	Variance of homoskedastic (HM) ε_{ijk} under l-level model
$\sigma_{u(l)}^2$	Variance of HM u_{ij} under l-level model with HM ε_{ijk}
$\sigma_{\eta(l)}^2$	Variance of HM η_i under l-level model with HM ε_{ijk}
$\sigma_{e(l)}^2$	Variance of HM e_{ijk} under l-level model
$\sigma_{\varepsilon(l).ijk}^{2(ht)}$	Variance of heteroskedastic (HT) ε_{ijk} under l-level model
$\sigma_{u(l)}^{2(ht)}$	Variance of HT u_{ij} under l-level model with HT ε_{ijk}
$\sigma_{\eta(l)}^{2(ht)}$	Variance of HT η_i under l-level model with HT ε_{ijk}
s_i	Index set of sampled quantities in area i
r_i	Index set of non-sampled quantities in area i
$\hat{\boldsymbol{\beta}}_{(l)}$	Ordinary least squares estimator (OLSE) of $\boldsymbol{\beta}_{(l)}$

$\hat{\boldsymbol{\beta}}_{(l)}^{gls}$	Generalized least squares (GLS) estimator of $\boldsymbol{\beta}_{(l)}$ (BLUE) for HM ε_{ijk}
$\hat{\boldsymbol{\beta}}_{(l)}^{gls(ht)}$	GLS estimator of $\boldsymbol{\beta}_{(l)}$ (BLUE) for HT ε_{ijk}
t	Quantile (usually poverty line)
α	Parameter of FGT poverty index, α =0,1,2
$F_i(t)$	Distribution function (DF) at quantile t for i^{th} area
$F_{0i}(t)$	Head Count Rate (HCR) at poverty line t for i^{th} area $\left(F_{0i}(t)=F_i(t)\right)$
$F_{1i}(t)$	Poverty Gap (PG) at poverty line t for i^{th} area
$F_{2i}(t)$	Poverty Severity (PS) at poverty line t for i^{th} area
$\hat{F}_{\alpha i}^{ELL.2L}$	ELL estimator of $F_{\alpha i}(t)$ for i^{th} area based on 2-level model (HH, cluster)
$\hat{F}_{\alpha i}^{ELL.3L}$	ELL estimator of $F_{\alpha i}(t)$ for i^{th} area based on 3-level model (HH, cluster, area)
$\hat{F}_{\alpha i}^{EBP}$	EBP estimator of $F_{\alpha i}(t)$ for i^{th} area based on 2-level model (HH, area)
$\hat{F}_{\alpha i}^{MQ}$	MQ estimator of $F_{\alpha i}(t)$ for i^{th} area based on MQ model (HH, area)
$mse\{.\}$	Mean squared error (MSE) estimator of $\hat{F}_{\alpha i}^{.}$
$I(.)$	Indicator function
$\bar{Y}_i$	Mean of response variable for i^{th} area
$\mathrm{Var}_{(l)}\left(\bar{Y}_i\right)$	Variance of $\bar{Y}_i$ under l-level model
$\hat{\mathrm{V}}_{(l)}\left(\bar{Y}_i\right)$	Variance estimator of $\bar{Y}_i$ under l-level model
$E_l[.]$	Expectation of an estimator under l-level model
MELL	Modified ELL MSE estimator
$\boldsymbol{\alpha}$	Vector of regression coefficients of "alpha" model for heteroskedasticity
$\hat{\sigma}_{\varepsilon(l).ijk}^{2.ELL}$	Estimator of $\sigma_{\varepsilon(l).ijk}^{2(ht)}$ via ELL approach
$\hat{\sigma}_{\varepsilon(l).ijk}^{2.STR}$	Estimator of $\sigma_{\varepsilon(l).ijk}^{2(ht)}$ via stratification-based moment approach
PELL	ELL estimators based on parametric bootstrap for HM/HT ε_{ijk}
NPELL	ELL estimators based on non-parametric bootstrap (NPB) for HM ε_{ijk}
SPELL	ELL estimator based on semi-parametric bootstrap for HT ε_{ijk}
CDSM	Smearing based CD estimator via NPB for HM/HT ε_{ijk}
CDMC	Monte Carlo simulation based CD estimator via NPB for HM/HT ε_{ijk}

CHAPTER ONE

1. Introduction

The demand for official statistics on disaggregated sub-populations has increased remarkably in recent years. Typically, such sub-populations are the micro-level administrative units of a country or a large region. Classical direct estimation methods, for example design-based Horvitz-Thompson estimation (Horvitz and Thompson, 1952) or model-assisted generalized regression estimation (Särndal *et al.*, 2003), can only be regarded as efficient if domain-specific sample sizes are large. When the target domains are lower level administrative units, which are usually ignored in the survey sampling design, domain-specific sample sizes are usually too small for reliable inference. In these situations, small area estimation (SAE) methods can be used to estimate the parameters of interest with greater accuracy in these target domains.

SAE methods use regression models to obtain estimates of a parameter for a particular small area by so-called "borrowing strength from other areas". This is typically achieved by assuming that the same regression model holds in every small area, and that there are area-specific random effects and/or contextual effects that account for between-area variability in the model residuals. These models can be specified using area-level or unit-level data, and the approach to SAE based on them is usually referred to as model-based SAE (Rao, 2003). The most extensively used model-based SAE methods are empirical best linear unbiased prediction (EBLUP), empirical Bayes or best (EB) methods and hierarchical Bayes (HB) methods. Generally, development of these methods is developed mainly for inference about small area linear parameters (e.g. means or totals) of the target variable.

1.1 Small Area Poverty Estimation

One of the most important practical applications of model-based SAE methods is in poverty mapping. Poverty indicators are typically complex non-linear functions of welfare variables (Betti *et al.*, 2006) and so classical small area methodologies for totals or means cannot be straightforwardly applied to poverty estimation. The most commonly used poverty indicators are poverty incidence (also known as head count rate/ratio, HCR), poverty gap (PG) and poverty severity (PS). These indicators are often referred to as FGT measures, after Foster, Greer and Thorbecke (1984) where they were first suggested. Let E_{ik} be the per capita household expenditure of household (HH) k in area i and let t be an externally determined poverty line. The FGT poverty indicators for area i are then defined as $F_{\alpha i} = N_i^{-1} \sum_{k=1}^{N_i} \left(\frac{t - E_{ik}}{t} \right)^{\alpha} I\left(E_{ik} < t \right)$ with $\alpha = 0,1,2$ where N_i is the total number of HHs in area i. These indicators are non-linear functions of the finite population income distribution in an area and so the main problem when estimating them is to find an efficient estimator of the corresponding finite population small area income distribution function, which we denote by $F_i(t) = N_i^{-1} \sum_{k=1}^{N_i} I\left(E_{ik} < t \right)$. As an aside, we note that $F_i(t) = F_{0i}$.

The first systematic approach to using unit-level survey and census data model-based SAE methods for poverty estimation is described in Elbers, Lanjouw, and Lanjouw (2003). Their approach is often also referred to as the World Bank method or the ELL method. The ELL method utilizes survey data to develop a nested error regression model (Battese *et al.*, 1988) with HHs at level-one and clusters at level-two. Simulated census values of the welfare variable (*E*) are generated from the fitted model based on known census values of the model covariates using either a parametric or a semi-parametric bootstrap procedure. Area-specific values of poverty indices are then calculated using these simulated census values by suitable aggregation and the procedure is repeated a large number of times. The average and the variance of the simulated estimates for an area are considered as the corresponding area estimate and mean squared error (MSE). Note that under the ELL approach, the fitted regression model for *E* includes a random effect for a cluster but not one for an area. That is, the approach assumes between-cluster heterogeneity as well as between-area homogeneity.

An alternative SAE approach to poverty mapping is described in Molina and Rao (2010) and is based on the empirical Bayes or best prediction (EBP) approach to SAE. This method estimates the poverty indices of interest via Monte Carlo simulation of their empirical best predictors, i.e. the conditional expectations of these indices given the observed sample values of *E*. In particular, the approach assumes that a suitably transformed value of the welfare variable *E* (typically its logarithm) follows a nested error regression model with normally distributed random errors. The method generates simulated values of *E* by making independent draws from the conditional distribution of the non-observed values of *E* given the observed values of *E* and the associated population covariates or auxiliary variables. The main difference between the EBP method and the ELL method is that the former assumes between-area heterogeneity and (typically) between-cluster homogeneity. The EBP method also depends on the assumption that random errors in the model are Gaussian, though this assumption has been relaxed subsequently (Diallo and Rao, 2014; Elbers and Van der Weide, 2014).

Another approach to SAE uses M-quantile models (Breckling and Chambers, 1988) to characterise between-area variability, as first proposed by Chambers and Tzavidis (2006) and extended to poverty estimation by Tzavidis *et al.* (2008). Unlike the ELL and EBP methods, MQ method is free from parametric distributional assumptions and automatically provides outlier-robust inference. It also has the advantage that the fitted M-quantile model does not depend on the small area geography, so there are no boundary issues when distinguishing clusters from target small areas as in the ELL and EBP methods.

The ELL method is by far the most widely used SAE method in poverty estimation (it has been applied in over 60 countries), and is particularly favoured in developing countries due to its theoretical and practical flexibility (Elbers and Van der Weide, 2014). In comparison to the EBP and MQ methods, MSE estimation for ELL-based poverty estimates is also straightforward, requiring comparatively less time and computational resources. However, this method has also been criticised because of its dual assumption of cluster-heterogeneity and area-homogeneity when calculating small area poverty estimates and their estimated MSEs. In practice it is extremely unlikely that the area-homogeneity assumption holds, and in such situation the ELL method leads to unbiased poverty estimates with underestimated MSE (Tarozzi and Deaton, 2009). Such underestimated MSEs have the potential to give policy makers the impression that the ELL poverty estimates are more reliable than they actually are, and hence lead to incorrect decisions based on identification of the most vulnerable small areas where more aid may be required.

A number of approaches to overcoming this problem have been proposed by Elbers *et al.* (2008). Since the estimated location effect (determined by overall variation at higher levels than HH-level) cannot be separated into area-level and cluster-level effects in the ELL method, one has to assume that this effect is either entirely a cluster-level effect (optimistic assumption) or entirely an area-level effect (conservative assumption). Tarozzi and Deaton (2009) comment that the "conservative" assumption could lead to imprecise and unusable MSE estimates, while the "optimistic" assumption is obviously necessary for the validity of the ELL methodology, but leads to estimated MSEs that are biased low. At present, there is no accepted way has been established yet to follow in such situations where area-variability in the welfare data cannot be adequately explained by the explanatory and contextual variables.

1.2 Contribution of the Thesis

This thesis makes four contributions to our understanding of the ELL method for poverty estimation and how it should be applied. The first contribution, detailed in **Chapter Three**, is to present a detailed numerical comparison of the statistical performance of the ELL, EBP and MQ approaches in the context of a realistic simulation built on a recent Bangladesh poverty study. An important feature of this study is that it emulates a common characteristic of poverty data in developing countries, where within-cluster variability in the distribution of the HH income (or welfare) variable *E* dominates its between-cluster variability.

The next contribution, set out in **Chapter Four**, is to develop a MSE estimation method for ELL-type poverty estimates that is robust to the presence of significant unexplained between-area variability in the distribution of *E*. This builds on the observation that in the numerical study described in the previous chapter, MSE is shown to be underestimated when the underlying area-homogeneity assumption is violated. The basic idea behind the proposed MSE estimation method is to develop a correction factor that ensures a robust variance estimator of the area-specific mean is unbiased under a true 3-level model and also approximately unbiased under a 2-level working model. The proposed modification to the ELL MSE estimation method is then to base MSE estimation on a separate bootstrap procedure that uses the correction factor to generate cluster-specific random errors that more realistically capture the potential between-area variability unaccounted for by the explanatory variables included in the ELL regression model. This new ELL-based MSE estimation method is referred to as robustified or modified ELL (MELL) methodology in what follows.

The third contribution is based on the fact that in any realistic application of the ELL method, cluster-specific (level-two) random errors are typically assumed to be homoskedastic (i.e. the 'location effect' common to all HHs in the same cluster has the same distribution irrespective of cluster) but HH-specific (level-one) errors are allowed to be heteroskedastic (Elbers *et al.*, 2003). In particular, it is standard to model the variances of level-one errors using a parametric logistic function (called the "alpha" model) based on the assumption that these variances are a monotone smooth function of one or more explanatory variables. Modelling heteroskedasticity in this way is known to be difficult, and prone to model misspecification due to its dependency on an adequate set of explanatory variables that can explain the heteroskedasticity. Consequently this method requires a careful search for potential explanators or their transformations. Moreover, most poverty mapping studies that use this model (BBS and UNWFP, 2004; World Bank, 2013; Haslett, 2013) have found that its explanatory power (r-squared value) is usually very small.

In this thesis we propose a semi-parametric approach to estimating the level-one heteroskedastic error variances using a stratification-based method of moments (MOM) approach, referred to as STR in subsequent chapters. This has the advantage of not requiring specification of a set of suitable explanators for the level-one heteroskedasticity. An iterative generalized least square (IGLS) method is then used to determine the level-one error variances used in the ELL simulations.

The thesis also explores an alternative to the ELL method for estimating area-specific FGT poverty indicators. This uses the smearing-based prediction method proposed by Chambers and Dunstan (1986) (hereafter referred as CD) for model-based estimation of the finite population distribution function under a linear superpopulation model. Lombardía *et al.* (2005) have utilized the CD approach to estimate the distribution function for a finite uncorrelated population with unknown heteroskedastic errors, and Chambers and Pratesi (2013) have suggested how the CD approach can be adapted for poverty estimation in the context of a predictive model with area specific random effects model (Rao, 2003) or a M-quantile model (Chambers and Tzavidis, 2006).

This thesis develops a new small area poverty estimation method following this CD approach by assuming a finite population conforming to a 2-level superpopulation model with unknown heteroskedasticity at level-one as in the ELL methodology. In this method, the homoscedastic level-two error variance is estimated first and then the level-one heteroskedastic error variances are estimated utilizing the proposed STR semi-parametric estimation method. These estimated variances are then combined with the CD smearing approach to estimate the small area distribution function of E. Area-specific poverty measures and their MSEs are finally calculated via a non-parametric bootstrap procedure that resamples from this estimated small area distribution function. Note that like the ELL method, this CD based method will also produce underestimated MSEs if between-area variability is ignored. Consequently, it needs to be modified using the same MELL-based approach to overcome this issue.

The fourth, and final, contribution of this thesis is to adapt the methodology on robust MSE estimation and poverty estimation under heteroskedasticity developed in earlier chapters to an actual poverty mapping study, in this case one that was recently carried out in Bangladesh. This application is important since it shows that the ideas developed in this thesis have practical, as well as theoretical, impact.

1.3 Thesis Outline

This thesis is organized into seven chapters. A review of related literature covering the main research areas of the thesis set out in **Chapter Two**. This starts with an overview of recent developments and directions in the field of small area estimation, followed by a discussion on the application of SAE methods in poverty estimation. The review also covers variance component estimation, modelling heteroskedasticity in multilevel populations, and the use of the bootstrap for estimating the MSEs of estimates of complex non-linear parameters.

Chapter Three compares the ELL, EBP and MQ methods from both a theoretical as well as a numerical perspective. A number of numerical experiments have already been done to compare these three poverty estimation methods (Molina and Rao, 2010; Tzavidis, *et al.*, 2013; Souza *et al.*, 2015), but clear advice on how one should go about selecting an appropriate SAE method in a real situation is not yet available. The chapter uses a model-based simulation study based on real Bangladesh data to illustrate how an appropriate SAE method for poverty mapping can be determined in a developing country context. The study also clearly demonstrates that all the three methods can lead to underestimated MSEs in this case.

Variance component estimation via MOM is first discussed in **Chapter Four** and then a robust estimator of the variance of an area-specific mean is developed for the situation where a 3-level model is misspecified due to ignoring the third level. Based on this robust variance estimator, standard ELL methodology is then modified to obtain a robust estimate of MSE when the area-homogeneity assumption of ELL is violated. Note that this development (which includes a parametric bootstrap variation) is based on an assumption of homoskedastic random errors at all levels. A number of model-based simulation studies are then used to illustrate the performances of the standard and modified ELL estimators of MSE.

Chapter Five then addresses the issue of level-one heteroskedasticity in implementation of the ELL method. A stratification-based approach (STR) to accounting for this heteroskedasticity is proposed and compared with the "alpha model" approach of the ELL methodology in this situation. This comparison allows for different types of heteroskedasticity functions (monotone and non-monotone). In addition, the impact of misspecification of the heteroskedasticity model on both the ELL and CD-based poverty estimation methods (CDSM) is examined numerically under different hypothetical but practical scenarios. Finally, following the non-parametric bootstrap procedure of Marchetti *et al.* (2012), an easy to compute Monte Carlo simulation-based (CDMC) alternative to the computationally intensive CDSM method is proposed and is also investigated in these numerical studies.

Chapter Six brings together these research threads of the previous two chapters to develop modified versions of ELL- and CD-based methods that allow for heteroskedastic level-one errors, and lead to robust MSE estimates. These methods are then applied in a Bangladesh poverty mapping study. In particular, the Bangladesh Population and

Housing Census 2000 and the Household Income and Expenditure Survey 2000 are used to compare the different estimation methods.

Finally, the thesis concludes with a summary of its major findings, and some possible extensions and directions for future research in **Chapter Seven**. Proofs for all the theoretical results discussed in the paper, including variance component estimation under homoskedasticity and heteroskedasticity, and MSE estimation via the modified ELL method, are presented in appendices.

CHAPTER TWO

2. Background and Literature Review

This chapter provides an overview of small area estimation (SAE), recent developments in SAE theory and its application to complex poverty analysis. Standard small area methodologies for linear parameters are discussed first in order to motivate SAE methods, and the application of these methods to estimation of poverty indicators with their extensions is then reviewed.

2.1 Small Area Estimation

The term "small area" is typically used to describe geographic sub-areas of a country (e.g. states, counties, districts, municipalities) or small groups within a population of interest, often defined in terms of demographic, socio-economic and other characteristics (e.g. by sex, age, race, income status, business group). The sample sizes of these areas in national surveys are usually too small for adequately precise estimation based on methods that are considered appropriate at population level. Such estimates are often referred to as direct estimates. The standard direct estimators of a population quantity of interest are usually asymptotically design-unbiased or design-consistent but with large sampling errors for areas with small sample sizes. Use of indirect estimation methods can improve sub-population estimates in these cases. These are methods that use a model for the population characteristic of interest to generate estimates that have much smaller variances than the direct estimates. Since the model parameters are usually estimated from the entire sample (which is usually large), this approach is often said to “borrow strength” across all the small areas in order to improve the precision of the estimate for a given (small) area. These methods are reviewed in Ghosh and Rao (1994), Pfeffermann (2002, 2013), Rao (2003), Jiang and Lahiri (2006), and Rao and Molina (2015).

Traditional direct estimators for small areas are based on design-based (Cochran, 1977) and model-assisted (Särndal *et al.*, 1992) approaches to survey inference, while indirect estimators use the model-based approach (Brewer, 1963; Royall, 1970; Valliant *et al.*, 2000). Under the design-based approach a probability sample is drawn and the values of the target variable Y are considered to be fixed characteristics of the population elements. That is, the known sample values of Y (denoted as $\mathbf{y}_s$) and the unknown non-sample values of Y (denoted as $\mathbf{y}_r$) are treated as fixed constants. In contrast, the model-based approach starts from the assumption that the distribution of population values of Y is the realisation of a stochastic process which can be modelled. Provided the sampling method is non-informative for the parameters of this model, a sample drawn from the population will then have the same distribution of Y for the sampled and non-sampled observations, and so the model parameters can be estimated from the sample.

A direct small area estimator such as the Horvitz-Thompson estimator (Horvitz and Thompson, 1952) of the total of Y for an area i $\left(Y_i\right)$ uses only the sample Y values within area i. As a consequence, the variance of this estimator is of order $O\left(n_i^{-1}\right)$ where n_i is the sample size in area i. This can be large when n_i is small. A further problem is that no direct estimates are possible for areas with zero sample size. Design-based model-assisted estimation methods can be employed to increase the efficiency of the Horvitz-Thompson estimator by incorporating auxiliary information correlated with Y via a regression model. The Generalized Regression (GREG) estimation is an example of the design-based model-assisted approach to SAE (Särndal *et al.*, 1992; Rao, 2003). For a binary response variable, a logistic GREG (LGREG) estimator can be used, based on a logistic model formulation (Lehtonen and Veijanen, 1998). Lehtonen *et al.* (2009) also propose a number of semi-direct estimators based on the model calibration approach of Wu and Sitter (2001) and on the GREG. Model calibration has also been used to estimate small area poverty incidence in Lehtonen and Vejanen (2012).

In the model-based approach to SAE, a regression model for Y is first developed based on the survey data and known auxiliary variables. The predicted values of Y generated by this model and the auxiliary information are then used to predict linear parameters such as the means or totals for the target small areas. These SAE methods can be split into two groups according to their use of implicit and explicit models. Methods that use implicit models link related small areas through supplementary data from census and/or administrative records, whereas methods that use explicit models account for variability between small areas and between units in the areas through variation in the auxiliary data. In many cases, these models also include random area effects that enable them to also account for between-area variability that is not present in the auxiliary data.

2.1.1 Classical Small Area Estimation

Synthetic estimators, composite estimators and James-Stein estimators are some common examples of classical small area estimators that are based on implicit models. **Synthetic estimation** was first used by the US National Center for Health Statistics (1968) in order to calculate state-level disability estimates. Its basic assumption is that small areas belonging to a larger region should have the same characteristics as the larger region. Reliable direct estimates of these stable characteristics for the larger area can then be used to define an indirect estimator for the smaller areas (Gonzalez, 1973). Synthetic estimators are based on the implicit model that the distribution of the survey variable Y is the same in every small area once their differences in these characteristics are appropriately accounted for. These types of small area estimates are widely used by practitioners, in large part because they do not assume explicit models (Datta, 2009), being essentially a form of standardisation. They are easy and inexpensive to calculate but are essentially biased unless their assumption of similar small areas making up the larger area is true. Various synthetic estimation approaches and their application to small area problems are discussed by Levy (1979) and Rao (2003).

A **composite estimator** is defined as the weighted average of a direct estimator and a synthetic estimator, and serves as a way of balancing the large standard error of the direct estimator and the bias of synthetic estimator (Ghosh and Rao, 1994). The rational here is that the synthetic estimator outperforms the simple direct estimator in terms of smaller mean squared error (MSE) when area-level sample sizes are small, whereas the direct estimator similarly outperforms the synthetic estimator when the sample sizes are large. The **James-Stein estimator** (James and Stein, 1961), also known as the "shrinkage estimator" is a special case of a composite estimator with weights that minimizes the MSE of this estimator (Efron and Moris, 1973). Like synthetic estimators, composite estimators provide biased but precise estimates depending on appropriate weight selection.

2.1.2 Random Effects Model-based Small Area Estimation

The assumption that the different small areas making up a larger area are similar is usually false, even after controlling for known auxiliary variables. Consequently, many SAE methods use random area effects to account for this between-area heterogeneity. These random effects models can be classified into two broad groups on the basis of the type of data that are available for the small areas. Unit-level models assume that unit-specific values of a study variable and associated unit-specific as well as area-specific covariates are available (Battese *et al.*, 1988); while area-level models only require access to area-specific direct estimates and their associated area-specific covariates (Fay and Herriot, 1979). Both unit-specific and area-specific covariates are typically collected from census and geographic databases (e.g. GIS). In many cases these models are linear in the available covariates, and are therefore special cases of the linear mixed model (Rao, 2003). The main approaches to SAE built around these models are based on best linear unbiased prediction (BLUP) or its empirical version (EBLUP), empirical Bayes or empirical best (EB) estimation, and hierarchical Bayes (HB) methods. Further details on SAE methods using BLUP, EBLUP, EB and HB are available in Ghosh and Rao (1994), Rao (2003), Chambers and Clark (2012), and Rao and Molina (2015).

In most SAE applications the parameter of interest is the area-specific mean $\left(\bar{Y}_i\right)$ or total $\left(Y_i\right)$ for a target small domain. The first estimators of these parameters under a random effects (mixed) model were based on the BLUP method proposed by Henderson (1950) which assumes a linear mixed model with known variance components. In reality, variance components are unknown and can be estimated by maximum likelihood (ML) or restricted maximum likelihood (REML), or the method of moments (MOM). Substitution of the estimated variance components in the BLUP estimator leads to the

EBLUP estimator (Harville, 1991), which can be expressed as a weighted combination of a direct estimator and a regression-synthetic estimator. This idea is used in both unit-level and area-level SAE methods.

The basic area-level small area model is the Fay-Herriot (FH) model. It is typically used for estimation of area-specific totals and means. Fay and Herriot (1979) first used this model to estimate per capita income for small places (population less than 1,000) in the US based on data collected in the 1970 Census of Population and Housing. The FH model was also used to estimate mean income under the EURAREA project (http://www.statistics.gov.uk/eurarea) in Europe. The main problem associated with this method is its assumption that the sampling error variances of the direct area estimates are known. These sampling error variances are often estimated from unit-level survey data from the respective small areas. Since the area-specific sample sizes are usually very small, such estimated sampling error variances can be very unstable. As a consequence, it is usually recommended that such estimated sampling variances are first smoothed using an appropriate statistical smoothing technique (Wolter, 1985). The basic FH model has been extended by several authors to include non-normality of the random errors, spatial correlation of the random area-effects, and presence of outliers in the survey data (Rao and Molina, 2015).

When auxiliary information is available for all individuals (units) of a population, a unit-level random effects model can be used for SAE. This is typically based on the nested-error regression model, first suggested by Battese, Harter and Fuller (1988) (hereafter BFH). These authors used this unit-level model to predict corn and soybean crop areas for 12 counties in north-central Iowa using farm-interview data and satellite information. Under the assumption of normally distributed level-specific random errors, the regression parameters and variance components can be easily estimated via ML or

REML. The area mean or total can then be estimated using EBLUP, EB and HB methods based on a linear mixed model and assuming a suitable distribution (typically Gaussian) for the random effects (Rao, 2003).

Chambers and Tzavidis (2006) have proposed an alternative SAE method based on the structure of a unit-level model. Their M-quantile (MQ) model uses the information about between-area heterogeneity contained in the empirical sample level conditional distribution of the dependent variable, and therefore does not make any parametric assumptions. Furthermore, the parameters of the model are estimated using robust iteratively re-weighted least squares and so are insensitive to the presence of unit-level outliers in the sample data.

There are many extensions to the basic FH area-level and BFH unit-level models in the literature. Cressie (1993) describes a spatial extension that allows the area effects to be correlated spatially (as a function of the distance), and so borrows strength "over space" from the neighboring or similar areas given a simultaneous autoregressive (SAR) error process assumption for these effects (e.g. Petrucci and Salvati, 2006 & 2008; Coelho and Pereira, 2011). If time series data are available (e.g. from previous surveys) for the small areas then models can be defined to take account of inter-temporal correlations and hence borrow strength "over time" (Rao and Yu, 1994). Marhuenda *et al.* (2013) propose a spatio-temporal FH model that is an extension of the Rao and Yu (1994) temporal model. Sinha and Rao (2009) address the sample outliers issue from the perspective of both area- and unit-level linear mixed models. A number of researchers have also applied state-space models in small area problems (e.g. Singh *et al.*, 1994; Feder *et al.*, 2000). Extensions to MQ-based SAE that allow for spatially correlated area effects (e.g. Salvati *et al.*, 2008; Salvati *et al.* 2012) and for the presence of both area-level and unit-level outliers (Chambers *et al.*, 2014) have been proposed. Reviews of recent developments in

SAE methodology are also available from Pfeffermann (2002, 2013), Jiang and Lahiri (2006), and Rao and Molina (2015).

2.2 Small Area Estimation in Poverty Mapping

SAE methodology has essentially been developed to estimate linear functions of the population values of a response variable (e.g. its total and mean). The basic small area estimators based on EBLUP, EB or HB methods are all well suited for this purpose. However in many applications, the parameters of interest are complex non-linear functions of the response variable. In particular, poverty indicators are non-linear functions of the values of a unit-level (say household) welfare variable. In such cases, these basic SAE methods require modification.

The basic poverty indicators derived from household per capita income and expenditure are poverty incidence, poverty gap, and poverty severity. These were first defined by Foster, Greer and Thorbecke (1984), and are referred to as FGT measures from now on. Poverty incidence (also referred to as the head count ratio, HCR) is the proportion of individuals with income below a specified poverty line; poverty gap (PG) is the average relative distance from the poverty line for those individuals with incomes below the poverty line, and poverty severity (PS) is the average of the squares of these distances. There are also fuzzy monetary (and non-monetary) poverty measures (Cheli and Lemmi, 1995; Betti *et al.*, 2006) described in the literature, particularly those based on multidimensional poverty measures (such as lack of health, nutrition, education, energy, access to land, decision making power), which are sometimes used instead of monetary and unidimensional poverty measures (Henninger, 1998). These non-monetary poverty indicators are based on ranking individuals with respect to their welfare variable and do not require any poverty line. Inequality measures are also used to analyse the overall distribution of income, consumption or other attributes related to the welfare variable. The most common of these are the Gini index (Gini, 1912), the Sen index (Sen, 1976), and the Sen-Shorrocks-Thon index (Shorrocks, 1995). From this perspective, the PS can be considered as a combined measure of poverty and inequality.

Small area estimation of FGT poverty measures is the focus of this thesis. From a statistical perspective, estimation of FGT poverty indicators depends on prediction of the area-specific empirical distribution function (i.e. the proportion of people with income less than any particular pre-specified value). The better the prediction of this area-specific distribution function, the better the estimation of associated poverty indicators. In the last decade, a number of SAE methods have been developed to estimate complex poverty indicators, particularly the FGT indicators. These are often referred to as "poverty mapping methods". Below we review the SAE approach to poverty mapping and discuss its application.

2.2.1 Poverty Mapping: Motivations and Applications

Poverty analysis based on national and aggregated level indicators often masks the actual poverty conditions within disaggregated administrative units. Aggregate level poverty indicators are useful for evaluating and monitoring the overall human well-being and poverty situation of a country but are impractical for local policy making and interventions. Geographically disaggregated level indicators provide information about the spatial distribution of inequality and poverty within a country or a large state of a country. These spatially disaggregated indicators are typically published via a "poverty map", which allows the visualization of the incidence and magnitude of poverty across the region (country, state) of interest.

A poverty map is a tool to capture the heterogeneity in poverty and inequality within a country. The disparities in living standards at disaggregated level also suggest the variation in geographic factors that influence the poverty. Multiple dimensions of poverty can also be displayed in a single map (such as HCR with the positions of schools, hospitals, police stations, etc.), which therefore provides a spatial distribution of the determinants of poverty as well. Demombynes and Özler (2005) have calculated poverty estimates for the police jurisdictions in South Africa and showed how incidence of crime is related to levels of local inequality. Policy makers can utilize the spatial distributions of poverty indicators and their determinants in the selection and design of area-specific interventions. This information can also be used to determine how much an area is lagging behind and how many additional resources (such as subsidy for local infrastructure, employment) are required for reducing poverty and inequality across the country. A poverty map also helps to allocate aid during periods of national emergency, as after a flood or an earthquake.

A poverty map for finer geographic regions rather than for larger regions is suitable for effective interventions for several reasons including (i) reducing the leakage of aid to non-poor people (Type I error) and increasing the coverage of poor people (smaller Type II error); (ii) optimizing the design and administrative cost (Bigman and Fofack, 2000); (iii) reflecting regional variations in health care, food safety, education and employment facilities within the intervention programs; and (iv) increasing the effectiveness of targeting poor people who are in small homogeneous domains compared to heterogeneous large region (Datta and Ravallion, 1993).

A poverty map is a powerful way of presenting information on poverty and inequality within a country in a way that is understandable to non-professional viewers and also one that encourages stakeholders to participate in local decision making and in negotiations with government organizations. A poverty map improves communication between national and local decision makers, and as a consequence leads to decentralization and improved effectiveness of national poverty alleviation intervention programs.

On a global scale, there are several large projects that have aimed to generate poverty maps at finer administrative units based on SAE methods. The first (still the largest) of these projects is the Small Area Income & Poverty Estimates (SAIPE) project conducted by the US Census Bureau (Citro *et al.*, 1997). This is an ongoing program in which the US Census Bureau estimates the number of poor school-age children within states and counties each year. These estimates are then used for the administration of federal programs and the allocation of federal funds to local jurisdictions (Maiti and Slud, 2002).

The World Bank initiated development of an expenditure-based SAE methodology in 1996 (Hentschel and Lanjouw, 1996) and finalized its application to poverty mapping in 2002 (Elbers *et al.*, 2002, 2003). This methodology is known as the "World Bank" methodology. It has been applied more than 60 countries (Elbers and Van der Weide, 2014), and particularly in the developing countries of Asia, Africa, and South America. The World Bank method is now accepted as defining the industry standard SAE methodology for creating poverty maps that focus on finer level administrative units (Haslett and Jones, 2010). Some other well-known projects are SAMPLE (Small Area Methods for Poverty and Living Conditions Indicators) and AMELI (Advanced Methodology for the European Laeken Indicators) which were supported by European Commission and which were mainly aimed at investigating and developing suitable SAE methods for small area poverty estimation.

Poverty mapping is now an important tool for decision makers when investigating and discussing social, economic, and environmental problems. Decision makers frequently use poverty maps to identify areas where development has lagged and where investments in infrastructure and services could have the greatest impact. Henninger and Snel (2002) detail the application of poverty maps using case studies of 14 countries from Africa, Asia, and Latin America. Some important application of poverty maps are: poverty reduction strategy in Nicaragua, improving geographic targeting of annual expenditure in Guatemala, designing emergency response and food aid programs (e.g., creating a geo-referenced strategy for comprising cholera outbreak in KwaZulu Natal province of South Africa in 2001 and identifying the poorest communities in Cambodia for distributing World Food Program food aid under "food for work" interventions), local decision-making in "Minas Gerais" state of Brazil (e.g., distributing statewide tax revenues toward poorer municipalities), and improving transparency in public decision-making by raising awareness of poverty in Panama.

2.2.2 Poverty Estimation based on Area-Level Model

Area-level models were used to define the first efficient small area poverty estimation methods. In particular, these models underpin the SAIPE project, where an area-level model-based FH estimator is used to estimate the number of poor school-age children by state, county, and school district (Maiti and Slud, 2002). However, these estimators can be criticized. The effective sample size in this case (and hence an assessment of the reliability of the estimated model parameters) is the number of small areas (D), which is usually much smaller than the number of observations (n) in the survey data. There is also the problem that the method assumes that sampling error variances are known, whereas they are in fact unknown. Furthermore, estimates of this error variance are unstable (due to small n_i in survey data). However, on the positive side, the FH estimator is free from the confidentiality issues that arise with unit-level data and is also less sensitive to unit-level outliers.

FH-type estimators have been used to estimate the proportion of population below a poverty line (i.e. the HCR; see Quintano *et al.*, 2007; Hansen, *et al.* 2011). In order to estimate other FGT poverty indices, separate FH models linking the corresponding direct estimates and potential area-specific explanatory information are required (for example, see Marhuenda *et al.*, 2013, and Guadarrama *et al.*, 2015). When estimating PG and PS, application of the FH model requires one to examine the underlying assumption of a linear relationship between the direct estimates and the area-level explanatory variables. Due to the non-linear behaviour of these poverty indicators, finding an appropriate transformation for fitting a linear area-level model is critical. A further issue is that FGT indicators and other poverty inequality indicators vary with the poverty line. This means that indicator-specific area-level modelling is required for a specific poverty line. In particular, it is usually not possible to fit a single area-level random effect model that can be used with all three FGT measures (HCR, PG, & PS). A multivariate FH model (Benavent and Morales, 2016) has been developed to simultaneously estimate all three FGT indices but this requires area-specific direct estimates with corresponding sampling error variances from an underlying unit-level survey data set.

2.2.3 Poverty Estimation based on Unit-Level Model

The first unit-level SAE method applied to poverty mapping was the World Bank methodology. In this method, a nested-error regression model with random effects at cluster-level (primary sampling unit) is fitted to a survey data set using the logarithm of household per capita income or expenditure as the response variable, and with covariates defined by explanatory variables that are available in both survey and census data sets. The fitted model parameters are used to impute the response values for all census units using the available census information, and the area-specific FGT parameters are then calculated by appropriately aggregating these imputed values. A parametric or non-parametric bootstrap procedure can be used to obtain the standard errors of these FGT estimates. The ELL method assumes that the only significant source of higher level variation is between-cluster variation, which is equivalent to the assumption that it is possible to incorporate a large enough number of explanatory variables in the two-level regression model to ensure negligible between-area variation. As a consequence, the ELL method will fail to provide efficient estimates when there is non-negligible between-area variation in the distribution of the response variable. Though this method is robust to departures from the assumption of normal errors and is not computationally intensive, it is not robust to the presence of outliers and is not robust to model misspecification. However, it does have the advantage that since the model is fitted at household (unit) level and then simulated, the same predicted model is used to estimate poverty indicators at any level of the population (Guadarrama *et al.*, 2015).

An alternative to the ELL method was introduced by Molina and Rao (2010). This is based on the EB prediction (EBP) method and assumes normally distributed random effects at area-level rather than at cluster-level in the nested-error regression model. The predicted non-sample values $\mathbf{y}_r$ (r stands for non-sampled units) of the response variable are derived from the conditional distribution of these unobserved values in $\mathbf{y}_r$ given the (observed) sampled values in $\mathbf{y}_s$ (s stands for sampled units). Thus, like ELL approach, response values are imputed for all census units and then the area-specific complex functions of $\mathbf{y}$ are calculated. The EBP method is sensitive to departures from its normality assumption of random errors, and also requires correct transformation of the welfare variable to normality. The normality assumption has recently been relaxed, with Diallo and Rao (2014) considering use of the skewed normal distribution for the random errors. Elbers and Van der Weide (2014) have also proposed an EBP method assuming the distribution of random errors can be modelled as a normal mixture.

Molina, Nandram and Rao (2014) have proposed a Hierarchical Bayes (HB) method as an alternative to the EB method of poverty estimation. Since the HB method depends on the posterior distribution of the target parameters, it does not require use of the bootstrap to estimate its MSE. Instead, these MSE estimates can be calculated from the posterior variances. Note that other summary measures of the posterior distribution such as credible intervals (the Bayesian confidence interval used for interval estimation) can be easily calculated under HB method. However, the HB method is highly dependent on its model assumptions and is not extendable to more complex models without the use of bootstrap procedures. Advantages and disadvantages of the ELL, EBP and HB estimation methods have been discussed in Guadarrama *et al.* (2015).

Tzavidis *et al.* (2008) proposed an M-quantile (MQ) approach to poverty mapping as an alternative to the ELL and EBP methods. The MQ method is less sensitive to outliers and does not require geographical specification to develop the MQ regression model. The method relies on an area-specific MQ coefficient instead of the area-specific random effect associated with a multilevel model. This is calculated by averaging the MQ coefficients of units belonging to the area of interest, where the MQ coefficient of a unit is the value of q where the observed response and the fitted M-quantile model coincide. As with the EBP method, the application of the MQ method described in Tzavidis *et al.* (2008) assumes between-area variability instead of between-cluster variability.

In all three methods (ELL, EBP and MQ), values of the response values are imputed for all census units using a bootstrap procedure, which can be parametric or non-parametric. MSE estimates are calculated from a single set of bootstrap realisations in the ELL method, but from separate bootstrap realisations in both the EBP and the MQ methods. Furthermore, separate estimation methods are required in EBP and MQ methods for those small areas not represented in the sample survey data. This is an important advantage for the ELL method (at least in developing countries), since in many cases target small areas are not covered in the nationally representative survey used to fit the model for *Y*. In more developed countries, more spread out samples tend to be the norm, and in these cases the EBP or MQ methods can perform better (Elbers and Van der Weide, 2014). The ELL, EBP and MQ methods are discussed in more detail in **Chapter Three**.

In unit-level small area poverty estimation methods, the main aim is to predict the empirical distribution function of the skewed response variable *Y*. In this context, any method of predicting this distribution function can be utilized in order to estimate the FGT poverty indicators (Chambers and Pratesi, 2013). One such method is the smearing-based prediction method of Chambers and Dunstan (1986, hereafter CD). Note that the MQ SAE method uses this smearing approach. In general, however, any other prediction model can be substituted.

Linear mixed models are usually developed under the assumption of homoskedastic random errors. However, this homoskedasticity assumption may be violated because of the presence of outliers in the distribution of response variable, model misspecification, asymmetry of the response variable and use of incorrect transformation (Verbeek, 2008). Heteroskedasticity may also be present in aggregated data where the variance of response variable varies with the size of groups (Gordon, 2012). The violation of the homoskedasticity assumption need not produce bias, but may lead to invalid statistical inferences. Jiang and Nguyen (2012) show that homoskedasticity-based ML estimators of fixed effects and within-cluster correlation remain consistent under heteroskedasticity, but the EBLUP estimator of the small area mean is then non-optimal. A number of heteroskedasticity-consistent estimators are available for the linear model (Hayes and Cai, 2007); however to the best of our knowledge, the literature on heteroskedasticity is still limited to the class of linear mixed models. Bouhlel (2013) contains a comprehensive discussion on heteroskedasticity under the linear mixed model.

When applying the ELL method, level-one (household) errors are often assumed to be heteroskedastic, with the cluster-specific random errors assumed to be homoskedastic. Under this approach, the homoskedastic level-two variance component is first estimated and then the level-one heteroskedastic error variances are estimated using a logistic-type link function of the squared residuals (ELL, 2002). This model can depend on explanatory variables or their transformations and is known as the "alpha model". Fitting such an "alpha model" is a critical part of the exploration of potential explanatory variables and their transformations. However, the explanatory power of the "alpha" model is typically very low, with r-square values typically below 5% in many studies, including those carried out in Bangladesh, Nepal, and Cambodia (Haslett, 2013). This parametric approach to heteroskeadasticity modeling in the ELL method and some alternative non-parametric approaches are discussed in **Chapter Five**.

In the EBP method the nested errors are usually assumed to be homoskedastic. However, Van der Weide (2014) has proposed an EBP method for a linear mixed model which allows heteroskedasticity at level-one under an assumption of normality by adapting the results of Huang and Hidiroglou (2003) on GLS estimation and EB prediction with sampling weights. The assumption of normal errors in the EBP method can be relaxed by following the Elbers and Van der Weide (2014) approach, where these errors are assumed to follow a normal mixture. This is now incorporated into the traditional ELL method in the World Bank POVMAP (version 2.5) software.

2.3 Mean Squared Error (MSE) Estimation for Small Area Poverty Estimates

The accuracy of small area estimates is usually measured by their mean squared error (MSE). In this context, Kackar and Harville (1984) proposed an MSE estimator for EBLUP-type estimators based on Taylor approximation under normality assumption of random errors. Prasad and Rao (1990) proposed an MSE estimator for the EBLUP based on linear mixed models with block-diagonal covariance structure, which used moment estimators of the variance components. This approach has been extended by Datta and Lahiri (2000) to the situation where ML/REML variance component estimators are used.

An alternative to the use of asymptotic analytic approximations for the MSE of the EBLUP estimates is to adopt a resampling method. Availability of high-speed computers and powerful software has increased the use of resampling methods (e.g. jackknife and bootstrap techniques) both for estimating complex parameters and for estimating their MSEs. These procedures are attractive due to their conceptual simplicity, lack of parametric assumptions, and applicability to complex statistical models (Molina *et al.*, 2009). Jiang *et al.* (2002) proposed a unified jackknife method for estimating MSE of the EB predictor for the class of generalized linear mixed effect models. González-Manteiga *et al.* (2008a) proposed a bootstrap MSE estimator of the EBLUP that can approximate the Prasad-Rao MSE estimator under a linear mixed model. Under a generalized linear mixed model, González-Manteiga *et al.* (2007) also provide a bootstrap estimator based on a wild bootstrap design. Hall and Maiti proposed MSE estimation methods based on parametric (2006a) and non-parametric (2006b) bootstrap procedures under a linear mixed model as an alternative to analytic MSE estimators.

For poverty estimation based on the FH area-level model, poverty indices with their accuracy are typically estimated via EBLUP-type and the corresponding analytic MSE estimators respectively (Slud and Maiti, 2006). In this context, the analytic MSE estimation procedures of Kackar and Harville (1984), Prasad and Rao (1990), Datta and Lahiri (2000), and the bootstrap MSE estimation procedures of Pfeffermann and Glickman (2004), González-Manteiga *et al.* (2005), Hall and Maiti (2006a), Opsomer *et al.* (2008) can be employed to estimate the MSEs of the poverty estimates calculated under the area-level model.

Under unit-level model, analytic MSE estimation for estimated FGT poverty measures is often intractable, and so bootstrap-based MSE estimators are typically used. In the ELL method, both FGT estimates and their corresponding MSE estimates are calculated based on the same parametric or non-parametric (semi-parametric for heteroskedastic level-one errors) bootstrap procedure. Similarly, a parametric bootstrap procedure for the EBP method based on González-Manteiga *et al.* (2008a) and a non-parametric bootstrap procedure for MQ method based on Marchetti *et al.* (2012) can be used to estimate the MSEs of the resulting poverty estimates. However, it should be noted that both the EBP and the MQ bootstrap MSE estimators are complex and computationally intensive compared to the ELL bootstrap MSE estimator. MSE estimation procedures for the ELL, EBP and MQ methods are discussed in detail in **Chapter Three**.

Finally, we note that although the ELL method has been widely used (particularly in developing countries), it also been criticised due to its assumption of area-homogeneity (Tarozzi and Deaton, 2009). If the potential between-area variability cannot be captured by the fixed effects in the fitted regression model (i.e. there is model misspecification), then the estimated MSE produced by the ELL method always understates the true MSE. Alternative versions of the ELL method that address this issue have been proposed by Elbers *et al.* (2002) and Elbers *et al.* (2008). However, these have been criticised by Tarozzi and Deaton (2009) due to the posssibility of unstable and inconsistent results. We discuss the MSE estimation problem in more detail in **Chapter Four**.

CHAPTER THREE

3. Alternative Methods for Poverty Estimation in Developing Countries

Small area estimation (SAE) methods are now widely used as indirect statistical methods for improving geographic distribution of poverty indicators. Since the turn of the century, three unit-level SAE techniques: the ELL method of Elbers, Lanjouw, and Lanjouw (2003) also known as World Bank method, the Empirical Best Prediction (EBP) method of Molina and Rao (2010) and the M-Quantile (MQ) method of Tzavidis *et al.* (2008) have all been used to estimate micro-level FGT poverty indicators (Foster, Greer, and Thorbecke, 1984). These methods vary in terms of their underlying model assumptions and specifically their differences in how they use random effects to model the distribution of income across the geography of interest. They consequently perform best when the real data actually agree with their underlying assumptions. In this Chapter we compare the performances of the three methods in terms of the poverty estimates and associated standard errors they produce in an empirical situation where both between-area heterogeneity and between-cluster heterogeneity are present in the superpopulation model. Our comparison study is based on a simulated dataset which explicitly replicates a data scenario for a developing country. In particular, we compare the ELL approach based on a three-level nested-error regression model with the standard ELL approach based on a two-level model in this situation.

The Chapter is organized as follows: **Section 3.1** briefly discusses the ELL, EBP, and MQ methods of SAE and their applications to poverty estimation; **Section 3.2** describes the FGT poverty indicators and details how these three unit-level SAE methods are used in poverty studies; **Section 3.3** then compares these SAE methods in terms of some key issues as far as poverty estimation is concerned; **Section 3.4** outlines the construction of the census dataset and the sampling design to select a survey sample from it (with the associated dataset). This construction mimics real datasets from Bangladesh. An empirical-based simulation study based on the constructed pseudo census data is also illustrated in Section 3.4. **Section 3.5** discusses the findings of the simulation study; and finally **Section 3.6** concludes with a discussion of some outstanding statistical issues regarding the selection of a SAE method for poverty mapping in developing countries.

3.1 Background

Recently national and local administrative authorities place more importance on small area statistics for policy development and implementation. The traditional national surveys are conducted to obtain national and regional level statistics and so statistics at micro-level administrative units are not estimable. Direct estimation method is not possible to obtain estimate with efficient precision due to very small sample size (even zero) for a significant number of micro-level administrative units. Model-based small area estimation (SAE) method provides the opportunity to calculate the small area parameters with greater efficiency by the so called "borrowing strength" across all the small areas (Rao, 2003). The basic idea of SAE method is to develop a regression model to a survey data and then utilize the known values of the auxiliary variables available in census and/or administrative database for each small area.

Model-based SAE methods are mainly classified into two broad groups based on data availability. Unit-level SAE method is used when data is available at unit-level; otherwise area-level SAE method is applied. In case of unit-level SAE method, the survey data include main variable of interest and some related explanatory variables which must be in the form of census/administrative database. A linear mixed model is then developed using the survey data and the developed model is utilized to obtain the small area statistics using the auxiliary information available for all population units in census dataset. While in area-level SAE method, the weighted direct estimates of the small area parameters obtained from a survey data and area-level explanatory variables obtained from census data are utilized to develop a linear mixed model. This linear mixed model is then used to calculate small area statistics for both sampled and non-sampled areas in the survey.

The basic model-based SAE methods are concerned with estimation of small area linear parameters such as means and totals. The straightforward application of standard SAE method is not appropriate for the parameters which are non-linear function of the response variable such as quantiles and distribution functions. In poverty mapping study, the mostly used poverty indicators: poverty incidence, poverty gap and poverty severity are non-linear function of small area income or expenditure distribution function (Foster et al., 1984). Consequently, the modified SAE methods are required for estimating these non-linear poverty indicators at disaggregated administrative units.

The unit-level SAE method for poverty mapping was first developed by Elbers *et al.* (2003) which is known as World Bank method and in short ELL method. In the ELL method a 2-level nested-error regression model (Battese *et al.*, 1988) is developed using a survey data at household (HH) level and the developed model is utilized to impute HH-level response values for all census HHs, which are then aggregated over the corresponding small areas. The ultimate poverty indicators and their mean squared error (MSE) are calculated via a bootstrap procedure where the fitted model and the estimated regression parameters are used as inputs. Due to simplicity of the ELL methodology, it has been implemented in most of the developing countries including Thailand (Healy *et al.*, 2003), Cambodia (Fujii, 2004), South Africa (Alderman *et al.*, 2002), Brazil (Elbers *et al.*, 2004), Bangladesh (BBS and UNWFP, 2004), the Philippines (Haslett and Jones, 2005), Nepal (Haslett and Jones, 2006) and so on. This method has also been applied for estimating prevalence of child undernutrition at disaggregated administrative units (Haslett and Jones, 2006; BBS and UNWFP, 2004).

Molina and Rao (2010) utilized the method of empirical Bayes or best prediction (EBP) under a finite population to estimate the FGT measures. The EBP method produces estimators with minimum mean squared error or “best predictor” through Monte Carlo (MC) approximation under the assumption that transformed welfare variable follows a nested-error regression model with normally distributed random errors. In EBP method simulated values of the welfare variable are generated via independent draws from the conditional distribution of unobserved values given the observed values. The EBP method is based on a 2-level nested-error regression model with area-specific random effects (area variability), while the ELL method is also based on a 2-level model but with cluster-specific random effects (cluster variability).

Both the ELL and EBP methods are based on standard random effects model with strong distributional assumptions, formal specification of the random part, and are also non-robust to the presence of outliers in the values of response variable. An alternative approach to SAE is MQ method proposed by Chambers and Tzavidis (2006), and Tzavidis and Chambers (2007) based on the quantiles of the conditional distribution of response variable given the covariates (Breckling and Chambers, 1988). Unlikely ELL and EBP, MQ method is distribution free and automatically provides outlier robust inference. The MQ model-based approach has been brought in poverty mapping study by Tzavidis *et al.* (2008, 2010). Application of MQ methodology in poverty estimation relaxes the parametric assumptions of random errors in the traditional unit-level nested-error regression model.

The EBP method has been implemented mostly in European Union Member States (EU-MS) using the Income and Living Conditions (EU-SILC 2012) survey datasets. Molina and Rao (2010) applied EBPP to estimate province-level poverty incidence and poverty gap by sex. The MQ method has also been applied to estimate poverty indicators at local administrative in different European countries including Albania (Tzavidis *et al.*, 2008), Tuscany (Giusti *et al.*, 2009), Italy (Giusti *et al.*, 2011). Recently both methods have been applied to estimate multidimensional poverty at municipality-level in Mexico (Licona and Jimenez, 2015). The underlying assumptions of EBP and MQ methods indicate their suitability in more developed countries where more spread out samples are collected covering most of the target small areas (Elbers and Van der Weide, 2014).

Since the three SAE methods for poverty mapping are based on unique assumptions, all methods are not perfect for a specific condition. The ELL method has criticism of assuming area-homogeneity, while the EBP method assumes cluster-homogeneity. Like EBP, the MQ method does not care about cluster-specific variation since it ignores boundary issue in developing the MQ model. The ELL and EBP methods are likely to suffer from model misspecification particularly the parametric specification of random errors and the presence of outliers. In reality it is very tough to select an appropriate SAE method for poverty studies due to unknown behaviour of population at the target domains where sample information is poor (even none). The aim of this chapter is to compare and contrast the three unit-level SAE methods numerically via a simulation study built on the datasets of Bangladesh as a candidate of developing countries. Based on the simulation study, some benchmarks have also been made for the selection of SAE method for poverty estimation in the developing countries.

3.2 Small Area Methodologies for Poverty Estimation

Consider a finite population of size N divided into D small domains of size $N_1, N_2, ..., N_D$ and let E_{ik} be the income or expenditure for individual k belongs to i^{th} domain. If E_{ik} is less than a poverty line t, the k^{th} individual will be considered as below the 'poverty line'. The FGT poverty measures for domain i are then calculated as

$$F_{\alpha i} = N_i^{-1} \sum_{k=1}^{N_i} F_{\alpha ik};\ i = 1, 2, ..., D;\ \alpha = 0, 1, 2 \tag{3.1}$$

where $F_{\alpha ik} = \left(\frac{t - E_{ik}}{t}\right)^{\alpha} I\left(E_{ik} < t\right)$. The FGT poverty indicators are then defined as

$$F_{\alpha i} = \begin{cases} \text{Poverty Incidence} & if \quad \alpha = 0 \\ \text{Poverty Gap} & if \quad \alpha = 1 \\ \text{Poverty Severity} & if \quad \alpha = 2 \end{cases}$$

which clearly illustrate that poverty incidence (also known as Head Count Rate, HCR) is the proportion of population having income or expenditure level under the poverty line, poverty gap (PG) is the expected income shortfall from the poverty line, and poverty severity (PS) is the expected squared shortfall of income from the poverty line.

Now consider a random sample of size $n = n_1 + n_2 + ... + n_D$ is drawn from the population following a specified sampling design. The sizes of sampled and non-sampled parts of an area i are defined as n_i and $\left(N_i - n_i\right)$ respectively. Also s_i and r_i denote the sets of sample and out-of-sample individuals belonging to area i. In such case the unweighted and weighted direct poverty indicators are defined as $\hat{F}_{\alpha i} = n_i^{-1}\sum_{k=1}^{n_i} F_{\alpha ik}$ and $\hat{F}_{\alpha i}^{w} = \hat{N}_i^{-1}\sum_{k=1}^{n_i} w_{ik}F_{\alpha ik}$ respectively where $w_{ik} = N_i/n_i$ is the sampling weight for individual k in area i and $\hat{N}_i = \sum_{k=1}^{n_i} w_{ik}$ is a design-unbiased estimator of N_i. In our study, the census units are HHs and the variable of interest is per capita consumption expenditure, and so the accumulation needs to be weighted by HH size when we utilize (3.1). Then the ultimate FGT indicators will be $F_{\alpha i} = M_i^{-1}\sum_{k=1}^{N_i} m_{ik}F_{\alpha ik}$ where m_{ik} is HH size of k^{th} HH and $M_i = \sum_{k=1}^{N_i} m_{ik}$ is total population in i^{th} area. Though it is assumed that all the small areas have sampled units in theoretical development, in reality a significant number of small areas are not covered in sample (where $n_i = 0$). In such case, synthetic estimators are used to calculate poverty indices through model-based SAE methods based on the auxiliary information of those areas (Rao, 2003). The ELL, EBP and MQ methods of poverty estimation are elaborately discussed in the following subsequent sub-sections.

3.2.1 The World Bank (ELL) Method

In the World Bank poverty mapping methodology, a regression model is developed between the log-transformed response variable $Y_{ijk} = \log\left(E_{ijk}\right)$ and the explanatory variables $\left(X_{ijk}\right)$ available in a sample data of recent household income and expenditure survey. Here E_{ijk} is per capita expenditure for k^{th} HH lives in j^{th} cluster of i^{th} area. Standard methods of fitting regression model cannot be used due to implementation of complex sampling technique (say, stratification and/or cluster) for selecting the clusters and the HHs in the survey design. Moreover, since HHs are tend to be clustered together into villages or other small geographic or administrative units that are themselves relatively homogenous, it is common to have a cluster (primary sampling unit) effect on the response values. So a nested-error linear regression model (Battese *et al.*, 1988) is developed considering HHs at level one and clusters at level two as

$$\begin{gathered} y_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} + u_{ij} + \varepsilon_{ijk} \quad i = 1,\ldots,D \;\; j = 1,2,\ldots,C_i \;\; k = 1,2,\ldots,N_{ij} \\ u_{ij} \sim N\left(0, \sigma^2_{u(2)}\right) \;\; \varepsilon_{ijk} \sim N\left(0, \sigma^2_{\varepsilon(2)}\right) \end{gathered} \tag{3.2}$$

where cluster-specific and HH-specific errors u_{ij} and ε_{ijk} are assumed to follow normal distribution with constant variance components. The sub-script (2) of $\boldsymbol{\beta}_{(2)}$, $\sigma^2_{u(2)}$, and $\sigma^2_{\varepsilon(2)}$ stands for 2-level model. The model can be extended to 3-level hierarchical model if individual level data are available (say children within household).

Since poverty indicators are function of the distribution of y_{ijk}, not the function of $\mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)}$, ELL uses a bootstrap procedure to regenerate the conditional distribution of y_{ijk} by adding simulated values of cluster-specific $\left(u_{ij}^*\right)$ and HH-specific errors $\left(\varepsilon_{ijk}^*\right)$ to each estimated fitted value $\left(\mathbf{x}_{ijk}^T \hat{\boldsymbol{\beta}}_{(2)}\right)$. The ELL method provides unbiased estimates of poverty measures with their standard errors via a parametric bootstrap procedure as below.

Step 1: Fit model (3.2) to the sample data $\left(\mathbf{y}_s, \mathbf{x}_s\right)$ and obtain $\hat{\boldsymbol{\beta}}_{(2)}$, $\hat{\mathbf{v}}\left(\hat{\boldsymbol{\beta}}_{(2)}\right)$, $\hat{\sigma}^2_{u(2)}$, and $\hat{\sigma}^2_{\varepsilon(2)}$ using appropriate estimation method such as maximum likelihood (ML) or restricted maximum likelihood (REML) or method of moments (MOM).

Step 2: Generate the values of the regression parameters from the corresponding parametric distribution as $\boldsymbol{\beta}^*_{(2)} \sim N\left(\hat{\boldsymbol{\beta}}_{(2)}, \hat{\mathbf{v}}\left(\hat{\boldsymbol{\beta}}_{(2)}\right)\right)$.

Step 3: Generate the cluster-specific and HH-specific random errors independently and identically from the corresponding parametric distributions such as $u_{ij}^{*} \sim N\left(0, \hat{\sigma}_{u(2)}^{2}\right)$ and $\varepsilon_{ijk}^{*} \sim N\left(0, \hat{\sigma}_{\varepsilon(2)}^{2}\right)$ respectively for $i = 1, ..., D$, $j = 1, 2, ..., C_i$, and $k = 1, 2, ..., N_{ij}$. The random errors can be generated from the corresponding empirical distributions to relax the parametric assumptions of random errors.

Step 4: Generate L (say, L =500) independent and identically distributed bootstrap population income values $\left\{y_{ijk}^{*(l)};\ l = 1, 2, ..., L\right\}$ via the bootstrap super-population model $y_{ijk}^{*} = \mathbf{x}_{ijk}^{T}\boldsymbol{\beta}_{(2)}^{*} + u_{ij}^{*} + \varepsilon_{ijk}^{*}$.

Step 5: The parameter of interest for a particular small area is calculated by aggregating the generated income values belonging to the small area. The FGT poverty measures $F_{\alpha i}^{*(b)}$ are calculated from each bootstrap population and then the ELL estimates with their mean squared error (mse) are calculated as

$$\hat{F}_{\alpha i}^{ELL} = L^{-1}\sum_{l=1}^{L} F_{\alpha i}^{*(l)} \text{ and } mse\left(\hat{F}_{\alpha i}^{ELL}\right) = L^{-1}\sum_{l=1}^{L}\left\{F_{\alpha i}^{*(l)} - \hat{F}_{\alpha i}^{ELL}\right\}^{2}. \quad (3.3)$$

The HH-specific errors (also known as idiosyncratic errors) are allowed to be heteroskedastic in the ELL method. In this chapter, the HH-level errors are assumed homoskedastic for the comparison purposes. In either case of homoskedastic or heteroskedastic HH level errors, both parametric and semi-parametric bootstrap procedures can be applied in the ELL method.

In ELL methodology, the basic idea is to increase the predictive power of the fitted regression model (high R-squared value) and to reduce as much as possible the ratio of between-cluster variation to total variation $\hat{\sigma}_{u(2)}^{2}\left(\hat{\sigma}_{u(2)}^{2} + \hat{\sigma}_{\varepsilon(2)}^{2}\right)^{-1}$. For these reasons, more explanatory variables at different hierarchical levels such as HH, cluster and area are considered in the regression model. Synthetic behaviour and ignorance of between-area variation (as a consequence produce underestimated MSE) are the main criticisms of the ELL methodology.

An ELL method based on 3-level model (3-level ELL) instead of the standard 2-level model (2-level ELL) can be easily conducted assuming a 3-level structure of the super-population model. The procedure is almost similar as the established 2-level model-based ELL methodology where an additional area-level random error is generated assuming a parametric or empirical distribution function. The procedure based on a 3-level model solves the problem of underestimated MSE estimation if in reality the area-level variability exits. However, it is very tough to obtain unbiased and efficient estimates of all level-specific variance components due to lack of data at the considered levels. The ELL estimators based on 2-level and 3-level models are denoted as ELL.2L and ELL.3L respectively.

3.2.2 Empirical Best Prediction (EBP) Method

In the EBP method, small domain of interest (aggregate of clusters) rather than cluster is considered to have random effects on the response variable. The 2-level nested-error regression model is considered as

$$\begin{gathered} y_{ik} = \mathbf{x}_{ik}^T \boldsymbol{\beta}_{(2)} + \eta_i + \varepsilon_{ik} \quad i = 1,..,D \quad k = 1,..,N_i \\ \eta_i \sim N\left(0, \sigma^2_{\eta(2)}\right) \quad \varepsilon_{ik} \sim N\left(0, \sigma^2_{\varepsilon(2)}\right) \end{gathered} \tag{3.4}$$

where η_i and ε_{ik} are area- and HH-specific random errors. Here the FGT estimator is split into sample and non-sample part as

$$F_{\alpha i} = N_i^{-1} \left[\sum_{k \in s_d} F_{\alpha ik} + \sum_{k \in r_d} F_{\alpha ik} \right] \text{ for } \alpha = 0,1,2$$

and the non-sample part is then predicted based on a MC approximation. The best predictor of $F_{\alpha i}$ under the squared error loss is given by

$$\hat{F}_{\alpha i} = E_{\mathbf{y}_{r_i}}\left[F_{\alpha i} \mid \mathbf{y}_{s_i}\right] = \int F_{\alpha i} f\left(\mathbf{y}_{r_i} \mid \mathbf{y}_{s_i}\right) \mathbf{y}_{r_i}$$

where $f\left(\mathbf{y}_{r_i} \mid \mathbf{y}_{s_i}\right)$ is the joint density of $\mathbf{y}_{r_i}$ (vector of non-sample y in area i) given $\mathbf{y}_{s_i}$ (vector of sample y in area i). The basic procedures of EBP method to obtain the estimate of $F_{\alpha i}$ can be explained briefly as:

Step 1: Fit the nested-error model (3.4) to survey data $\left(\mathbf{y}_s, \mathbf{x}_s\right)$ by ML or REML or Henderson method III (Rao, 2003) to obtain the estimated regression parameters $\hat{\boldsymbol{\beta}}_{(2)}, \hat{\sigma}^2_{\eta(2)}$, and $\hat{\sigma}^2_{\varepsilon(2)}$.

Step 2: Generate L independent realization of $\mathbf{y}_{r_i}$ $\left\{\mathbf{y}^{(l)}_{r_i}, l = 1, 2, .., L\right\}$ from the conditional distribution $f\left(\mathbf{y}_{r_i} \mid \mathbf{y}_{s_i}\right)$ which is assumed to follow normal distribution with mean $\boldsymbol{\mu}_{i_{r|s}} = \mathbf{X}_{r_i}\boldsymbol{\beta} + \sigma^2_{u(2)}\mathbf{1}_{N_i - n_i}\mathbf{1}^T_{n_i}\mathbf{V}^{-1}_{si}\left(\mathbf{y}_{s_i} - \mathbf{X}_{s_i}\boldsymbol{\beta}\right)$ and variance-covariance matrix $\mathbf{V}_{i_{r|s}} = \sigma^2_{u(2)}\left(1 - \gamma_i\right)\mathbf{1}_{N_i - n_i}\mathbf{1}^T_{N_i - n_i} + \sigma^2_{\varepsilon(2)}\mathbf{I}_{N_i - n_i}$ where $\gamma_i^{-1} = \sigma^{-2}_{u(2)}\left(\sigma^2_{u(2)} + n_i^{-1}\sigma^2_{\varepsilon(2)}\right)$, $\mathbf{X}_{s_i}$ and $\mathbf{X}_{r_i}$ are matrices of sample and non-sample values of explanatory variables belong to i^{th} area. The MC approximation method is simplified by generating observations from $y^{(l)}_{ik} = \mathbf{x}^T_{ik}\hat{\boldsymbol{\beta}}_{(2)} + v^{*(l)}_i + \varepsilon^{*(l)}_{ik};\ ik \in r_i$ where $v^*_i \sim N\left(0, \hat{\sigma}^2_{u(2)}\left(1 - \hat{\gamma}_i\right)\right)$ is independent of $\varepsilon^*_{ik} \sim N\left(0, \hat{\sigma}^2_{\varepsilon(2)}\right)$.

Step 3: Calculate $F^{(l)}_{\alpha i}$ using the vector $\mathbf{y}^{(l)}_i = \left\{\mathbf{y}^T_{s_i}, \mathbf{y}^{T(l)}_{r_i}\right\}$ and then average over L replicates to obtain the EBP estimates as $\hat{F}^{EBP}_{\alpha i} \approx L^{-1}\sum_{l=1}^{L} F^{(l)}_{\alpha i}$.

In EBP methodology, the mean square error (MSE) estimator of $\hat{F}_{\alpha d}^{EBP}$ is calculated separately via a parametric bootstrap method (González-Manteiga *et al.*, 2008a). The steps are: (i) obtain $\hat{\boldsymbol{\beta}}_{(2)}$, $\hat{\sigma}_{\eta(2)}^{2}$, and $\hat{\sigma}_{\varepsilon(2)}^{2}$ following a suitable estimation method; (ii) generate domain effects $\eta_i^* \sim N\left(0, \hat{\sigma}_{\eta(2)}^2\right)$ and HH effects $\varepsilon_{ik}^* \sim N\left(0, \hat{\sigma}_{\varepsilon(2)}^2\right)$ independently, and then calculate bootstrap population income values y_{ik}^* via $y_{ik}^* = \mathbf{x}_{ik}^T \hat{\boldsymbol{\beta}}_{(2)} + \eta_i^* + \varepsilon_{ik}^*$ for all census HHs; (iii) calculate domain-specific $F_{\alpha i}^*$ by aggregating the bootstrap population values; (iv) then refit the nested-error model (3.4) with the bootstrap elements y_{ik}^* belonging to the sample and calculate the bootstrap model parameters denoted as $\hat{\boldsymbol{\beta}}_{(2)}^*$, $\hat{\sigma}_{\eta(2)}^{2*}$, $\hat{\sigma}_{\varepsilon(2)}^{2*}$; (v) obtain the bootstrap EBP estimator of $F_{\alpha i}^*$ denoted by $\hat{F}_{\alpha i}^{*\text{EBP}}$ through MC approximation method; (vi) repeat steps (ii)-(v) a large number of times, say, B =1000. Let $F_{\alpha i}^{*(b)}$ be the true value calculated at step (iii) and $\hat{F}_{\alpha i}^{*\text{EBP}(b)}$ be the EBP estimate calculated at step (v) in the b^{th} $\left(b = 1,..,B\right)$ replicate of the bootstrap procedure; (vii) then the estimated MSE of $\hat{F}_{\alpha i}^{EBP}$ is defined by $mse\left(\hat{F}_{\alpha i}^{EBP}\right) = B^{-1} \sum_{b=1}^{B} \left[F_{\alpha i}^{*(b)} - \hat{F}_{\alpha i}^{*\text{EBP}(b)} \right]^2$.

Molina and Rao (2010) pointed out that in the simplest case of estimating a small area mean, the EBP approach leads to $E\left(\bar{y}_i^*\right) = N_i^{-1}\left(\sum_{k \in s_i}^{n_i} y_{ik} + \sum_{k \in r_i}^{N_i - n_i} \mathbf{x}_{ik}^T \boldsymbol{\beta}_{(2)} + \eta_i^*\right)$, while the ELL method provides a synthetic regression estimator as $E\left(\bar{y}_i^*\right) = N_i^{-1} \sum_{ijk \in s_i + r_i}^{N_i} \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)}$ since $E\left(u_{ij}^*\right) = 0$ and $E\left(\varepsilon_{ijk}^*\right) = 0$ for a specific area. And so the ELL estimator would be less efficient than EBP estimator if the between area/cluster variation is higher comapred to the total variation. If the number of small areas is large or number of sampled small areas with fewer observation (for example, single cluster per sampled small area), however EBP method will essentially deduce to the ELL method (Haslett, 2013). In reality a large number of small areas are uncovered in the survey data set and then ultimately the EBP approach will produce the synthetic estimates like ELL.

3.2.3 M-Quantile (MQ) Method

The M-quantile of order q for the conditional distribution of a random variable y given a vector of P covariates $\mathbf{x}$ denoted by $Q_q(\mathbf{x};\psi)$ is defined as the solution of the estimating equation $\int \psi_q\left(y-Q_q(\mathbf{x};\psi)\right) f(y \mid \mathbf{x})\, dy=0$ where ψ denotes an asymmetric influence function. Suppose $(\mathbf{x}_k, y_k)$, $k=1,..,n$ denotes the observed sample values of $(\mathbf{x}, y)$ for an individual k. Then a linear MQ regression model for y_k given $\mathbf{x}_k$ is one where we assume that the q^{th} M-quantile satisfies $Q_q(\mathbf{x}_k;\psi)=\mathbf{x}_k^T \boldsymbol{\beta}_\psi(q)$.

For a specified q and continuous ψ, the MQ regression parameters $\boldsymbol{\beta}_\psi(q)$ are estimated by solving the estimating equations $\sum_{k=1}^{n} \psi_q\left[\varphi^{-1}\left\{y_k-\mathbf{x}_k^t \hat{\boldsymbol{\beta}}_\psi(q)\right\}\right] \mathbf{x}_k=0$ where $\psi_q(\vartheta)=2\psi(\vartheta)\left\{q * I(\vartheta>0)+(1-q) * I(\vartheta \leq 0)\right\}$ and φ is a suitable robust estimate of the scale of the residuals (say, Mean Absolute deviation (MAD) estimate *i.e.*, $\varphi=(0.6745)^{-1} * median \mid y_k-\mathbf{x}_k^t \hat{\boldsymbol{\beta}}_\psi(q) \mid$). Here ψ is an appropriately chosen influence function such as Chambers and Tzavidis's (2006) Huber Proposal 2 influence function $\psi(\vartheta)=\vartheta * I(-c \leq \vartheta \leq c)+c * \operatorname{sgn}(\vartheta)$, and Tzavidis and Chambers's (2007) Huber function $\psi(\vartheta)=\vartheta * I(-c \leq \vartheta \leq c)+c * \operatorname{sgn}(\vartheta) * I(|\vartheta|>c)$ with the tuning constant

$c = 1.345$. A straightforward application of an iterative weighted least squares (IWLS) algorithm provides the solution of the estimating equations.

The basic idea of SAE method based on MQ model is that unlikely in a mixed effects model the conditional variability related to the conditional distribution of y given $\mathbf{x}$ does not depend on the pre-defined hierarchical structure, rather than being characterized by the MQ coefficients of the population units. Also it is assumed that since MQ coefficients are determined at population level, population units within a small area have almost similar MQ coefficients. The MQ coefficient q_k for the population unit k with values y_k and $\mathbf{x}_k$ is obtained such that $Q_{q_k}\left(\mathbf{x}_{ik};\psi\right) = y_{ik}$. When the conditional M-quantiles are assumed to follow a linear model $Q_q\left(\mathbf{x};\psi\right) = \mathbf{x}^T\boldsymbol{\beta}_\psi\left(q\right)$, Tzavidis and Chambers (2007) proposed a bias corrected MQ predictor of area-specific mean $\left(\mu_i\right)$ based on smearing method of Chambers and Dunstan (1986) as

$$\hat{\mu}_i^{MQ.CD} = N_i^{-1}\left(\sum_{k\in s_i} y_{ik} + \sum_{k'\in r_i} \mathbf{x}_{ik'}^t \hat{\boldsymbol{\beta}}_\psi\left(\hat{\theta}_i\right) + \frac{N_i - n_i}{n_i}\sum_{k\in s_i}\left(y_{ik} - \hat{y}_{ik}\right)\right) \tag{3.5}$$

where $\hat{\theta}_i$ is the estimator of MQ coefficient $\left(\theta_i\right)$ for area i calculated by averaging the MQ coefficients of the units in area i, $\hat{\boldsymbol{\beta}}_\psi\left(\hat{\theta}_i\right)$ is the area-specific regression coefficients and $\hat{y}_{ik} = \mathbf{x}_{ik}^T\hat{\boldsymbol{\beta}}_\psi\left(\hat{\theta}_i\right)$ is a linear combination of the auxiliary variables.

Straightforward estimation of FGT poverty indicators (3.1) is also not possible in MQ method. Like EBP, estimation of these non-linear parameters corresponds to the problem of estimating the out-of-sample observations. Thus MQ estimator of FGT poverty indicators and quantiles can be written as

$$\hat{F}_{\alpha i}^{MQ} = N_d^{-1}\left[\sum_{k\in s_i} F_{\alpha ik} + \sum_{k^{/}\in r_i} \hat{F}_{\alpha ik^{/}}\right] \quad i = 1,2,\ldots,D \quad \alpha = 0,1,2 \tag{3.6}$$

where $\hat{F}_{\alpha ik^{/}} = n_i^{-1}\sum_{k\in s_i}\left(\frac{t-\hat{E}_{ikk^{/}}}{t}\right)^{\alpha} I\left(\hat{E}_{ikk^{/}} \le t\right)$ with $\hat{E}_{ikk^{/}} = \exp\left[\mathbf{x}_{ik^{/}}^T \hat{\boldsymbol{\beta}}_{\psi}\left(\hat{\theta}_i\right) + e_{ik}\right]$ and e_{ik} is the sample residuals defined by the fitted MQ model to $\left(y_s, x_s\right)$. Marchetti *et al.* (2012) proposed an alternative procedure to calculate (3.6) following a MC simulation approach parallel to the EBP approach. The basic steps are:

Step 1: Fit the MQ small area models to the survey data $\left(y_s, x_s\right)$ and obtain estimates of MQ parameters θ_i and $\boldsymbol{\beta}_{\psi}\left(\theta_i\right)$ for $i = 1,..,D$.

Step 2: Generate an out of sample vector of size $\left(N_i - n_i\right)$ using the estimated model parameters $\hat{\theta}_i$ and $\hat{\beta}_{\psi}\left(\hat{\theta}_i\right)$ in the MQ model $y_k^* = x_k^T \hat{\beta}_{\psi}\left(\hat{\theta}_i\right) + e_k^*;\, k \in r_i$ where e_k^* is drawn from the empirical distribution function of the model residuals.

Step 3: Repeat step 2 a large number of times L (say, L=1000) to calculate L estimate of $F_{\alpha i}$ $\left(F_{\alpha i}^{*(l)}; l = 1,2,..,L\right)$ combining sample $\left(y_{s_i}\right)$ and non-sample observations $\left(y_{r_i}^*\right)$ in each process. Average the L estimates of $F_{\alpha i}$ to obtain ultimate MQ estimate as $\hat{F}_{\alpha i}^{MQ} = \sum_{l=1}^{L} F_{\alpha i}^{*(l)}$.

Though the MSE of $\hat{\mu}_d^{MQ.CD}$ can be estimated analytically (Chambers *et al.*, 2011), the procedure can be unstable when area-specific sample sizes are very small (Marchetti *et al.*, 2012). A non-parametric bootstrap procedure is proposed by Marchetti *et al.* (2012) to estimate the MSE for not only small area mean but also poverty indicators and quantiles. The bootstrap procedure can be illustrated briefly only for the poverty indicators as: **(i)** fit an MQ small area model $y_{ik} = \mathbf{x}_{ik}^{T}\beta_{\psi}(\theta_i) + e_{ik}$ to sample data where e_{ik} is the unit-level random term with distribution function G for which no explicit parametric assumptions are considered; **(ii)** estimate the target small area parameters denoted by $\hat{F}_{\alpha i} = N_i^{-1}\left[\sum_{k\in s_i} Z(y_{ik},t,\alpha) I(y_{ik} \le t) + n_i^{-1}\sum_{k^{/}\in r_i}\sum_{k\in s_i} Z(\hat{y}_{ikk^{/}},t,\alpha) I(\hat{y}_{ikk^{/}} \le t)\right]$, where $\hat{y}_{ikk^{/}} = \mathbf{x}_{ik^{/}}^{T}\boldsymbol{\beta}_{\psi}(\hat{\theta}_i) + e_{ik}$, $e_{ik} = y_{ik} - \mathbf{x}_{ik}^{t}\boldsymbol{\beta}_{\psi}(\hat{\theta}_i)$, and $Z(y_{ik},t,\alpha) = \left(\frac{t-y_{ik}}{t}\right)^{\alpha}$; **(iii)** define an estimator denoted by $\hat{G}(u)$ of the distribution of residuals from either the empirical distribution function of the area-specific sample residuals (e_{ik}) or a corresponding smoothed distribution function; **(iv)** generate a bootstrap population consistent with MQ small area model by sampling e_{ik}^{*} from the $\hat{G}(u)$ either without consideration of the small area (unconditional) or with consideration of the small area (conditional) available in sample data via $y_{ik}^{*} = \mathbf{x}_{ik}^{t}\hat{\boldsymbol{\beta}}_{\psi}(\hat{\theta}_i) + e_{ik}^{*}$, $k = 1,2,...,N_i$, $i = 1,2,...,D$; **(v)** calculate the bootstrap population parameter as $F_{\alpha i}^{*} = N_i^{-1}\left[\sum_{k\in s_d^{*}} Z(y_{ik}^{*},t,\alpha) I(y_{ik}^{*} \le t) + \sum_{k^{/}\in r_d^{*}} Z(y_{ik^{/}}^{*},t,\alpha) I(y_{ik^{/}}^{*} \le t)\right]$; **(vi)** select a without replacement sample from the bootstrap population and estimate the bootstrap population parameter $\hat{F}_{\alpha i}^{*}$ using the estimator $\hat{F}_{\alpha i}$; **(vii)** generate B bootstrap populations following the step (iv) and draw R bootstrap sample from each bootstrap population using simple random sampling (without replacement) within the small areas so that $n_d^{*} = n_d$ and record the values $F_{\alpha i}^{*(b)}$ and $\hat{F}_{\alpha i}^{*(br)}$ with $b = 1,2,..,B$ and $r = 1,2,..,R$; **(viii)** estimate bias and variance of the estimated parameter $\hat{F}_{\alpha i}^{MQ}$ as $bias\left(\hat{F}_{\alpha i}^{MQ}\right) = B^{-1}R^{-1}\sum_{b,r}\left(\hat{F}_{\alpha i}^{*(br)} - F_{\alpha i}^{*(b)}\right)$ and

$\operatorname{var}\left(\hat{F}_{\alpha i}^{MQ}\right)=B^{-1}R^{-1}\sum_{b,r}\left(\hat{F}_{\alpha i}^{*(br)}-\bar{\hat{F}}_{\alpha i}^{*(br)}\right)^{2}$ where $\bar{\hat{F}}_{\alpha i}^{*(br)}=R^{-1}\sum_{r=1}^{R}\hat{F}_{\alpha i}^{*(br)}$; and **(ix)** then calculate the bootstrap MSE estimator of $\hat{F}_{\alpha i}^{MQ}$ as $mse\left(\hat{F}_{\alpha i}^{MQ}\right)=\operatorname{var}\left(\hat{F}_{\alpha i}^{MQ}\right)+bias\left(\hat{F}_{\alpha i}^{MQ}\right)^{2}$.

The description of the EBP and MQ method clearly indicates that both the methods require massive time in the estimation of RMSE compared to ELL method. To reduce the computational demand, the computation can be done separately by dividing both sample and population into several large regions. Ferretti and Molina (2012) suggest a fast algorithm for conducting the EBP method quickly.

3.3 Comparison of ELL, EBP and MQ Methods

The three unit-level SAE methods of poverty estimation are based on different assumptions and so will work better in different situations. There is no common analytic way to show which method is better. All the methods have some drawbacks particularly departure from the assumed distribution of random errors, misspecification of the multilevel model, presence of outliers in the distribution of target variable, and computational problem. A comparison of these SAE methods has been made below with respect to some technical issues noted in Table 3.1.

The basic difference of the three methods is the distribution of random errors: ELL assumes cluster-specific and EBP assumes area-specific random errors which follow some parametric distributions. The MQ method is free from distributional assumption of random effects; however the area-specific MQ coefficient behaves similar to area-specific random effects. The distributional assumption can also be relaxed in ELL method by using the empirical distribution of random errors in the bootstrap procedure.

All the three methods ignore either cluster- or area-level variability under a true 3-level superpopulation model. The ELL method assumes only between-cluster variation and so need to incorporate a large number of explanatory variables in the 2-level regression model to reduce the between-area variation. The ELL method may fail to provide estimates with efficient MSE if a negligible amount of between-area variation remains in the distribution of the response variable after inclusion of some contextual variables in the model specification. Similar problem is also expected in both EBP and MQ methods.

Naturally in the developing countries like Bangladesh the between-cluster variation is significantly higher than between-area variation. When the cluster level is ignored in the multilevel model specification, the cluster-level variance component would be merged with both the individual and area-level variance components (Tranmer and Steel, 2001a). In such situation, individual level and area-level variance components will mislead the distribution of the corresponding random errors. In similar manner, if the area-level random effect is found significant but ignored in ELL method, the area-level variance component will be merged with cluster variance component. So a careful diagnostic is necessary to select the ultimate multilevel model which will be used in the bootstrap procedure of the EBP and ELL methods.

Geographic specification of area and cluster are not needed to develop the MQ regression model. Area-specific MQ coefficient $\left(\theta_i\right)$ instead of area-specific random effect $\left(\eta_i\right)$ is calculated by taking the average of MQ coefficient of individual units $\left(q_k\right)$ belonging to the area and then area-specific MQ regression coefficients $\beta_\psi\left(\theta_i\right)$ are estimated. The area-specific regression coefficients $\beta_\psi\left(\theta_i\right)$ are used to predict response values for the individuals belong to the area. In this sense MQ method also considers the between-area variation and ignores the between-cluster variation. Thus similar to EBP method, the same question regarding the ignorance of cluster-variation in the distribution of response variable arises in MQ method.

The EBP method provides efficient estimates based on nested-error regression model under the normality assumption of target variable. For implementation of the EBP method, a correct transformation of the skewed expenditure or income variable is needed to achieve normality of the random errors. Diallo and Rao (2014) relax the normality assumption by considering a skew-normal distribution for the random errors. Both the ELL and MQ methods are free from such transformation to make the errors normally distributed. This is the main reason of applying generalized least squared (GLS) estimation method instead of ML or REML in the ELL method.

Both the ELL and EBP methods are sensitive to the presence of outliers in the distribution of response variable, while the MQ method provides robust estimates under the situation. Though the MQ method is distribution free of the random effects, it may be less efficient than EBP or ELL method if the model assumptions are unrealistically true (Tzavidis *et al.*, 2013).

The theoretical development of EBP and MQ methods are based on the matching of census and survey HHs which is quite impossible in real applications. The sampling ratios are usually found very small and so matching becomes redundant.

In terms of computation of MSE, the ELL method is economical and very fast compared to other methods. In both the EBP and MQ methods, the MSE estimation requires separate calculation with intensive computation. For big data of developing countries, implementation of the EBP and MQ methods might be a problem in terms of computational resources.

In reality, all small areas are not covered in the survey dataset and covered small areas comprise a few clusters. If only single cluster per area is available in the survey data set, the between-cluster variation and the between-area variation will be same and all the three methods will produce almost similar results since there is no chance of ignorance of any random effect in any method. Thus the differences among the three methods only possible to examine when both between-cluster and between-area variations exist in the distribution of response variable. In developing countries, it is observed that a significant number of small areas have single cluster and some areas have two or more clusters in the survey data. In such situation, there is always a possibility of having both cluster- and area- variability. From ELL point of view some area-specific contextual variable can be included in the model to remove the between-area variation.

Table 3.1: Basic comparison criteria of ELL, EBP and MQ methods

Issues regarding SAE Methods	**ELL.2L**	**ELL.3L**	**EBP**	**MQ**
Distributional assumptions of random errors:				
Parametric Bootstrap	Yes	Yes	Yes	-
Non-parametric Bootstrap	No	No	-	No
Ignored level in model specification	Area	None	Cluster	Cluster
Sensitivity of model misspecification	Yes	Yes	Yes	Yes
Transformation for normality	No	No	Yes	No
Influence of outliers	Yes	Yes	Yes	No
Necessity of matching census and survey HHs	No	No	Yes	Yes
Computational demand	Less	Less	High	High

Another issue arises in both the EBP and MQ methods with respect to prediction of income distribution for the small areas not covered in the survey data set. The prediction power for the non-sampled areas will not be as much as the sampled area in EBP and MQ method, while in ELL method the prediction power will be same for both sampled and non-sampled areas. For the non-sampled small areas, FGT measures and their associated MSEs are calculated based on the same bootstrap procedure in EBP and MQ methods similar to the ELL method. In case of MQ method, the parameters θ_i and $\boldsymbol{\beta}_{\psi}\left(\theta_i\right)$ are estimated based on $q = 50^{th}$ order MQ coefficients. Since without sample in an area all the estimators provide synthetic type estimates, it is expected that all estimators provide almost same estimates, however there is a doubt due to different types of generation of population in bootstrap procedure.

Previous studies to compare ELL, EBP, and MQ methods

A few studies have been done to compare these three methods simultaneously. Molina and Rao (2010) compared ELL and EBP methods through a simulation study considering only area-specific random effects in the population model and assuming very poor predictive power of the regression model. They mentioned that ELL method may provide worse result than the direct estimator in such situation.

Betti *et al.* (2007) conducted a simulation experiment to compare ELL and MQ methods utilizing two real datasets: the Living Standards Measurement Study (LSMS) 2002 and the Population 2001 Census of Albania. Using the survey data set, at first they developed a 2-level regression model considering HH at 1^{st} level and cluster at 2^{nd} level and then generate the simulated population by drawing the regression parameters from their sampling distribution and resampling the level-specific errors from the corresponding estimated residuals. They found the ELL method provides more biased estimates than MQ but the ranking of the poverty estimates doesn't influence much by the bias. They also noted that there were problem in MQ estimates when the small areas (districts) are not covered in the survey.

Tzavidis *et al.* (2013) explicitly discussed some technical issues of these SAE methods for poverty mapping. The authors conducted an empirical study as well as an application to Tuscany poverty data to compare ELL and MQ methods. They also found the MQ method is more efficient than the ELL method, sometimes even ELL method may provide poor outcome than the direct estimates as Molina and Rao (2010) found. The main reason of such results is the synthetic behavior of ELL method. However, the ELL method is not based on this framework. It assumes the hierarchical structure of the population where the lower levels have greater effect than the higher levels and so reduce the higher level impact on the model by including the corresponding level contextual variables.

Souza *et al.*, (2015) compare the ELL and EBP methods using 2010 Census data of Minas Gerais state of Brazil where the information of target variable (per capita HH income) is collected. They conducted a simulation study by selecting 400 samples following 2008-2009 Consumer Expenditure Survey (CES) from the known population and then the ELL and EBP methods are applied to estimate poverty incidence and poverty gap at municipality level. In the survey data set only 20% municipalities (195 out of 853) were available. They observed that the ELL estimator is performing better than the EBP estimator in terms of both relative bias (RB) and relative root mean squared error (RRMSE). Though both estimators provide synthetic estimates (most of the areas are out-of-sample), the EBP estimator shows higher overestimation than the ELL estimator. One of the main reasons might be the differences in generation of population in the bootstrap procedure: ELL generates population considering cluster-level variation, while EBP generates based on area-level variation.

3.4 An Empirical-based Simulation Study

A model-based simulation study has been planned with the aim of comparing the three SAE methods as well as selecting an appropriate SAE method for poverty mapping under the circumstances of developing countries. The first poverty mapping study in Bangladesh was conducted by BBS and UNWFP (2004) utilizing 2001 Bangladesh Population and Housing Census (hereafter referred as 2001 Census) and 2000 Household Income and Expenditure Survey (hereafter referred as 2000 HIES). The simulation study is based on these two datasets. A brief overview of the two datasets is given at first and then the generation procedure of pseudo census and the selection procedure of a sample from the census are explained in the following subsections.

3.4.1 Structure of Census and Survey Datasets

Sampling design of 2000 HIES was based on 1991 Census when Bangladesh was divided into 5 divisions (*Barisal*, *Chittagong*, *Dhaka*, *Khulna* and *Rajshahi*), 64 districts (*Zila*), and 507 sub-districts (*Upzila*). In 2001 Census *Sylhet* division was created from *Chittagong* division. The structure of 2000 HIES has been maintained in the simulation study. Instead of full census data, BBS selected 5% enumeration areas (EAs) from each sub-districts (the target small domains) by systematic sampling (BBS and UNWFP, 2004). The 5% census data covers all the administrative units up to sub-districts, while 2000 HIES covers 63 districts and 295 sub-districts (**Table 3.2**). Summary statistics of the number of EAs per sub-district given in **Table 3.3** indicate that at least 3 EAs per sub-district are available in the census data set while more than 75% sub-districts have single EA in the survey data set. In the simulation study only the 295 sub-districts sampled in 2000 HIES are considered for comparison purpose. It is noted that the EAs are the primary sampling units (PSU) in 2000 HIES and the clusters in our simulation study.

Sampling design of 2000 HIES

For 2000 HIES the country was stratified into 14 strata considering the residential classification of PSUs: Urban, Rural and Statistical Metropolitan Area (SMA) of 5 divisions (*Barisal* division had no SMA). Two-stage stratified sampling technique was followed to select 442 PSUs from 14 strata by stratified random sampling at the first stage and then 10 or 20 HHs were drawn from the selected PSUs by systematic sampling.

The number of HHs of the selected PSUs (complete list of HHs) ranges over 190-310 for rural and urban areas, and 100-287 for other urban areas. In rural and non-metropolitan urban areas 20 HHs were selected, while 10 HHs from metropolitan urban areas (SMA). A total of 7440 HHs were selected of which 7428 HHs are included in the study due to missing or incomplete information. Since 2000 HIES was conducted based on the 1991 Census, the HIES classification of PSU was different from that of 2001 Census. Census classification of PSUs shown in appendix **Table A3.1** is considered in the poverty mapping study.

Table 3.2: Summary statistics of Bangladesh administrative units in 2001 Census and 2000 HIES

	Division	District	Sub-district	EA	Household	Population ('000)
2001 Census	5	64	507	59990	25362321	130,523
5% 2001 Census	5	64	507	12908	1258240	6,156
2000 HIES	5	63	295	442	7428	38

Source: BBS and UNWFP (2004) and BBS (2001)

Table 3.3: Summary statistics of enumeration area (EA) by sub-districts in 2001 Census and 2000 HIES

	Number	Mean	SD	Min (%)	Max (%)
5% 2001 Census	12908	25.5	12.9	3 (0.02)	79 (0.007)
2000 HIES	442	1.5	1.9	1 (75.3)	10 (0.03)

Source: BBS and UNWFP (2004)

Sampling design for simulation study

In poverty analysis, matching between census HHs and survey HHs is an important issue for the selection of appropriate SAE method. Such matching in developing countries like Bangladesh, Nepal, and Bhutan is very difficult and unrealistic. For comparison of the three SAE methods, matching between survey and census units is maintained in the simulation study. This is the main reason of considering only 295 sub-districts in the simulations.

Since PSUs of a sub-district are classified into urban, rural and SMA, sub-districts are partitioned into unique portions (such as Urban PSUs, Rural PSUs, SMA PSUs) so that PSUs can be selected randomly according to their geographic location which confirms the structure of 2000 HIES. Appendix **Table A3.2** shows how the PSUs of 2000 HIES are selected over the country. In the simulation work, the same structure has been maintained when PSUs and HHs are selected. For simplicity in the selection process, some PSUs in 5% census data are merged with their neighbor PSUs to make the size at least 40.

Similar to 2000 HIES, 442 PSUs are selected with probability proportional to number of HHs and then HHs are selected randomly from the selected PSUs. The sampling ratio of sub-districts (based on 5% Census and 2000 HIES survey datasets) ranges from about 0.28% to 5.15% with mean 0.98% and median 0.83% which reveal a realistic picture of developing countries. The real sampling ratio should be much smaller than these if full census is considered.

Construction of pseudo census

In BBS and UNWFP (2004) study, 30 explanatory variables (including two-way interactions) at HH and sub-district levels (a list is given in appendix **Table A3.3**) were used to develop a regression model following a robust regression procedure which accounts complex survey design and provides consistent estimates of the covariance matrices. Instead of such complex regression procedure, multilevel analysis, one of most significant technique to account the hierarchical effect in the regression model, can be done to capture the variability in average per capita expenditure for consumption at different hierarchies. The sources of unequal selection probabilities (complex sampling design) can be controlled by incorporating design covariates in the model specification. However, if multilevel model is correctly specified but sampling weights are unaccounted by the covariates included in the regression model, a suitable weighting method is required to adjust the standard estimators of parameters for incorporating the design weights (Pfeffermann *et al.* 1998). For developing multilevel models for Bangladesh data, two design variables (division and urban-rural specification used for creating strata) as covariates and the hierarchies (household and clusters which are drawn with unequal probabilities) are incorporated in the regression model. It is expected that unequal selection probabilities are captured in the developed multilevel models and hence the estimated regression parameters and the variance components can be considered as reasonable and free from design-effect.

Log-transformed per capita consumption expenditure of HHs have been regressed on the selected explanatory variables to develop 2-level and 3-level models considering HHs, PSUs and sub-districts at first, second and third hierarchies respectively. The summary results of the multilevel models are given in **Table 3.4**. Null models indicate the significant contribution of PSUs $\left(\sigma_u^2\right)$ and sub-districts $\left(\sigma_\eta^2\right)$ in the variations of response variable. When auxiliary variables are included in the null model, both between-cluster and between-area variations are reduced. Though the variance component at sub-district level is very small, it is found still significant after including 30 explanatory variables and even 75% sub-districts have only single cluster. Both marginal and conditional R-squared values (according to Nakagawa and Schielzeth, 2013) are found slightly higher in 3-level model. So the 3-level model has been considered to generate the pseudo census values of per capita HH consumption expenditure in the simulation study. The 3-level model to generate log-transformed per capita HH expenditure is

$$\begin{gathered} \mathrm{y}_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(3)} + \eta_i + u_{ij} + e_{ijk};\ \ \boldsymbol{\beta}_{(3)} \sim MN\left(\hat{\boldsymbol{\beta}}_{(3)}, \hat{\mathbf{v}}\left(\hat{\boldsymbol{\beta}}_{(3)}\right)\right) \\ \eta_i \sim N\left(0, 0.0086\right),\ u_{ij} \sim N\left(0, 0.0186\right),\ e_{ijk} \sim N\left(0, 0.1135\right) \end{gathered} \tag{3.7}$$

where "MN" stands for multivariate normal and $\boldsymbol{\beta}_{(3)}$ is the vector of regression parameters under the 3-level model. The regression coefficients with their significance are given in appendix **Table A3.4**.

Table 3.4: Summary statistics of 2-level (2L) and & 3-level (3L) regression models fitted by restricted maximum likelihood method (REML)

Model	df	Marginal R^2	Conditional R^2	P-value	σ^2_{ε}	$\hat{\sigma}^2_u$	$\hat{\sigma}^2_{\eta}$
2L: Null	3	-	0.4102		0.2178	0.1515	-
3L: Null	4	-	0.3970	<0.01	0.2177	0.0770	0.0663
2L: Full	33	0.5941	0.6733		0.1135	0.0275	-
3L: Full	34	0.5962	0.6744	<0.01	0.1135	0.0186	0.0086

3.4.2 Simulation Process

In the simulation study, S=500 artificial populations are generated based on the superpopulation model (3.7) and then a sample has been drawn from each population to estimate the FGT measures applying four estimators: 2-level ELL (ELL.2L), 3-level ELL (ELL.3L), EBP and MQ. Parametric bootstrap procedure assuming homoskedastic random errors has been followed in the ELL methods to obtain FGT estimates by bootstrapping L=500 populations. In EBP and MQ methods, the out of sample response values are recalculated L=500 times for estimating FGT indicators by following parametric and non-parametric bootstrap procedure respectively. To compare the four estimators of FGT measures, relative bias (RB) and relative root mean squared error (RRMSE) are calculated based on these 500 artificial populations. The Spearman's rank correlation between the true and the estimated FGT measures has been calculated for each estimator in each simulation. The FGT poverty indicators are calculated considering upper poverty line (UPOVLN) estimated by the BBS (2003). The poverty line varies with strata which are slightly modified in the study of BBS and UNWFP (2004). The poverty lines shown in appendix **Table A3.5** are maintained to estimate the FGT poverty indicators in the simulation study.

In ELL methodology, the root mean squared errors (RMSEs) of estimated FGT measures are calculated easily and very quickly with the estimation of FGT measures. While the RMSEs are calculated separately in both EBP and MQ methods via more complex and time-consuming bootstrap procedures. Particularly for the large scale simulation work, they require extensive computational resources. To reduce the computation burden in this simulation study, RMSEs are calculated only for S=100 simulations. In ELL method, the RMSEs are calculated by bootstrapping L=100 populations in each simulation. To maintain the consistency of RMSE calculation, B=100 bootstrap population are generated in both EBP and MQ methods. In EBP method, for each bootstrap population FGT measures are calculated by generating L=100 out-of-sample vectors based on the response values of same sample units. While in MQ method, for each bootstrap population R=5 samples (more sample may reduce bias) are drawn via SRSWOR from each area and then FGT measures are estimated by generating L=100 out-of-sample vectors from each sample. A comparison of computation time has been shown in appendix **Table A3.6** to show the computational intensiveness of RMSE estimation in EBP and MQ methods compared to ELL method. The performances of the RMSE estimators are shown by comparing the area-specific RMSE with the simulated true RMSE based on the $S=100$ simulated true populations and by calculating coverage rates (CR) of nominal 95% confidence interval of the RMSE estimators.

3.5 Simulation Results

The distributions of area-specific RB and RRMSE expressed in percentages are shown in **Figure 3.1** and **Figure 3.2** respectively. The figures clearly indicate that both ELL.2L and ELL.3L estimators are performing better than the EBP and MQ estimators in terms of both RB and RRMSE for all the FGT indicators. The EBP estimator is likely to overestimate and the MQ estimator is likely to underestimate. These trends are obvious when FGT measure shifts from HCR to PS. No significant differences are observed in between the ELL estimators in either case of RB and RRMSE. Both the ELL estimators are providing lower and stable RBs and RRMSEs compared to the EBP and MQ estimators. The EBP estimator provides higher RB and RRMSE with large variation. Though MQ estimator provides downward RBs, their RRMSEs are found lower than those of EBP estimator.

Figure 3.1: Distribution of relative bias (RB, %) of ELL.2L, ELL.3L, EBP and MQ estimators of HCR, PG, & PS over 500 simulations

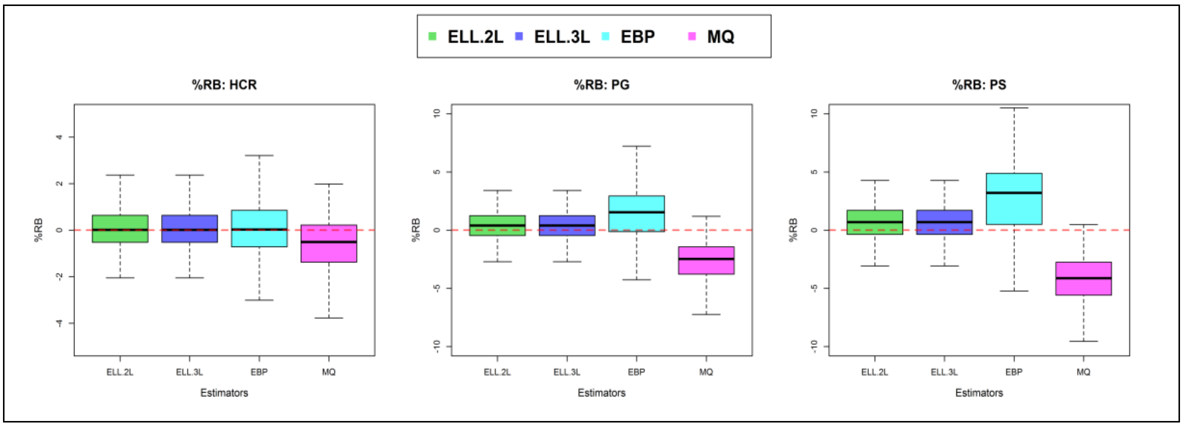
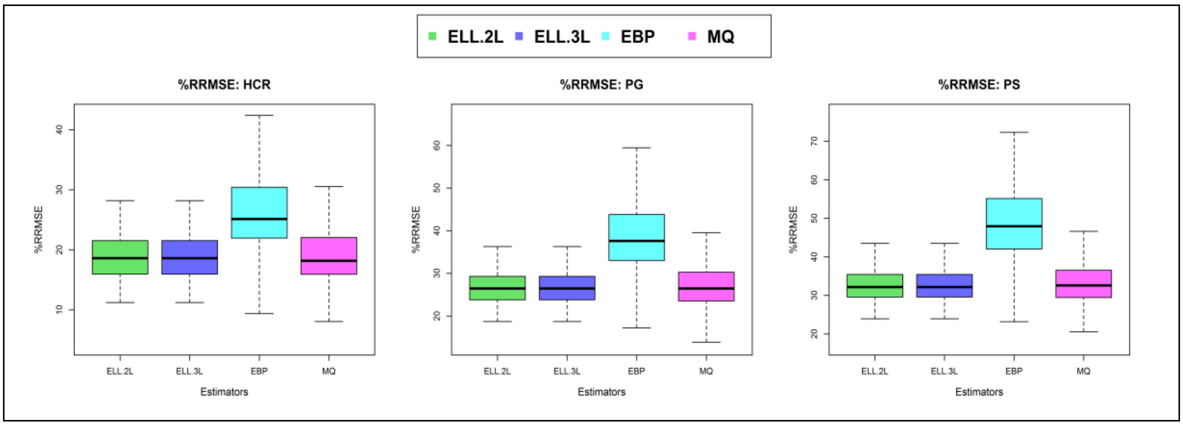
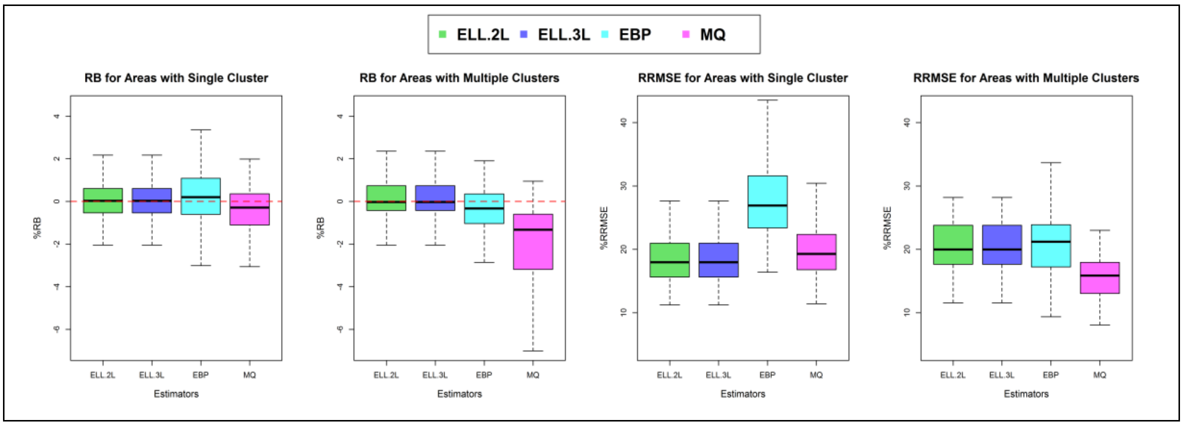


Figure 3.2: Distribution of relative root mean squared error (RRMSE, %) of ELL.2L, ELL.3L, EBP and MQ estimators of HCR, PG, & PS over 500 simulations



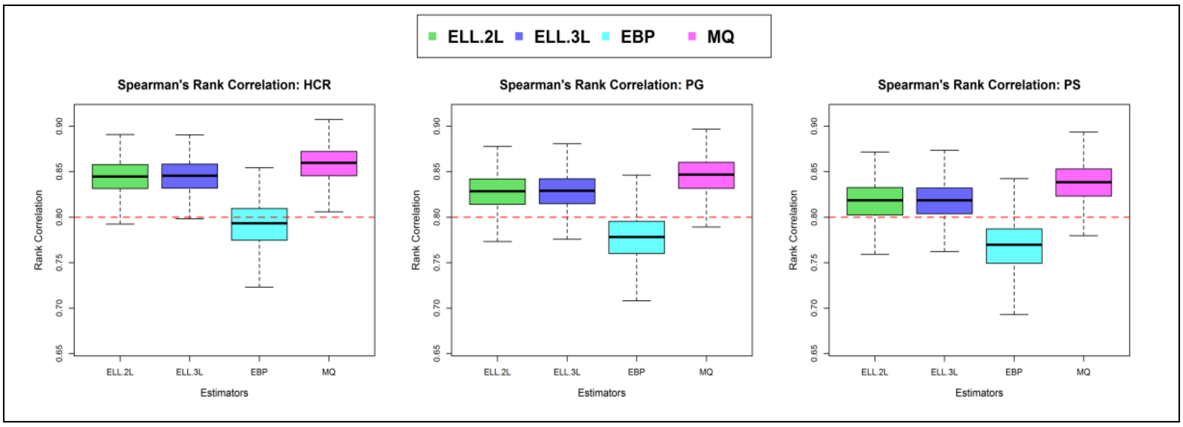
In the survey dataset, about 75% sampled areas have single cluster and the remaining have two or more clusters. So a substantial influence of the number of sampled PSUs per area is expected in the performances of FGT estimators. To examine the influence of this sample characteristic, RB and RRMSE of four estimators are plotted by dividing the areas according to single and multiple clusters sampled per area in the survey dataset. **Figure 3.3** shows that ELL estimators perform in the similar manner for both the situations but EBP and MQ estimators behave differently. The EBP estimator shows stable distributions of RB and RRMSE for areas with multiple clusters compared to those with single cluster. While MQ estimator behaves similar to ELL estimators for the areas with single cluster, and interestingly shows higher negative RBs and smaller RRMSE for the areas with multiple clusters. Thus the estimates calculated by EBP and MQ methods are significantly influenced by the number of sampled clusters per small area.

Figure 3.3: Distribution of relative bias (RB, %) and relative root mean squared error (RRMSE, %) of ELL.2L, ELL.3L, EBP and MQ estimators for HCR over 500 simulations by areas with single and multiple sampled clusters



The distribution of Spearman's rank correlations between the true and the estimated FGT measures are plotted in **Figure 3.4** to examine whether the bias issue does any change in the rank of the areas according to their poverty estimates. The figure shows that MQ estimator provides better performances in terms of ranking the sub-districts according to poverty situations, though it provides significant downward RB. The performances of MQ estimator also remain same for all the three FGT measures, while both the ELL and EBP estimators show a slightly downward trend of correlations with the degree of FGT measures. The EBP estimator shows the poorest correlations for all the FGT measures. These lower correlations and the higher RRMSEs of EBP estimator (**Figure 3.2**) indicate that the EBP estimators do not provide efficient FGT estimates in the considered scenario of Bangladesh.

Figure 3.4: Distribution of rank correlations between the true HCR, PG, & PS and their estimates by ELL.2L, ELL.3L, EBP and MQ estimators over 500 simulations



The area-specific averages of estimated RMSEs over the 100 simulations are plotted against the true simulated RMSE of FGT indicators in **Figure 3.5**. It is obvious that the estimated RMSE by the ELL.3L estimator will track the true simulated RMSE than the other estimators and the ELL.2L estimator will significantly underestimate the true RMSE. **Figure 3.5** exactly shows the expected trend. The estimated RMSEs obtained by EBP and MQ estimators also fail to track the true RMSE and their performances vary with the FGT indicators. With the increase of the degree of FGT measures, the estimated RMSEs of EBP estimator are likely to shift to the true RMSE while an opposite trend is observed for the MQ estimator.

The average estimated RMSEs are also plotted against population size in **Figure 3.6** (smoothed line) to see the influence of population size. It can also be seen that how the performances of the RMSE estimators vary with the number of sampled clusters per area in the survey data set. All RMSE estimators show downward trend with the population size as expected but the performances vary for the areas with large population from where multiple clusters are selected in the survey data set. Both the EBP and MQ estimators show stable RMSE for the areas with relatively smaller population but steadily declined RMSE for the larger areas. In case of EBP estimator, the differences between the true line and the estimated RMSE line gradually decrease with the degree of FGT indicator, but always show lower RMSE for the larger areas. The ELL estimators show almost similar trend of RMSE for all FGT indicators.

Figure 3.5: Average values of estimated root mean squared errors (RMSE) of estimated HCR, PG, & PS by ELL.2L, ELL.3L, EBP and MQ estimators against the true simulated RMSE over 100 simulations

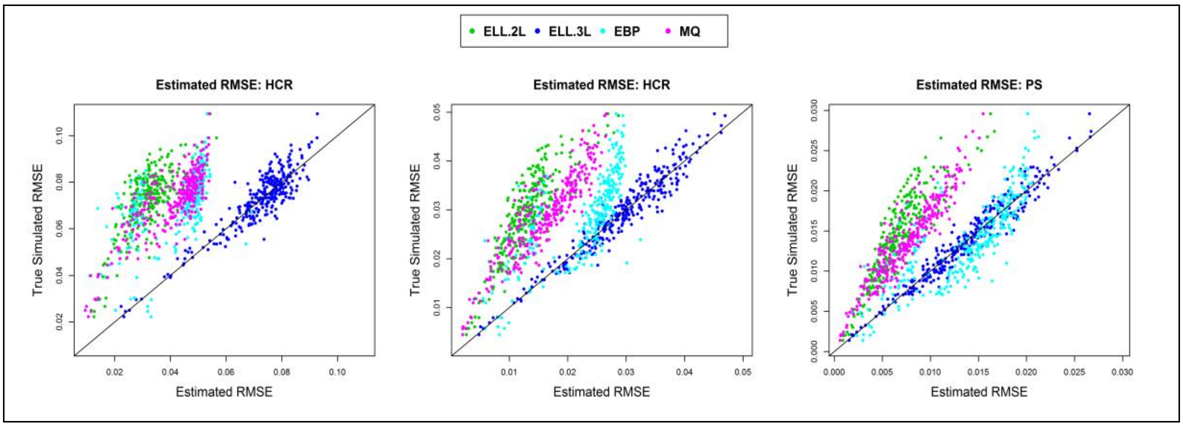
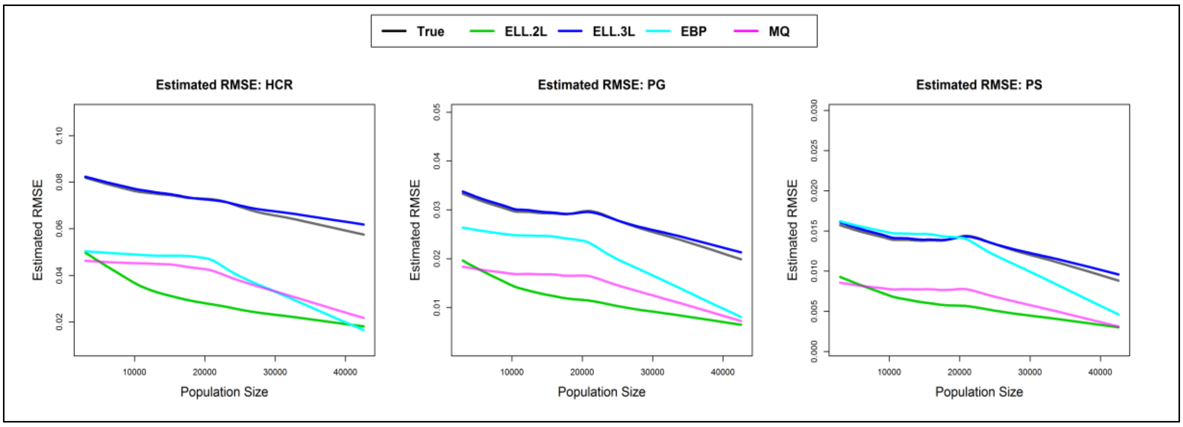
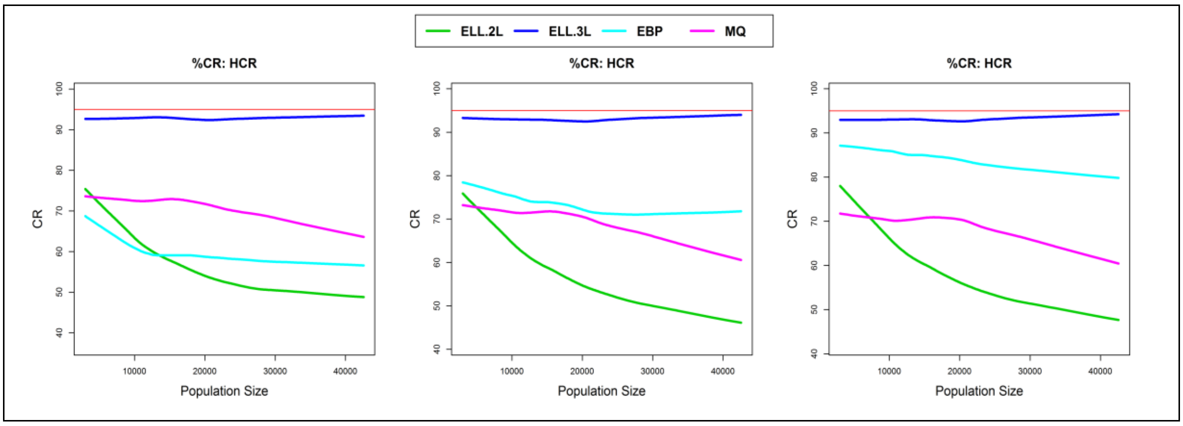


Figure 3.6: Average values of estimated root mean squared error (RMSE) of estimated HCR, PG, & PS by ELL.2L, ELL.3L, EBP and MQ estimators over 100 simulations against area-specific population size



The influence of estimated RMSEs is obvious in coverage rate (CR) of the estimators. **Figure 3.7** shows the smooth trend of coverage rate with the population size. Similar to the trend of estimated RMSE, the coverage rates decrease with the population size for all the estimators except the ELL.3L estimator. The ELL.2L estimator shows an exponential downward curve of coverage rates with the population size. The EBP and MQ estimators also show under coverage rates but performing better than the ELL.2L estimator. The performance of EBP estimator improves with the degree of FGT indicators, while no change has been observed for both the ELL.2L and MQ estimators. The performances of ELL.2L estimator in terms of estimated RMSE and CR are due to the influence of ignoring the between-area variability though it produces unbiased FGT estimate under this situation. For the EBP and MQ estimators, the lower estimated RMSE and lower coverage are due to failure of capturing the cluster-level variation.

Figure 3.7: Actual coverage rates (CR, %) of nominal 95% confidence intervals of ELL.2L, ELL.3L, EBP and MQ estimators of HCR, PG, & PS over 100 simulations against area-specific population size



3.6 Concluding Remarks

In the simulation study the census and survey datasets are formed to reveal a structure of Bangladesh (a representative of developing countries) where both cluster-homogeneity and area-homogeneity assumptions are violated. In such situation, the standard ELL estimator (ELL.2L) fails to capture the negligible but significant area variability, while EBP estimator ignores the higher cluster-level variation and merge the cluster-level variation with both individual-level and area-level variation. On the other side, MQ method fails to capture cluster-specific random effects in the prediction due to incapability of distinguishing the area-specific random effects from the cluster-specific effects.

The simulation results show that the standard ELL estimators perform better than the EBP and MQ estimators in terms of RB and RRMSE in the situation like Bangladesh where 75% small areas have single cluster in the sample and a negligible between-area variation exists in the population after combining a large number of explanatory variables in the regression model. As expected ELL.2L estimator underestimates the true RMSE and hence shows under coverage due to ignorance of area variability. The EBP and MQ estimators also show under coverage but perform better than ELL.2L estimator, which may be due to observed biases of EBP and MQ estimators.

The reasons for failure of EBP and MQ methods may be suspected as (a) higher between-cluster variation than between-area variation, (b) both HH- and area-specific random errors are generated from wrong distributions in EBP method, and MQ method wrongly considers cluster variation as area variation, (c) most of the small areas have single cluster in the survey data that may misguide the prediction of distribution function in both EBP and MQ methods. In such situation, 3-level model-based EBP and MQ methods can be thought as an alternative considering cluster as mid-level between HH and small area, or 2-level model-based EBP and MQ methods considering cluster rather than area as level-two since cluster variability is higher compared to area variability.

In terms of computational burden and matching the survey HHs with the Census HHs, the ELL method is safer, easier, faster, and economical than both the EBP and MQ methods. Moreover, most of the target administrative units are not available in survey data set and so separate calculation procedure is required in both EBP and MQ methods. For the non-sampled areas the prediction power may vary from the sampled areas in EBP and MQ methods due to different estimation procedure for both FGT measures and their MSEs. The poverty mapping study of Minas Gerais state of Brazil by Souza *et al.* (2015) is a perfect example where 80% target small areas were out of the survey data set.

The main criticism of basic ELL method is the underestimation of RMSE when area-homogeneity assumption is violated in reality. A 3-level model-based ELL method can be applied when the area variability is found significant in the regression model. From the simulation study, it is clear that the 3-level ELL method performs better than the traditional 2-level ELL method in terms of both FGT estimates as well as their estimated RMSE. Conversely, it is very sensitive in the sense that it may overestimate the true RMSE if there is no area effect indeed in the distribution of the response variable. Moreover, fitting a better 3-level model with appropriate variance component estimation method is critical in the situation like Bangladesh where availability of multiple clusters per small area in the survey data is a concerned matter to distinguish the cluster- and area-specific variability accurately. As an alternative a robust approach can be thought which will provide unbiased or approximately unbiased RMSE when 3-level population model is true one but the 2-level model is considered as the working model in ELL method. Such a robust ELL approach will be proposed in **Chapter Four**.

In the ELL method the regression model should have larger predictive power (as much as possible larger R-squared value) and so require many auxiliary variables. The lower the between-cluster and between-area variations, the better the performance of ELL method is expected due to its synthetic behaviour. In this regard, explanatory variables at different administrative levels should be included in the regression model that can capture variations at different levels. Having good R-squared value by inclusion of more explanatory variables in the model specification may not guarantee the capture of area variability that we found in the Bangladesh survey dataset. Thus proper care is needed in the application of naïve ELL methodology to free from the underestimated MSE and poor coverage rates in the absence of adequate contextual effects at PSU level.

Table A3.1: Classification of primary sampling unit (PSU) in 2000 HIES and 2001 Census

HIES PSU Classification (RMO)	HIES Stratum	Census PSU Classification (RMO)	Census Stratum
1=Rural	Stratum 1	1=Rural	Stratum 1
2=Non-metropolitan Urban	Stratum 2	2=Urban municipality	Stratum 2
4=Metropolitan Urban	Stratum 3	3=Other Urban	Stratum 1
5=Extra Metropolitan PSU selected for 2000 HIES	Stratum 3	4=Statistical Metropolitan Area (SMA)	Stratum 3

Table A3.2: Distribution of stratum, primary sampling unit (PSU), household (HH) in 2000 HIES by PSU Classification in 2000 HIES and 2001 Census

Stratum	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Division	Barisal		Chittagong			Dhaka			Khulna			Rajshahi			
No. PSU	26	10	60	10	32	69	10	70	29	8	22	68	12	16	442
No. HH	20	20	20	20	10	20	20	10	20	20	10	20	20	10	7440
HIES RMO	1	2	1	2	4,5	1	2	4,5	1	2	4,5	1	2	4,5	-
Census RMO	1,3	2	1,3	2	4	1,3	2	4	1,3	2	4	1,3	2	4	-

Table A3.3: List of auxiliary variables available in 2000 HIES and 2001 Census

Name	Meaning	Level	Name	Meaning	Level
electric	Has electricity	HH	*hhhprmed*	HH head not completed primary education	HH
ilattr_1	Sanitary latrine	HH			
ilattr_3	No latrine	HH	*child5p*	proportion of HH under 5	HH
iwater1	Drinking water from tap	HH	*literatep*	proportion of literate people in house	HH
ibuild_3	Semi-pucca house	HH	*femalep*	proportion of females in HH	HH
ibuild_4	Pucca house	HH	*ownaglnd*	Owns agricultural land	HH
owner	Own house	HH	*mhhsize*	Average HH size	Sub-district
rural	Designated as a rural area	HH	*depratio*	population under 15 or over 60 / population 15-59	Sub-district
workst2p	Proportion of employees/family helpers/other	HH	*paginc*	Proportion of HHs with agriculture as main income source	Sub-district
workst3p	Proportion of self-employed	HH	*idiv_1*	HHs in Barisal Division	Division
iincom_3	Main income source from transport, construction	HH	*idiv_2*	HHs in Chittagong Division	Division
num_hh	Number of HH members	HH	*idiv_4*	HHs in Khulna Division	Division
num_hh2	$(num\_hh-mean(num\_hh))^2$	HH	*idiv_5*	HHs in Rajshahi Division	Division

Note: variables beginning with 'i' are indicator variables

Table A3.4: Estimated regression coefficients of fitted 2-level and 3-level linear models by REML*

Variables	2-level Model				3-level Model			
	Est.	SE	t	p	Est.	SE	t	p
Intercept	6.94	0.14	49.54	0.00	6.88	0.15	44.82	0.00
electric	0.22	0.01	17.49	0.00	0.22	0.01	17.43	0.00
ilattr_1	0.11	0.01	8.13	0.00	0.11	0.01	8.12	0.00
ilattr_3	-0.13	0.01	-8.92	0.00	-0.13	0.01	-8.93	0.00
iwater_1	0.15	0.03	6.01	0.00	0.16	0.03	6.19	0.00
ibuild_3	0.11	0.02	5.65	0.00	0.11	0.02	5.82	0.00
rural	-0.09	0.04	-2.05	0.04	-0.08	0.04	-1.87	0.06
owner	0.12	0.01	8.15	0.00	0.12	0.01	8.02	0.00
ibuild_4	0.32	0.03	11.11	0.00	0.32	0.03	11.15	0.00
workst2p	-0.23	0.01	-17.62	0.00	-0.23	0.01	-17.68	0.00
workst3p	-0.14	0.02	-6.91	0.00	-0.15	0.02	-7.11	0.00
iincom_3	-0.07	0.01	-7.00	0.00	-0.07	0.01	-6.99	0.00
num_hh	-0.08	0.00	-21.79	0.00	-0.08	0.00	-21.77	0.00
num_hh2	0.01	0.00	11.16	0.00	0.01	0.00	11.16	0.00
hhhprmed	-0.16	0.02	-9.12	0.00	-0.16	0.02	-9.06	0.00
literatep	0.39	0.02	24.07	0.00	0.39	0.02	24.15	0.00
child5p	-0.53	0.03	-17.72	0.00	-0.53	0.03	-17.66	0.00
mhhsize	0.09	0.03	2.72	0.01	0.10	0.04	2.93	0.00
depratio	-0.38	0.09	-4.32	0.00	-0.39	0.09	-4.12	0.00
paginc	0.07	0.07	0.98	0.33	0.03	0.07	0.41	0.69
idiv_1	-0.04	0.04	-1.19	0.24	-0.03	0.04	-0.75	0.45
idiv_2	0.15	0.03	4.84	0.00	0.15	0.03	4.24	0.00
idiv_4	-0.17	0.03	-5.77	0.00	-0.16	0.03	-4.86	0.00
idiv_5	-0.17	0.03	-6.01	0.00	-0.14	0.03	-4.65	0.00
female*rural	-0.19	0.03	-6.49	0.00	-0.19	0.03	-6.49	0.00
ibuild_3*rural	0.15	0.03	5.11	0.00	0.15	0.03	5.02	0.00
owner*ibuild_4	0.10	0.03	2.99	0.00	0.10	0.03	3.17	0.00
rural*workst3p	0.19	0.02	8.25	0.00	0.20	0.02	8.43	0.00
rural*num_hh	0.01	0.00	1.97	0.05	0.01	0.00	1.92	0.06
rural*num_hh2	0.00	0.00	-2.67	0.01	0.00	0.00	-2.65	0.01
rural*hhhprmed	0.06	0.02	2.97	0.00	0.06	0.02	2.95	0.00
R^2	**M: 0.5941; C: 0.6733**				**M: 0.5962; C: 0.6744**			

* M: Marginal R^2 , C: Conditional R^2

Table A3.5: Upper (UPOVLN) and lower (LPOVLN) poverty lines in 2000 HIES by strata and districts

Stratum	Poverty Line		District by RMO (type of residence) ID
	LPOVLN	UPOVLN	
1	545.58	615.74	Rural Area of all districts in Barisal division
2	608.82	802.92	Urban Area of all districts in Barisal division
3	572.40	738.42	1201,1301,1901,3601,5801,9001,9101
3	582.43	718.94	301,1501,2201,3001,4601,5101,7501,8401
4	693.5	818.11	302,1202,1302,1902,2202,3002,3602,4602,5102, 5802,7502,8402,9002,9102
4	702.46	971.08	1502
5	702.46	971.08	SMA of all districts in Chittagong & Sylhet Divisions
6	540.24	590.63	2901,3501,3901,4801,5401,7201,8201,8601,8901,9301
6	547.63	659.44	2601,3301,5601,5901,6101,6701,6801
7	521.06	629.01	2902,3502,3902,4802,5402,5602,5902,6102,6802,7202,8202, 8602,8902,9302
7	648.65	893.14	2602,3302,6702
8	648.65	893.14	SMA of all districts in Dhaka Division
9	526.51	623.81	Rural Area of all districts in Khulna division
10	608.82	802.92	Urban Area of all districts in Khulna division
11	608.82	802.92	SMA of all districts in Khulna Division
12	509.56	581.99	1001,2701,3201,3801,4901,5201,7301,7701,8501,9401
12	585.58	689.59	6401,6901,7001,7601,8101,8801
13	557.05	725.74	Urban Area of all districts in Rajshahi division
14	557.05	725.74	SMA of all districts in Rajshahi Division

Source: BBS and UNWFP (2004)

Table A3.6: Computation time of ELL, EBP and MQ mean squared error (MSE) estimators via parallel computation using 50 cores

Estimator	Bootstrap, B	Resample, R	Out-of-sample Generation, L	Minutes
ELL	100	-	100	1.00
EBP	100	1	100	62.50
MQ	100	5	100	1500.00

CHAPTER FOUR

4. Robust Mean Squared Error Estimation of ELL Poverty Estimates

The ELL methodology is the small area estimation method developed by the World Bank for poverty mapping, and is widely used in developing countries. However, it has been criticized because of its assumption of negligible between-area variability when used to calculate small area poverty estimates. In particular, the mean squared errors of these estimates are significantly underestimated when this between-area variability cannot be adequately explained by the explanatory and contextual variables specified in the model. In this chapter a method of mean squared error (MSE) estimation for ELL-type estimates is proposed which is robust to the presence of significant unexplained between-area variability. Simulation studies have been carried out to assess the performance of the proposed method in comparison to standard ELL methodology when the area homogeneity assumption is violated.

This chapter is organized as follows: **Section 4.1** briefly demonstrates the ELL methodology with the underlying assumptions and their criticisms; **Section 4.2** describes the moment-based variance component estimation method when a 3-level model is perfectly specified and also when it is misspecified by ignoring a level; **Section 4.3** defines a robust variance estimator of area mean under this type of model misspecification; **Section 4.4** describes the steps of the standard ELL methodology and the robustified or modified ELL methodology; **Section 4.5** explains how the proposed modified ELL methodology performs via a simulation experiment; and at last some tentative conclusions are set up in **Section 4.6**.

4.1 Background

Over the last two decades, the poverty mapping small area estimation methodology developed by the World Bank (Elbers, Lanjouw and Lanjouw; 2002, 2003; henceforth ELL) has been applied in many developing countries. The basic idea underpinning ELL methodology is to combine data from a household (HH) survey with the HH records of a recent census or of an administrative registrar, though any HH-level linkage of these datasets is not required. The ELL method provides poverty estimates, with their estimated precision, at a specified local area level. Since the ELL method is based on an area homogeneity assumption - the probability of being poor given the explanatory variables in a small area is the same as in the larger region (Tarozzi and Deaton, 2009); the main criticism leveled against it is its assumption of random effects at the survey cluster level rather than at the local area level. To reduce the impact of possible between-area heterogeneity, a large number of explanatory and contextual variables are typically included in the regression model.

Tarozzi and Deaton (2009) claim that the underlying area homogeneity assumption of the ELL method may not be always true and hence may lead to misleading inference when this assumption is violated. They also question the assumption of homoscedastic, independent and identically distributed cluster random effects since small areas within a large area are likely to be interrelated. In such cases random cluster effects may be correlated if model regressors fail to adequately capture this between-cluster correlation. In response to these criticisms, Demombynes *et al.* (2007) show that the issues raised by Tarozzi and Deaton (2009) may be addressed partially if the area-level contextual variables are included in the model.

As far as criticism of the area-homogeneity assumption is concerned, Elbers *et al.* (2008) report results from a validation study based on the Census 2000 dataset of Minas Gerais state in Brazil, which is large enough to have heterogeneous local areas where the homogeneity assumptions must fail. They check the conditional independence assumption by developing a state-level model with municipality-level regressors and 853 municipality-level models with HH-level and enumeration-level regressors, and conclude that the state-level model is adequate for estimating welfare at area level as long as it captures local heterogeneity by including appropriate area-level regressors.

Inter-cluster and intra-cluster correlation coefficients are important components of the simulation phase of the ELL method. In particular, the variance of the welfare predictions obtained following this simulation may be understated when these correlations are large and are not explicitly accounted for in the regression model (Tarozzi and Deaton, 2009). Since the estimated location effect (variation due to higher levels of a population model) cannot be separated into area-level and cluster-level effects in the ELL method, one has to assume that this effect is either entirely a cluster-level effect - an optimistic assumption that rules out any correlation at higher level or is entirely an area-level effect - a conservative assumption that can lead to an upwardly biased mean squared error (MSE) estimates (Elbers *et al.*, 2002; Demombynes *et al.*, 2007). As Tarozzi and Deaton (2009) note, the conservative assumption could lead to imprecise (and hence unusable) estimates, while the optimistic assumption is obviously necessary for the ELL methodology to be valid, but could lead to downwardly biased MSEs if incorrect.

Elbers *et al.* (2008) applied both optimistic (i.e. the standard) and the conservative methods in the Minas Gerais study. Confidence interval generated under the optimistic method were mostly found to be narrower those generated by the conservative method. In particular, the authors found that 42% municipalities could be statistically distinguished from one another at a 95% level of confidence using the optimistic method with 35% using the conservative method. The authors conclude that the ELL method performs well even under some violation of its assumptions. However, neither method provides a good solution for the problem of between-area heterogeneity. Elbers *et al.* (2008) mention that the conservative method can be applied in the situation where some spatial correlation of errors remains after controlling for between-cluster heterogeneity. However, as noted earlier, Tarozzi and Deaton (2009) argue that this conservative method may then produce unstable estimates. Consequently, the question of which approach a researcher should adopt under different circumstances remains unanswered.

Molina and Rao (2010) also criticise the between-area homogeneity assumption and argue that a better random-effects specification for poverty estimation should have area-specific random effects than cluster-specific random effects. That is, these authors prefer the assumption of cluster-homogeneity to one of area-homogeneity. However, ignoring significant cluster-heterogeneity then raises the question of valid MSE estimates.

In this chapter a different approach to MSE estimation has been proposed when there is possible between-area heterogeneity associated with an ELL-based poverty estimation exercise. The basis of this approach is the analytical relationship between moment-based estimators of variance components under a 3-level hierarchical model and those obtained when fitting a 2-level model to the same data (i.e., when either level-two or level-three is ignored). This relationship allows us to identify a robust variance estimator of the area mean that is unbiased under the 3-level model and is also approximately unbiased under the 2-level model. This robust variance estimator is then to adjust the estimate of the level-two variance component used in the ELL simulation procedure. It should be noted that this adjustment is aimed purely at improving the estimated MSEs of the poverty estimates produced by the ELL method. Standard ELL methods are still used to calculate confidence intervals for poverty measures, since under our approach only the cluster-level variance component is modified in the ELL bootstrap procedure.

4.2 Estimation of Variance Components

Variance components of a multilevel model are usually estimated using maximum likelihood (ML), restricted ML (REML), Henderson method III, and moment-based least square (Searle *et al.*, 1992). Moment-based variance component estimators (MOM) are usually unbiased if there is an adequate sample size at each level of the multilevel model; otherwise estimators of higher level variance components may be biased and can be even negative. However, MOM estimators of variance components can also be analytically specified and are hence suited to an investigation of the relationship between the variance component estimates derived under a correctly specified multilevel model and those calculated under an incorrectly specified multilevel model corresponding to where a level of the true model is either ignored or is not available in the survey data (Tranmer and Steel, 2001a). Here the MOM variance component estimation method is focused.

4.2.1 Estimation of Variance Components under Perfectly Specified Model

Let y_{ijk} indicates the value of a target variable Y for k^{th} HH (level-one) belonging to j^{th} cluster (level-two) in i^{th} area (level-three). A 3-level linear model for Y then can be written as

$$y_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(3)} + \eta_i + u_{ij} + \varepsilon_{ijk};\ i = 1, 2, ..., D;\ j = 1, 2, ..., C_i;\ k = 1, 2, ..., N_{ij} \quad (4.1)$$

where $\eta_i \sim N\left(0, \sigma_{\eta(3)}^2\right)$, $u_{ij} \sim N\left(0, \sigma_{u(3)}^2\right)$ and $\varepsilon_{ijk} \sim N\left(0, \sigma_{\varepsilon(3)}^2\right)$ are level-specific zero-mean homoskedastic random errors respectively. In what follows, the sub-script (l) is used to indicate any parameter under a correctly specified l-level model. The covariance between values of Y for two distinct HHs is $\sigma_{\eta(3)}^2 + \sigma_{u(3)}^2$ if the HHs are from same cluster, $\sigma_{\eta(3)}^2$ if the HHs are from different clusters but from the same area, and zero otherwise. Suppose that a sample of n HHs is randomly drawn from this population using a two-stage cluster sampling technique, with clusters randomly sampled at the first stage and HHs within clusters randomly sampled at the second stage. Let C_{s_i} denote the number of clusters sampled in area i, with $C_s = \sum_{i \in s} C_{s_i}$ denoting the total number of sampled clusters, and n_{ij} $\left(i = 1, .., D;\ j = 1, .., C_{s_i}\right)$ is the number of HHs selected randomly from each selected clusters. The linear regression model implied by (4.1) can be fitted to the sample data using least squares method, in which case the moment-based estimates of the area-, cluster- and HH-specific random effects can be calculated as

$$\hat{\eta}_i = n_i^{-1} \sum_{j=1}^{C_{s_i}} \sum_{k=1}^{n_{ij}} \hat{e}_{ijk} = n_i^{-1} \sum_{j=1}^{C_{s_i}} n_{ij} \hat{u}_{ij}\ ,\ \hat{u}_{ij} = n_{ij}^{-1} \sum_{k=1}^{n_{ij}} \hat{e}_{ijk} \text{ and } \hat{\varepsilon}_{ijk} = \hat{e}_{ijk} - \hat{\eta}_i - \hat{u}_{ij}$$

respectively where $\hat{e}_{ijk} = y_{ijk} - \hat{y}_{ijk}$ and $\hat{y}_{ijk}$ denotes the least squares fitted value.

Under the model (4.1), expectations of the sample residual variances $\left(s^{(1)}, s^{(2)}, s^{(3)}\right)$ calculated at each level of the population can be expressed in terms of population variance components $\left(\sigma^2_{\eta(3)}, \sigma^2_{u(3)}, \sigma^2_{\varepsilon(3)}\right)$ as

$$E_3\begin{bmatrix} s^{(1)} \\ s^{(2)} \\ s^{(3)} \end{bmatrix} = E_3\begin{bmatrix} \dfrac{\sum_{ijk\in s}\left(\hat{e}_{ijk}-\hat{\bar{e}}_{...}\right)^2}{n-1} \\ \dfrac{\sum_{ij\in s} n_{ij}\left(\hat{\bar{e}}_{ij.}-\hat{\bar{e}}_{...}\right)^2}{C_s-1} \\ \dfrac{\sum_{i\in s} n_i\left(\hat{\bar{e}}_{i..}-\hat{\bar{e}}_{...}\right)^2}{D-1} \end{bmatrix} = \begin{bmatrix} 1 & \dfrac{n-\bar{n}_0^{(2)}}{n-1} & \dfrac{n-\bar{n}_0^{(3)}}{n-1} \\ 1 & \dfrac{n-\bar{n}_0^{(2)}}{C_s-1} & \dfrac{n-\bar{n}_0^{(3)}}{C_s-1} \\ 1 & \dfrac{\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}}{D-1} & \dfrac{n-\bar{n}_0^{(3)}}{D-1} \end{bmatrix}\begin{bmatrix} \sigma^2_{\varepsilon(3)} \\ \sigma^2_{u(3)} \\ \sigma^2_{\eta(3)} \end{bmatrix}$$

where $\bar{n}_{0i}^{(2)} = n_i^{-1}\sum_{j=1}^{C_{s_i}} n_{ij}^2$, $\bar{n}_0^{(2)} = n^{-1}\sum_{ij\in s} n_{ij}^2$, $\bar{n}_0^{(3)} = n^{-1}\sum_{i\in s} n_i^2$, and E_3 indicates expectation under the true 3-level model. The right hand side of the expectation above can be written as $\mathbf{A}_{(3)}\boldsymbol{\Lambda}_{(3)}$, where $\mathbf{A}_{(3)}$ is the coefficient matrix and $\boldsymbol{\Lambda}_{(3)}$ is the vector of true variance components. If $\mathbf{A}_{(3)}$ is non-singular, unbiased estimators of the variance components are given by the following expression, see also Tranmer and Steel (2001a, 2001b).

$$\hat{\boldsymbol{\Lambda}}_{(3)} = \left[\hat{\sigma}^2_{\varepsilon(3)} \quad \hat{\sigma}^2_{u(3)} \quad \hat{\sigma}^2_{\eta(3)}\right]^T = \mathbf{A}_{(3)}^{-1}\left[s^{(1)} \quad s^{(2)} \quad s^{(3)}\right]^T. \tag{4.2}$$

Unbiased estimators of the variance components of a perfectly specified 2-level model can be defined similarly. Detail estimation methods under correctly specified 2-level and 3-level models are given in **Appendix A.1** and **Appendix A.2** respectively**.** Under a perfectly specified 2-level, the estimators of corresponding variance components are expressed as below.

$$\begin{bmatrix} \hat{\sigma}^2_{\varepsilon(2)} \\ \hat{\sigma}^2_{u(2)} \end{bmatrix} = \begin{bmatrix} \dfrac{n-1}{n-C_s}s^{(1)} - \dfrac{C_s-1}{n-C_s}s^{(2)} \\ -\dfrac{n-1}{n-C_s}\dfrac{C_s-1}{n-\bar{n}_0^{(2)}}\left(s^{(1)}-s^{(2)}\right) \end{bmatrix}$$

Expectations of these estimators under the 3-level model become $E_3\left(\hat{\sigma}^2_{\varepsilon(2)}\right)=\sigma^2_{\varepsilon(3)}$ and $E_3\left(\hat{\sigma}^2_{u(2)}\right)=\left[\sigma^2_{u(3)}+R\sigma^2_{\eta(3)}\right]$ where $R=\left(n-\overline{n}_0^{(2)}\right)^{-1}\left(n-\overline{n}_0^{(3)}\right)<1$. Thus the estimator $\hat{\sigma}^2_{\varepsilon(2)}$ is still unbiased under the 3-level model.

4.2.2 Estimation of Variance Components under Misspecified Model

In many practical applications of the ELL methodology, a 2-level model is fitted instead of a 3-level model because most of the small areas of interest have just one sampled cluster. In such situation, a fitted 3-level can be numerically unstable. However, it is clear from the development above that ignoring a level in a hierarchy may significantly influence estimates of poverty measures due to the use of biased estimators of variance components. The influence of an ignored level in the model hierarchy may be lessened by including corresponding contextual variables at the regression model specification. However, the influence of the ignored level will still remain in the fitted parameters of the model particularly in the estimated variance components (Tranmer and Steel, 2001a). In this section an attempt has been made to show how the variance component of an ignored hierarchy is involved in the estimation of other variance components.

Suppose level-three of the 3-level population model (4.1) is ignored and a 2-level model is fitted to the survey data. Unbiased estimators of HH- and cluster-specific variance components under this 2-level model are expressed as

$$\begin{bmatrix}\hat{\sigma}^2_{\varepsilon(3/3)}\\ \hat{\sigma}^2_{u(3/3)}\end{bmatrix}=\mathbf{A}_1^{-1}\begin{bmatrix}s^{(1)}\\ s^{(2)}\end{bmatrix} \text{ with } \mathbf{A}_1=\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}=\begin{bmatrix}1 & (n-1)^{-1}\left(n-\overline{n}_0^{(2)}\right)\\ 1 & (C_s-1)^{-1}\left(n-\overline{n}_0^{(2)}\right)\end{bmatrix}$$

where $\mathbf{A}_1$ is a sub-set of $\mathbf{A}_{(3)}$ and the sub-script (3/3) is used to denote an estimator under a **3-**level model where level-three is ignored. Expectations of these variance component estimators under the (correct) 3-level model (4.1) are

$$E_3\begin{bmatrix}\hat{\sigma}^2_{\varepsilon(3/3)}\\ \hat{\sigma}^2_{u(3/3)}\end{bmatrix}=\mathbf{A}_1^{-1}E_3\begin{bmatrix}s^{(1)}\\ s^{(2)}\end{bmatrix}=\mathbf{A}_1^{-1}\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\end{bmatrix}\begin{bmatrix}\sigma^2_{\varepsilon(3)}\\ \sigma^2_{u(3)}\\ \sigma^2_{\eta(3)}\end{bmatrix}=\mathbf{A}_1^{-1}\mathbf{A}_{1,2}\mathbf{\Lambda}_{(3)}$$

where $\mathbf{A}_{1,2}$ consists of first and second rows of $\mathbf{A}_{(3)}$. After simplification, these expectations become

$$E_3\begin{bmatrix}\hat{\sigma}^2_{\varepsilon(3/3)}\\ \hat{\sigma}^2_{u(3/3)}\end{bmatrix}=\begin{bmatrix}\sigma^2_{\varepsilon(3)}\\ \sigma^2_{u(3)}+R\sigma^2_{\eta(3)}\end{bmatrix} \qquad (4.3)$$

where $R<1$. Thus when level-three of a 3-level model is ignored, $\hat{\sigma}^2_{\varepsilon(3/3)}$ is still an unbiased estimator of $\sigma^2_{u(3)}$. However, the expectation of $\hat{\sigma}^2_{u(3/3)}$ always less than the sum of $\sigma^2_{u(3)}$ and $\sigma^2_{\eta(3)}$. These expectations are exactly same as $E_3\left(\hat{\sigma}^2_{\varepsilon(2)}\right)$ and $E_3\left(\hat{\sigma}^2_{u(2)}\right)$ respectively shown in sub-section 4.2.1. In particular, we note that equality holds only if survey data are such that every level-three unit contains a single level-two unit since in that case $\bar{n}_0^{(2)}=\bar{n}_0^{(3)}$.

Now if level-two of the 3-level model is ignored, unbiased estimators of the remaining variance components are given by

$$\begin{bmatrix}\hat{\sigma}^2_{\varepsilon(3/2)}\\ \hat{\sigma}^2_{\eta(3/2)}\end{bmatrix}=\mathbf{A}_2^{-1}\begin{bmatrix}s^{(1)}\\ s^{(3)}\end{bmatrix} \text{ with } \mathbf{A}_2=\begin{bmatrix}a_{11} & a_{13}\\ a_{31} & a_{33}\end{bmatrix}=\begin{bmatrix}1 & (n-1)^{-1}\left(n-\bar{n}_0^{(3)}\right)\\ 1 & (D-1)^{-1}\left(n-\bar{n}_0^{(3)}\right)\end{bmatrix}.$$

Expectations of these estimators under the correct 3-level model can be approximated as

$$E_3\begin{bmatrix}\hat{\sigma}^2_{\varepsilon(3/2)}\\ \hat{\sigma}^2_{\eta(3/2)}\end{bmatrix}\approx\begin{bmatrix}\sigma^2_{\varepsilon(3)}+\left(1-\dfrac{\bar{n}_0^{(2)}}{\bar{n}_0^{(3)}}\right)\sigma^2_{u(3)}\\ \dfrac{\bar{n}_0^{(2)}}{\bar{n}_0^{(3)}}\sigma^2_{u(3)}+\sigma^2_{\eta(3)}\end{bmatrix}$$

since $\dfrac{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}{n-D} \approx 1-\dfrac{(n-1)\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_{0}^{(2)}\right)-(D-1)\left(n-\bar{n}_{0}^{(2)}\right)}{(n-D)\left(n-\bar{n}_{0}^{(3)}\right)} \approx 1-\dfrac{\bar{n}_{0}^{(2)}}{\bar{n}_{0}^{(3)}}$.

Similarly, if level-one of the 3-level model is ignored, unbiased estimators of the remaining two variance components are

$$\begin{bmatrix}\hat{\sigma}_{u(3/1)}^{2}\\ \hat{\sigma}_{\eta(3/1)}^{2}\end{bmatrix} = \mathbf{A}_3^{-1}\begin{bmatrix}s^{(2)}\\ s^{(3)}\end{bmatrix} \text{ with } \mathbf{A}_3 = \begin{bmatrix}a_{22} & a_{23}\\ a_{32} & a_{33}\end{bmatrix} = \begin{bmatrix}\dfrac{n-\bar{n}_{0}^{(2)}}{C_s-1} & \dfrac{n-\bar{n}_{0}^{(3)}}{C_s-1}\\ \dfrac{\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_{0}^{(2)}}{D-1} & \dfrac{n-\bar{n}_{0}^{(3)}}{D-1}\end{bmatrix}$$

and their expectations under the true 3-level model are approximated as

$$E_3\begin{bmatrix}\hat{\sigma}_{u(3/1)}^{2}\\ \hat{\sigma}_{\eta(3/1)}^{2}\end{bmatrix} \approx \begin{bmatrix}\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)^{-1}(C_s-D)\sigma_{\varepsilon(3)}^{2}+\sigma_{u(3)}^{2}\\ \left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)^{-1}\left(D-C_s n^{-1}\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)\sigma_{\varepsilon(3)}^{2}+\sigma_{\eta(3)}^{2}\end{bmatrix}$$

since $-\dfrac{(C_s-1)\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_{0}^{(2)}\right)-(D-1)\left(n-\bar{n}_{0}^{(2)}\right)}{\left(n-\bar{n}_{0}^{(3)}\right)\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)} \approx \dfrac{D-C_s n^{-1}\sum_{i\in s}\bar{n}_{0i}^{(2)}}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}$.

Please see Appendix A.3 for details concerning the derivation of these approximations**.**

When level-two of the 3-level model is ignored, expectations of the remaining variance component estimators ($\hat{\sigma}_{\varepsilon(3/2)}^{2}, \hat{\sigma}_{\eta(3/2)}^{2}$) include a term depend on $\sigma_{u(3)}^{2}$. When $\sigma_{u(3)}^{2}$ is large compared to $\sigma_{\eta(3)}^{2}$, the term containing $\sigma_{u(3)}^{2}$ in $E_3\left[\hat{\sigma}_{\eta(3/2)}^{2}\right]$ will have significant impact. Similarly when level-one of the 3-level model is ignored, the expected values of $\hat{\sigma}_{u(3/1)}^{2}$ and $\hat{\sigma}_{\eta(3/1)}^{2}$ contain terms depending on $\sigma_{\varepsilon(3)}^{2}$. Since $\sigma_{\varepsilon(3)}^{2}$ is typically much larger than the other two variance components, the terms can have a significant impact, leading to biased estimators of the second and third variance components.

In the following section, we show how these analytic relationships help us to define robust estimators of the variance components when a multilevel model is misspecified because an important level in the model is ignored. Recollect that level-three (the area of interest) is always ignored when implementing the ELL methodology (Elbers *et al.*, 2003), while level-two (cluster) is ignored when implementing the Empirical Best method of Molina and Rao (2010) as well as the M-quantile method of Chambers and Tzavidis (2006). Consequently, we expect these methods may lead to biased variance components estimation (and so biased MSE estimation) when used for poverty mapping.

4.3 Variance Estimator of Area Mean under Misspecified Model

The design of a household survey is usually based on a suitable hierarchical geography for a target population (e.g. a large region or a country). Consequently, it is inevitable that this hierarchy will be reflected in the distribution of a response variable measured in the survey. In the construction of working model for this variable, one needs to account for those levels in this hierarchy that correspond to significant components of the total variation of this variable across the target population. If a significant level (i.e., one that contributes significantly to total variation) is ignored in the working model, then this misspecification can lead to biased inference. In this section we show how the estimated variance of an unweighted area mean $\bar{Y}_i$ under a working 2-level model can be corrected for its bias when in fact the actual population model is 3-level.

To start, suppose that the 3-level model (4.1) holds, then the variance of an area-specific unweighted mean of the response variable $\bar{Y}_i$ and its plug-in estimator can be written as

$$\mathrm{Var}_{(3)}\left(\bar{Y}_i\right) = \sigma^2_{\eta(3)} + \bar{n}^{(2)}_{Ui}\sigma^2_{u(3)} + N_i^{-1}\sigma^2_{\varepsilon(3)} \text{ and } \hat{\mathrm{V}}_{(3)}\left(\bar{Y}_i\right) = \hat{\sigma}^2_{\eta(3)} + \bar{n}^{(2)}_{Ui}\hat{\sigma}^2_{u(3)} + N_i^{-1}\hat{\sigma}^2_{\varepsilon(3)}$$

where $\bar{n}^{(2)}_{Ui} = N_i^{-2}\sum_{j=1}^{C_i} N_{ij}^2 < 1$.

Now suppose that an incorrect 2-level working model is assumed. Under this incorrect 2-level model, the variance of $\bar{Y}_i$ and its plug-in estimator are defined as

$$\mathrm{Var}_{(2)}\left(\bar{Y}_i\right) = \sigma^2_{u(2)}\bar{n}^{(2)}_{Ui} + N_i^{-1}\sigma^2_{\varepsilon(2)} \text{ and } \hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right) = \hat{\sigma}^2_{u(2)}\bar{n}^{(2)}_{Ui} + N_i^{-1}\hat{\sigma}^2_{\varepsilon(2)}.$$

The expected value of these variance estimators under the correct 3-level model become $\mathrm{E}_3\left[\hat{V}_{(3)}\left(\bar{Y}_i\right)\right] = \sigma^2_{\eta(3)} + \bar{n}^{(2)}_{Ui}\sigma^2_{u(3)} + N_i^{-1}\sigma^2_{\varepsilon(3)}$ and $\mathrm{E}_3\left[\hat{V}_{(2)}\left(\bar{Y}_i\right)\right] = \left(R\sigma^2_{\eta(3)} + \sigma^2_{u(3)}\right)\bar{n}^{(2)}_{Ui} + N_i^{-1}\sigma^2_{\varepsilon(3)}$ which indicates that the expected value $\mathrm{E}_3\left[\hat{V}_{(2)}\left(\bar{Y}_i\right)\right]$ always underestimates the true variance $\mathrm{Var}_{(3)}\left(\bar{Y}_i\right)$ since $\bar{n}^{(2)}_{Ui} < 1$ and $R < 1$. An area-specific adjustment to $\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)$ that generally ensures an unbiased or approximately unbiased estimator of $\mathrm{Var}_{(3)}\left(\bar{Y}_i\right)$ is then

$$\hat{\mathrm{V}}^{M}_{(2)}\left(\bar{Y}_i\right) = \left\{\left(1/\bar{n}^{(2)}_{Ui}\right)\hat{\sigma}^2_{\eta(3)} + \hat{\sigma}^2_{u(3)}\right\}\bar{n}^{(2)}_{Ui} + N_i^{-1}\hat{\sigma}^2_{\varepsilon(2)} \tag{4.4}$$

which is unbiased under the true 3-level model. Note that the modified variance estimator (4.4) is also robust to model over-specification, where a 3-level model is taken as the working model, but in fact there is no level-three effect. In this case, we expect $\hat{s}^2_{h(3)}$ to be very small (and close to zero) if in fact a 2-level model actually holds, and so the first term of (4.4) will be negligible. A robust variance estimator for a weighted area mean can be easily obtained similarly which is illustrated in **Appendix A.5**.

4.4 Standard and Robustified ELL Methodology

The basic idea underpinning standard application of the ELL methodology for poverty mapping is Monte Carlo simulation of the population values of the HH income/expenditure. This simulation is based on a 2-level nested-error regression model linking logarithm of HH income to HH and cluster characteristics in the survey data. Two key assumptions are non-informative sampling given these characteristics and similar definitions for them in the survey and in the census. Model parameters are first estimated from income data collected in a household income and expenditure survey, and fitted income values are generated for the entire target population from this fitted model using the census values of these characteristics. It is standard to assume that the domains of interest (the small areas) are homogeneous and that between-area variation is due to between-cluster variation. Consequently, a 2-level regression model (HH as level-one, cluster as level-two) is used, ignoring the domains. If there is a negligible but significant domain effect, the ELL method can lead to unbiased estimates of poverty indicators, but with underestimated MSEs (Tarozzi and Deaton, 2009).

In what follows, it is useful to set out the basic steps of the standard ELL method. Let y_{ijk} denote the logarithm of HH oncome and let $\mathbf{x}_{ijk}$ denote the vector of HH and cluster characteristics that "explain" y_{ijk}. The parametric bootstrap version of this approach is as follows: **Step 1:** Fit a 2-level model to the survey values of y_{ijk} and $\mathbf{x}_{ijk}$ and hence calculate estimates of regression coefficients and variance components using a suitable estimation method (e.g. ML, REML, MOM, etc.). **Step 2:** Assign a cluster-specific error u_{ij}^{*} to each census cluster by making a random draw from a suitable parametric distribution, say $N\left(0,\hat{\sigma}_{u(2)}^{2}\right)$. **Step 3:** Assign a HH-specific error ε_{ijk}^{*} to each census HH by making independent random draw from a suitable parametric distribution, say $N\left(0,\hat{\sigma}_{\varepsilon(2)}^{2}\right)$. **Step 4:** Simulate the vector of regression parameters $\boldsymbol{\beta}_{(2)}^{*}$ by making a random draw from the sampling distribution of its estimator, i.e. the multivariate normal distribution with mean vector $\hat{\boldsymbol{\beta}}_{(2)}$ and covariance matrix $\hat{\mathbf{v}}\left(\hat{\boldsymbol{\beta}}_{(2)}\right)$. Step 5: Simulated population values for y_{ijk} via $y_{ijk}^{*}=\mathbf{x}_{ijk}^{T}\hat{\boldsymbol{\beta}}_{(2)}^{*}+u_{ij}^{*}+\hat{\varepsilon}_{ijk}^{*}$ and do exponentiation to recover simulated population values of HH income. The values of area specific parameters such as mean, quantiles, poverty indicators defined by this simulated population are recorded.

Step 6: Repeat steps (2)-(5) a large number of times (say, ***B***=1000). The mean and variance of the simulated area-specific parameters of interest are then used as their estimates and estimated MSEs respectively.

Now suppose after a proper statistical scrutiny, it is observed that the domains have a small but significant effect which has been ignored in the modeling. In consequence, implementation of the ELL procedure is likely to lead underestimated MSE. We focus on correcting this underestimation by repeating the ELL simulations, but with adjusted variance components that should lead to better MSE estimates. Since moment estimator of the level-one variance component $\hat{\sigma}^2_{\varepsilon(2)}$ is unbiased even when level-three variance component is ignored in the fitted model, an appropriate adjustment to the level-two variance component $\hat{\sigma}^2_{u(2)}$ is required. Three such adjustments are proposed below.

Adjustment 1: This adjustment assumes that we do not know the exact area-level population sizes (here number of HHs) N_i and cluster-level population sizes N_{ij}, and so we use the corresponding sample values as n_i and n_{ij} respectively. In this adjustment, the area-specific term $\overline{n}^{(2)}_{Ui}$ of $\hat{\mathrm{V}}^M_{(2)}\left(\overline{Y}_i\right)$ in (4.4) is replaced by its sample version $\overline{n}^{(2)}_{si} = n_i^{-2}\sum_{j\in s} n^2_{ij}$. Based on this adjustment, the estimated level-two level variance component $\hat{\sigma}^2_{u(2)}$ is replaced by $\hat{\hat{\sigma}}^2_u = k_1\hat{\sigma}^2_{u(2)}$ in the ELL procedure where $k_1 = \hat{\sigma}^{-2}_{u(2)}\left[\hat{\sigma}^2_{u(3)} + \hat{\sigma}^2_{\eta(3)} D^{-1}\sum_{i=1}^{D}\left(1/\overline{n}^{(2)}_{si}\right)\right]$. Note that the adjustment works better if the area-specific sample ratio is very close to the corresponding population ratio i.e. $\overline{n}^{(2)}_{si} \cong \overline{n}^{(2)}_{Ui}$ for $i = 1, 2, ..., D$; otherwise the correction will still underestimate the true variance. Also the population sizes for all small areas may not be available in the survey data set.

Adjustment 2: If the area- and cluster-specific population sizes N_i and N_{ij} are known, we can substitute these in the first adjustment. That is, we replace $\hat{\sigma}^2_{u(2)}$ by $\hat{\hat{\sigma}}^2_u = k_2 \hat{\sigma}^2_{u(2)}$ in the ELL procedure where $k_2 = \hat{\sigma}^{-2}_{u(2)} \left[\hat{\sigma}^2_{u(3)} + \hat{\sigma}^2_{\eta(3)} D^{-1} \sum_{i=1}^{D} \left(1 / \bar{n}^{(2)}_{Ui} \right) \right]$. Theoretically Aadjustment 2 will work better than Adjustment 1 provided area-specific population sizes (N_i) are known.

Adjustment 3: The procedure used for Adjustment 2 will give higher weight (k_2) to smaller areas and comparatively lower weight to larger areas. To reduce this variation, areas can be assigned to **H** strata according to their population size and then Adjustment can be carried out separately in each stratum. That is, the single level-two variance component used in simulation **Step 2** of the standard ELL method is replaced by stratum-specific adjusted level-two variance components of the form $\hat{\hat{\sigma}}^{2(h)}_u = k_3^{(h)} \hat{\sigma}^2_{u(2)}$ where $k_3^{(h)} = \hat{\sigma}^{-2}_{u(2)} \left[\hat{\sigma}^2_{\eta(3)} D^{-1}_{(h)} \sum_{i=1}^{D_{(h)}} \left(1 / \bar{n}^{(2)}_{Ui} \right) + \hat{\sigma}^2_{u(3)} \right]$ and $h = 1,..,H$. This method will work better if the population can be split into several large sub-populations on the basis of their population size.

Adjustment 4: This is the "Optimistic" method of Elbers *et al.* (2008, page 29). In this adjustment, a 3-level model is fitted with areas considering to level-three to obtaining the level-two and level-three variance components. Their sum $\tilde{\sigma}^2_{u(2)} = \left(\hat{\sigma}^2_{u(3)} + \hat{\sigma}^2_{\eta(3)} \right)$ is then used as new level-two variance component in the ELL simulation procedure. This adjustment is equivalent to the adjustment $\hat{\hat{\sigma}}^2_u = k_0 \hat{\sigma}^2_{u(2)}$ with $k_0 = \hat{\sigma}^{-2}_{u(2)} \left(\hat{\sigma}^2_{u(3)} + \hat{\sigma}^2_{\eta(3)} \right)$.

Note that in this case, the expectation of the variance estimator of area mean $\hat{V}_{(2)}\left(\bar{Y}_i\right)$ under the 3-level model becomes $\mathrm{E}_3\left[\hat{V}_{(2)}\left(\bar{Y}_i\right)\right]=\left(\sigma^2_{\eta(3)}+\sigma^2_{u(3)}\right)\bar{n}^{(2)}_{Ui}+N_i^{-1}\sigma^2_{\varepsilon(3)}$ where $\bar{n}^{(2)}_{Ui}<1$ implying that $\mathrm{E}_3\left[\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)\right]<\mathrm{E}_3\left[\hat{\mathrm{V}}_{(3)}\left(\bar{Y}_i\right)\right]$. That is, underestimation of the MSEs of the poverty estimates is still likely.

Adjustment 5: This is the "Conservative" method proposed by Elbers *et al.* (2008, page 29). In this adjustment, instead of bootstrapping a cluster-specific 2-level model (HH & cluster), an area-specific 2-level model (HH & area) is bootstrapped with area-level variance component $\tilde{\sigma}^2_{\eta(2)}=\left(\hat{\sigma}^2_{u(3)}+\hat{\sigma}^2_{\eta(3)}\right)$. In other words, the simulated location effect is at the area-level instead of at the cluster-level as in the standard ELL method. In this case the expectation of the variance estimator of the area mean $\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)$ under the 3-level model becomes $E_3\left[\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)\right]=\sigma^2_{\eta(3)}+\sigma^2_{u(3)}+\sigma^2_{\varepsilon(3)}N_i^{-1}$, which is greater than $E_3\left[\hat{\mathrm{V}}_{(3)}\left(\bar{Y}_i\right)\right]=\sigma^2_{\eta(3)}+\sigma^2_{u(3)}\bar{m}^{(2)}_{Ui}+\sigma^2_{\varepsilon(3)}N_i^{-1}$ since $\bar{n}^{(2)}_{Ui}<1$. That is, we expect this adjustment may lead to overstated MSEs when there is between-area variability in the data.

It is important to note that the modified approach to MSE estimation for the ELL-based poverty mapping defined by the adjustments above follows the same steps as the standard ELL method. The only difference is the use of an adjusted level-two variance component when simulating the location effect at **Step 2**. It is also noted that the purpose of the adjustment is solely to produce better MSE estimates. All other outputs including poverty estimates generated by the standard ELL simulation process remained unchanged. If the household-specific weights (say, HH size) are considered to calculate a weighted area mean, the correction will be similar by considering the number of population per area instead of number of HHs. The details are shown in **Appendix A.5**.

4.5 Numerical Evaluations

Two simulation studies have been conducted to explore the MSE estimation performance of the adjusted ELL-type MSE estimation methods described in the previous section. These are referred to as modified ELL (MELL) methods below. We focus on most commonly used FGT poverty indicators (Foster *et al.*, 1984) discussed in **Chapter Three**, i.e. head count rate (HCR), poverty gap (PG) and poverty severity (PS). The area-specific FGT poverty indicator is defined as $F_{\alpha i} = N_i^{-1} \sum_{j=1}^{C_i} \sum_{k=1}^{N_{ij}} \left(1 - E_{ijk}/t\right)^{\alpha} I\left(E_{ijk} < t\right)$ where E_{ijk} and t are per capita HH income and poverty line respectively. Area-specific values of HCR, PG, and PS are obtained by putting $\alpha = 0,1,2$ in $F_{\alpha i}$ respectively. Note that F_{0i} is the value of the area-specific income distribution at t.

The motivation of the modified ELL procedure is the analytic form of variance estimator of area-specific mean when the response variable is assumed to follow a normal distribution. In ELL procedure the log-transformed income/expenditure variable is also assumed to follow a normal distribution. Consequently the first simulation study explores the performances of these adjustments when the response variable Y is normally distributes and the interest is in estimation of the area-specific mean and the area-specific distribution (DF) of Y. The second simulation study is more realistic in that it considers MSE estimation when the logarithm Y of the response variable E (i.e. HH income) follows a normal distribution, and where the interest is in estimation of the area-specific values of the poverty indices HCR, PG, and PS defined by E.

4.5.1 Simulation Process

A 3-level hierarchical population consisting of D=75 small areas, C=1650 clusters, and H=180,450 HHs is used in the simulations. The number of clusters per area is allowed to vary between 15-29 clusters with corresponding cluster sizes varying between 96-120 HHs. The small areas are defined so that they can be partitioned into 5 strata according to their size (here the number of HHs per area). A two-stage random sampling procedure is used to first select 2-4 clusters randomly from each area and then randomly select 10 HHs from each of the selected clusters. Using the sampling procedure, a random sample of size n=2250 HHs is drawn in each simulation (see appendix **Table A4.1** for detailed regarding the simulated population and sample structure). The following three models are then used to generate the population values of the response variable:

$$y_{ijk} = 20 + \eta_i + u_{ij} + \varepsilon_{ijk};\ \varepsilon_{ijk} \sim N(0, 0.80);\ u_{ij} \sim N(0, 0.15);\ \eta_i \sim N(0, 0.05) \tag{4.5}$$

$$y_{ijk} = 20 + u_{ij} + \varepsilon_{ijk};\ \varepsilon_{ijk} \sim N(0, 0.80);\ u_{ij} \sim N(0, 0.20) \tag{4.6}$$

$$\begin{aligned} \log(y_{ijk}) &= \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(3)} + \eta_i + u_{ij} + \varepsilon_{ijk}; \\ &\varepsilon_{ijk} \sim N(0, 0.20);\ u_{ij} \sim N(0, 0.035);\ \eta_i \sim N(0, 0.015) \end{aligned} \tag{4.7}$$

where the values of the explanatory variables are drawn from a multivariate normal distribution as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_0 & \mathbf{X}_1 & \mathbf{X}_2 \end{bmatrix};\ \mathbf{X}_0 = \mathbf{1}_N;\ \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0.50 \\ 0.75 \end{pmatrix}, \begin{pmatrix} 1.5 & 0.10 \\ 0.10 & 0.95 \end{pmatrix}\right];\ \boldsymbol{\beta} = \begin{bmatrix} 6 & 0.5 & -0.55 \end{bmatrix}^T.$$

In the first simulation study, MSEs of the estimated area-specific means and DFs at the 10$^{\text{th}}$, 25$^{\text{th}}$, 50$^{\text{th}}$, 75$^{\text{th}}$, and 90$^{\text{th}}$ percentiles of the population distribution of the response variable are estimated for the first two populations (4.5) & (4.6). The second simulation study then considers estimation of the FGT poverty indicators with their estimated MSEs for each area based on the population model (4.7) and using the 10$^{\text{th}}$ and 25$^{\text{th}}$ percentiles of the overall income distribution as to define poverty lines. The standard MSE estimator generated under the 2-level model assumed by ELL (denoted ELL.2L with $K=1$ in what follows) together with its optimistic (Opt.ELL, $K=k_0$) and conservative (Cons.ELL / CON2L) versions, and the three modified 2-level model-based ELL estimators (MELL1: $K=k_1$, MELL2: $K=k_2$, & MELL3: $K=k_3$) are used to calculate estimated MSEs for the estimates of the area-specific parameters of interest. All HH-specific weights are set to unity, so all adjustment factors are calculated based on unweighted area means. A standard ELL estimator based on a 3-level working model (denoted as ELL.3L) is also used to estimate MSEs for comparison purposes.

The Monte Carlo simulations were repeated 1000 times with 1000 ELL simulations for each simulated population. In each simulation a sample is drawn from the simulated population and estimates of the area-specific parameters (Mean, DFs, FGT indicators) with their estimated MSEs are calculated. The differences between these estimated MSEs and the corresponding simulation-based actual MSEs are summarized using relative bias (RB) and relative root MSE (RRMSE), with both performance measures averaged over the simulations as well as over the areas of interest. Corresponding actual coverage rates (CR) of nominal 95% Gaussian-type confidence intervals are also shown. Note that the 2-level model based ELL estimates of target parameters are used for calculating the coverage rates for all MSE estimators based on a 2-level model.

An investigation was also carried out to see how MSE estimation performance is affected by whether or not the sample estimate of the level-three variance component is significantly different from zero. For this purpose, the 1000 simulated populations were divided into two groups based on the significance of the estimate of $\sigma^2_{\eta(3)}$ obtained by fitting a 3-level model to the sample data. This was done by carrying out a likelihood ratio test (LRT) of the hypothesis $H_0 : \sigma^2_{\eta(3)} = 0$. Note that the LRT is conservative in this case since the hypothesized value of $\sigma^2_{\eta(3)}$ is on the boundary of the feasible parameter space (Pinheiro and Bates, 2006). The nested 2-level and 3-level models are fitted using REML to valid the LRT (Snijders and Bosker, 2012). The reported p-value of LRT was then halved to minimize the boundary problem according to Pinheiro and Bates (2006).

4.5.2 Simulation Results

The values of RB shown in **Table 4.1** clearly show that the MSE estimator ELL.2L under standard 2-level ELL severely underestimates the actual simulated MSEs for ELL-based estimates of area means generated under the 3-level model (4.5). Furthermore, the optimistic and the sample-based adjusted MSE estimators corresponding to Adjustment 4 (Opt.ELL) and Adjustment 1 (MELL1) respectively, i.e. 2L(K=k_0) and 2L(K=k_1) in Table 4.1, similarly underestimates these MSEs. However, this underestimation is corrected by the population-based adjustment (MELL2: Adjustment 2 with K=k_2) and the stratification-based adjustment (MELL3: Adjustment 3 with K=k_3). Furthermore, the superior performance of these alternative MSE estimators extends to RRMSE and CR. In particular, we note that MELL2 and MELL3 perform very similar to the “true” MSE estimator, ELL.3L, which is calculated using the actual 3-level model (4.5) used to generate the population data. Finally, we see that the conservative version of the 2-level model-based MSE estimator (CON2L in Table 4.1) severely overestimates as expected.

In contrast, the standard ELL MSE estimator ELL.2L outperforms the other MSE estimators when the 2-level model (4.6) underpins the population data. All the other MSE estimators in this case tend to overestimate the MSE of the estimated area mean, with the optimistic (Opt.ELL) and sample-based adjusted (MELL1) estimators performing very similar to ELL.2L estimator. The more conservative adjustments associated with MELL2 and MELL3 estimators again perform similarly to the 3-level model-based MSE estimator ELL.3L, while the conservative 2-level model based estimator Cons.ELL is clearly of no practical use. Furthermore, when we restrict our assessment to simulated 3-level populations where $H_0 : \sigma^2_{\eta(3)} = 0$ fails to be rejected, we see no significant change in these results. See appendix **Table A4.2**.

Table 4.1: Area averaged values of relative bias (RB, %), relative root mean squared error (RRMSE, %), and actual coverage rate (CR, %) of nominal 95% confidence intervals generated by MSE estimates for estimated area-specific means under the 3-level (3L) and 2-level (2L) linear models

Model	Performance Measure	Standard and Modified ELL						
		3L	CON2L	2L(K=1)	2L(K=k_0)	2L(K=k_1)	2L(K=k_2)	2L(K=k_3)
3-Level Linear	RB	2.31	252.7	-80.77	-80.74	-72.62	4.56	1.60
	RRMSE	9.59	61.63	19.35	19.34	17.43	10.73	9.54
	CR	93.17	99.97	60.30	60.35	68.49	93.02	93.08
2-Level Linear	ARB	82.54	2033.87	12.83	12.82	19.41	81.81	81.86
	RRMSE	13.56	198.20	1.97	1.97	2.67	13.71	13.45
	CR	96.81	100.00	94.86	94.87	95.46	96.85	96.80

Turning on to Table 4.2, we see that the performances of the adjusted ELL-type MSE estimators for estimates of area-specific DFs under the 3-level population model (4.5) vary significantly according to the percentile at which the DF is calculated. The MELL2 and MELL3 MSE estimators show similar amounts of overestimation at extreme percentiles and similar amounts of underestimation in the middle of the distribution. In contrast, the RB values for ELL.2L, Opt.ELL and MELL1 show significant underestimation at all percentiles. Not surprisingly, ELL.3L records the best RB performance overall, while Cons.ELL is again the worst performer. RRMSE and CR performances are in line with these observations. Overall, MELL2 and MELL3 are clearly the best performers of all the 2-level model-based ELL type MSE estimators considered here, with RRMSE values that are comparable with those recorded by

ELL.3L. We also note that in this case the performance of MELL3 improves when the estimated value of $\sigma^2_{\eta(3)}$ is significant. See appendix **Table A4.2.**

Table 4.2: Area averaged values of relative bias (RB, %), relative root mean squared error (RRMSE, %), and actual coverage rate (CR, %) of nominal 95% confidence intervals generated by MSE estimates for estimated area-specific distribution functions (DFs) at different percentiles (***q***) under the 3-level (3L) linear model

Model	***q***	**Performance Measure**	**Standard and Modified ELL**						
			3L	CON2L	2L(K=1)	2L(K=k_0)	2L(K=k_1)	2L(K=k_2)	2L(K=k_3)
3-Level Linear	0.10	RB	6.02	296.20	-77.58	-77.52	-65.74	41.00	36.98
		RRMSE	1.90	13.16	3.35	3.35	2.86	2.82	2.60
		CR	93.93	99.75	64.63	64.61	75.12	96.09	96.03
	0.25	RB	3.74	266.92	-79.53	-79.49	-72.32	-29.95	-31.45
		RRMSE	3.07	20.73	6.08	6.08	5.48	2.71	2.73
		CR	93.51	99.93	60.92	60.92	68.26	88.58	88.63
	0.50	RB	2.71	255.55	-80.22	-80.19	-74.48	-46.30	-47.27
		RRMSE	3.69	24.64	7.64	7.64	7.10	4.53	4.61
		CR	93.2	100.0	59.8	59.8	66.4	84.5	84.4
	0.75	RB	3.06	264.27	-79.68	-79.64	-72.50	-30.31	-31.82
		RRMSE	3.06	20.58	6.10	6.10	5.56	2.74	2.76
		CR	93.57	99.89	60.89	60.84	68.26	88.66	88.60
	0.90	RB	4.61	290.52	-77.90	-77.84	-66.18	39.59	35.52
		RRMSE	1.88	12.99	3.38	3.38	2.89	2.79	2.56
		CR	93.87	99.70	64.44	64.51	75.01	96.03	96.03

In **Table 4.3** we see that if the population actually follows the working 2-level model (4.6), then all the MSE estimators tend to overstate the true MSEs of area-specific DFs at all the percentiles considered. This is similar to the behaviour observed for estimates of area-specific means in this situation (**Table 4.1**). The MSE estimators again fall into two groups - ELL.2L, Opt.ELL and MELL1 recording smaller upward biases (and RRMSE values) than ELL.3L, MELL2 and MELL3. While, the Cons.ELL performs very poor.

Finally, we see in **Table 4.4** that the performances of MELL2 and MELL3 in a realistic poverty estimation situation seem acceptable. The biases of these methods (compared with that of 3L, the gold standard here) get increasingly more positive as the index α of the FGT poverty measure increases. However, this seems preferable to the consistent (and large) negative biases and low values of CR associated with ELL.2L, Opt.ELL and

MELL1, even though these MSE estimates sometimes record smaller RRMSE values. We also see that MSE estimation performances when the poverty line is defined by the 25th percentile are generally better than those where this line is defined by the 10th percentile. This is unsurprising since in the former case more sampled HHs contribute to estimation of the poverty index as well as its MSE. However, it also clearly indicates that MSE estimation for ELL estimates of extreme poverty indices may be problematic unless great care is taken to ensure that there is no model misspecification.

Table 4.3: Area averaged values of relative bias (RB, %), relative root mean squared error (RRMSE, %), and actual coverage rate (CR, %) of nominal 95% confidence intervals generated by MSE estimates for estimated area-specific distribution functions (DFs) at different percentiles (***q***) under the 2-level (2L) linear model

Model	***q***	Performance Measure	Standard and Modified ELL						
			3L	CON2L	2L(K=10)	2L(K=k_0)	2L(K=k_1)	2L(K=k_2)	2L(K=k_3)
2-Level Linear	0.10	RB	71.3	1955.8	12.69	12.73	20.66	101.29	101.53
		RRMSE	2.29	37.39	0.45	0.45	0.61	3.37	3.31
		CR	96.34	100.0	94.56	94.55	95.22	96.57	96.55
	0.25	RB	78.93	1995.1	13.27	13.30	18.96	62.45	62.29
		RRMSE	4.23	63.92	0.65	0.65	0.84	3.15	3.09
		CR	96.68	100.0	94.79	94.80	95.30	96.57	96.58
	0.50	RB	82.59	2013.8	13.94	13.96	18.80	51.72	51.54
		RRMSE	5.38	78.62	0.75	0.75	0.95	3.02	2.96
		CR	96.89	100.0	94.99	94.94	95.40	96.66	96.62
	0.75	RB	78.87	1992.29	13.21	13.24	18.95	62.52	62.32
		RRMSE	4.23	63.74	0.64	0.64	0.83	3.15	3.09
		CR	96.75	100.0	94.93	94.91	95.39	96.67	96.68
	0.90	RB	69.76	1932.03	11.71	11.74	19.70	100.01	99.98
		RRMSE	2.27	36.99	0.43	0.43	0.59	3.36	3.29
		CR	96.35	100.0	94.54	94.59	95.23	96.58	96.57

Interestingly, both MELL2 and MELL3 perform better than ELL.3L in those simulated populations where the hypothesis of level-three variance component cannot be rejected based on the sample data. See the appendix **Table A4.2**. This suggests one should take extra care in defining the MSE estimator when the area-level variance component is insignificant. In such cases the modified estimators MELL2 and MELL3 are likely to be a better choice even though they are based on an incorrect model. Note that if a 2-level log-normal model is true instead a 3-level log-normal model, the modified MSE estimators for poverty indicators will perform similar to the MELL estimators for DFs that are observed in **Table 4.3**.

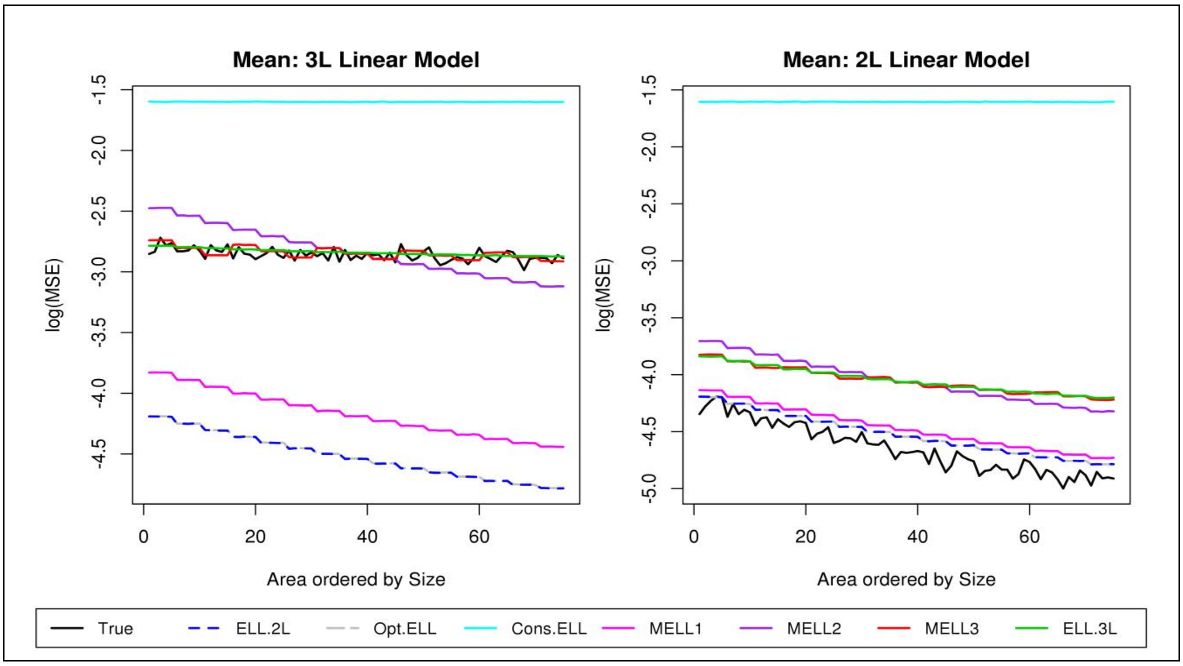
Table 4.4: Area averaged values of relative bias (RB, %), relative root mean squared error (RRMSE, %), and actual coverage rate (CR, %) of nominal 95% confidence intervals generated by MSE estimates for area-specific HCR, PG, and PS at poverty lines (***t***) corresponding to the 10th and 25th percentiles under the 3-level (3L) log-normal model

Indicator	t	Performance Measure	Standard and Modified ELL						
			3L	CON2L	2L(K=1)	2L(K=k_0)	2L(K=k_1)	2L(K=k_2)	2L(K=k_3)
HCR	0.10	RB	1.92	204.41	-80.12	-80.08	-70.63	37.11	33.12
		RRMSE	0.91	5.32	2.03	2.03	1.79	1.66	1.52
		CR	94.12	99.73	61.76	61.82	71.15	96.10	96.20
	0.25	RB	1.46	201.36	-82.05	-82.01	-74.10	-10.8	-13.18
		RRMSE	1.59	9.27	3.69	3.69	3.34	1.44	1.30
		CR	93.85	99.88	58.75	58.82	67.40	92.24	92.24
PG	0.10	RB	2.03	209.10	-79.67	-79.62	-68.38	120.35	114.16
		RRMSE	0.34	2.01	0.74	0.74	0.64	1.52	1.44
		CR	94.3	99.65	62.76	62.82	73.41	97.94	97.91
	0.25	RB	1.84	206.20	-81.93	-81.82	-72.75	27.27	23.53
		RRMSE	0.75	4.38	1.69	1.69	1.51	1.19	1.08
		CR	93.98	99.79	59.33	59.48	68.97	95.45	95.57
PS	0.10	RB	1.94	209.52	-78.28	-78.23	-65.56	216.52	209.15
		RRMSE	0.18	1.03	0.37	0.37	0.31	1.35	1.29
		CR	94.43	99.57	64.85	64.92	76.01	98.60	98.57
	0.25	RB	1.97	208.3	-81.23	-81.19	-71.07	68.13	63.21
		RRMSE	0.44	2.59	0.98	0.98	0.86	1.24	1.16
		CR	94.11	99.72	60.53	60.57	70.74	96.96	96.99

Figures 4.1 to 4.3 show the area-specific performances of the MSE estimators in the two simulation studies. Here log-scale area-specific estimated MSEs are averaged over simulations and then plotted against areas that are ordered according to population size. Figure 4.1 shows the performance of the MSE estimators for the area mean. This confirms that the adjusted MELL3 MSE estimator overcomes the underestimation problem of the standard ELL-based MSE estimator ELL.2L when the population data follow a 3-level model. Figure 4.2 shows that MELL2 and MELL3 MSE estimators for estimated area-specific DFs behave similarly to the ELL.3L MSE estimator, with overestimation at the two extreme percentiles and underestimation at the middle three

percentiles under the 3-level population. Overestimation at all percentiles under the 2-level model as anticipated is also clear. Figure 4.3 confirms that both the MELL2 and MELL3 MSE estimators overcome the underestimation problem of the standard ELL.2L estimator in the more realistic log-normal simulation that focuses poverty estimation. Figure 4.3 also shows that the underestimation associated with the standard ELL.2L MSE estimator increases while the overestimation of MELL2 and MELL3 estimators decreases with an increase in the poverty line.

Figure 4.1: Average of log-scale estimated MSE over simulations of area-specific estimated means under the 3-level (3L) and the 2-level (2L) linear models



All three figures show that the underestimation associated with estimated MSEs calculated via ELL.2L under a 3-level model is considerably larger than the corresponding overestimation by ELL.3L, MELL2 and MELL3 MSE estimators under a 2-level model. It is also worth noting that both underestimation of the true MSE under a 3-level model and overestimation under a 2-level model by the ELL.2L MSE estimator increases with the area-specific population size (Figures 4.1 and 4.2). The MELL3 MSE estimator appears to have the capacity to resolve this underestimation problem.

Figure 4.2: Average of log-scale estimated MSE over simulations of area-specific estimated distribution functions (DFs) under the 3-level (3L) and 2-level (2L) linear models

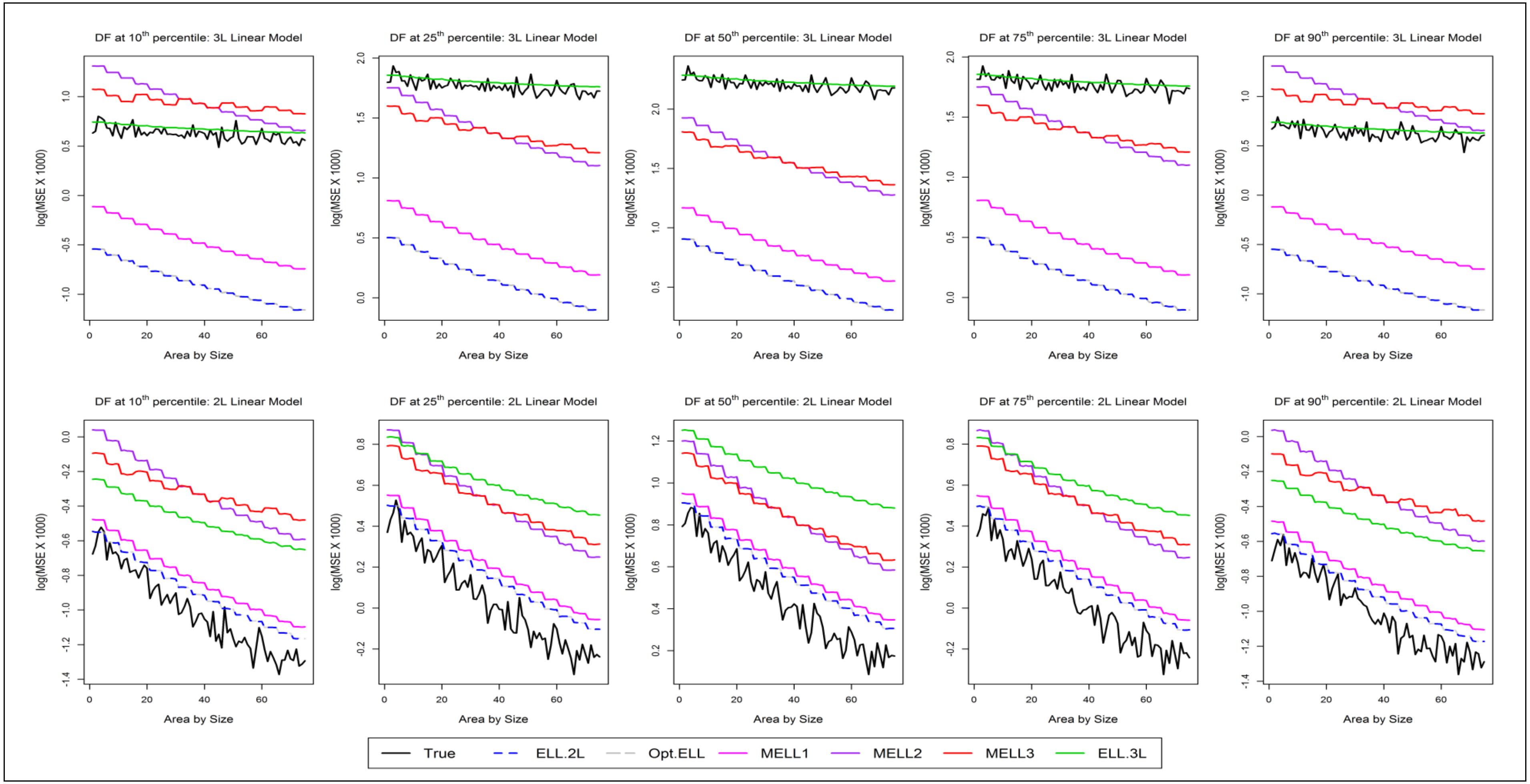
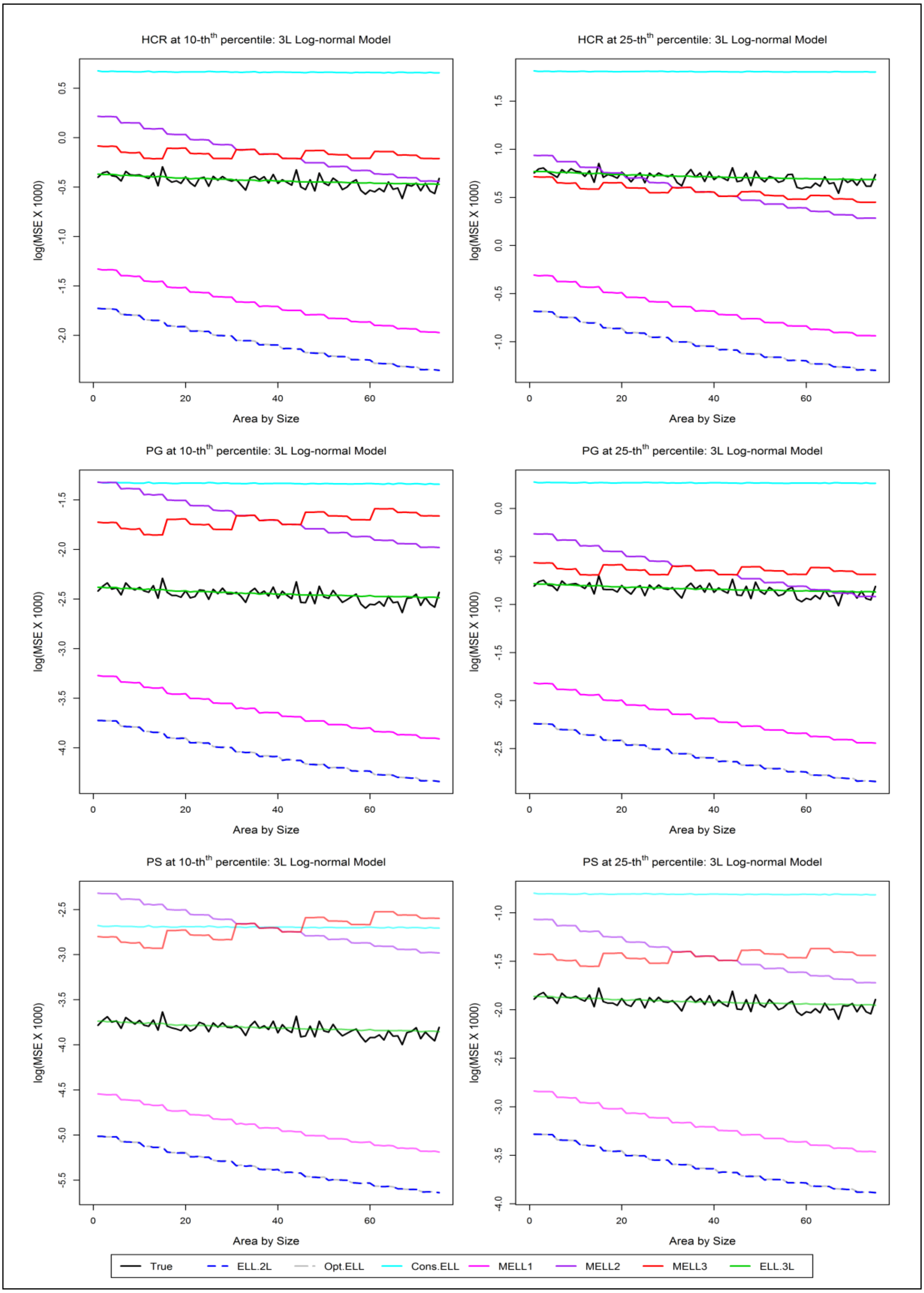


Figure 4.3: Average of log-scale estimated MSE over simulations of area-specific estimated HCR, PG, and PS at poverty lines corresponding to the 10^{th} and 25^{th} percentiles under the 3-level (3L) log-normal model



4.6 Concluding Remarks

In this paper we show how naive application of the standard 2-level model-based ELL method can lead to incorrect MSE estimation if its implied assumption of area homogeneity is incorrect. We propose a robust version of this MSE estimation method based on the relationship between estimated variance components under an incorrect 2-level model and corresponding estimates under a correct 3-level model. Our simulation studies provide some evidence that our modified MSE estimation method adequately corrects the bias of standard ELL-based MSE estimates of area-specific means, DFs, and FGT poverty measures. Our simulation studies support a modification corresponding to a simple stratified bias adjustment procedure, based on correcting the bias in the MSE estimator of the estimated area mean under a 2-level model-based ELL procedure. This adjustment works reasonably when used to correct the bias of the ELL estimators of the MSEs of FGT poverty indicators. We also note that our proposed bias adjustment procedure may be slightly conservative if in fact the underlying model has a 2-level structure (instead a 3-level structure) since then the MSE adjustment should be small and positive.

Note that the basic Monte Carlo approach to estimating MSE implicit in our modified MSE estimation method can be parametric (as described in Section 4.4) or non-parametric. In the latter case the estimated cluster-level residuals, the estimated cluster level residuals $\left(\hat{u}_{ij}\right)$ should be scaled so that the ratio of the Monte-Carlo variances $\mathrm{var}\left(\tilde{u}_{ij}\right)/\mathrm{var}\left(\hat{u}_{ij}\right)$ approximates the ratio of corresponding estimated variance components $\hat{\hat{\sigma}}_u^2/\hat{\sigma}_{u(2)}^2$ where $\tilde{u}_{ij(2)}=\hat{u}_{ij(2)}\hat{\hat{\sigma}}_u[\left(C_s-1\right)\sum_{ij\in s}\hat{u}_{ij(2)}^2]^{-1/2}$. This approach can be easily extended if the level-one (HH) errors are allowed to be heteroskedastic, as in most applications of ELL methodology.

Given that the proposed modification involves fitting a 3-level model to the sample data, a natural question to ask is why not use a 3-level ELL approach instead of a 2-level approach when a significant area effect is detected. However, our experience is that the adjusted ELL-type MSE estimation based on the (apparently) incorrect 2-level model that we propose often provides more stable MSE estimates than comparable MSE estimation based on the correct 3-level model. This seems particularly true in applications where a large number of sampled areas (e.g. 75% in Bangladesh study conducted by BBS and UNWFP, 2004) have a single sampled cluster. In such situations a 3-level ELL approach can be quite unstable.

Furthermore, the robustness of a 3-level approach to non-normality of area level errors, e.g. due to missing area-level contextual information in the model, is unknown. In contrast, the population size stratification used in MELL3 provides some robustness against this heterogeneity, though the issue of optimal choice of stratum boundaries for MSE estimation in this case remains an open question. Moreover, in terms of computation process, two separate bootstrap procedures are needed in the modified ELL methodology: basic bootstrap for poverty estimates and modified bootstrap for MSE estimates.

Finally, we note that the geographic location of the target small areas has been ignored in the adjustment methods proposed in this paper. Extension of these ideas to where there is spatial correlation between units at different levels of the data hierarchies in the survey and in the census is a topic for further research.

Table A4.1: Population and sample structure for the simulations studies described in section 4.5.1.

Population Structure (No. of Area, D **: 75**)	
No. of Clusters by Area No. of Total Cluster, C **= 1650**	15,15,15,15,15,16,16,16,16,16,17,17,17,17,17,18,18,18,18,18,19,19,19,19,19, 20,20,20,20,20,21,21,21,21,21,22,22,22,22,22,23,23,23,23,23,24,24,24,24,24, 25,25,25,25,25,26,26,26,26,26,27,27,27,27,27,28,28,28,28,28,29,29,29,29,29
No. of Units per Cluster by Area No. of Total Units, N **= 180450**	96,97,98,99,100,96,97,98,99,100,96,97,98,99,100,101,102,103,104,105,101, 102,103,104,105,101,102,103,104,105,106,107,108,109,110,106,107,108,109, 110,106,107,108,109,110,111,112,113,114,115,111,112,113,114,115,111,112, 113,114,115,116,117,118,119,120,116,117,118,119,120,116,117,118,119,120
Sample Structure (No. of Sampled Area, D **: 75**)	
No. of Sampled Clusters by Area No. of Total Cluster, C_s **= 225**	2,3,4,2,3,4,2,3,4,2,3,4,2,3,4,2,3,4,2,3,4,2,3,4,2,3,4,2,3,4,2,3,4,2,3,4,2, 3,4,2,3,4,2,3,4,2,3,4,2,3,4,2,3,4,2,3,4,2,3,4,2,3,4,2,3,4,2,3,4,2,3,4,2,3,4
No. of Sampled Units per Cluster No. of Total Units, n **= 2250**	10

Table A4.2: Area averaged values of relative bias (RB, %) for estimated area-specific means, distribution functions (DFs) at different percentiles (q) and FGT poverty indicators at poverty lines t corresponding to the 10th and 25th percentiles under the 3-level models based on significance of $\sigma^2_{\eta(3)}$.

Model	Parameter	Cases	q/t	3L	CON2L	2L(K=1)	2L(K=k$_0$)	2L(K=k$_1$)	2L(K=k$_2$)	2L(K=k$_3$)
3-Level Linear Model	Mean	$\sigma^2_{\eta(3)} \neq 0$	-	12.39	250.83	-80.88	-80.84	-71.72	14.96	11.59
		$\sigma^2_{\eta(3)} = 0$	-	-44.41	266.41	-79.97	-79.95	-76.5	-43.52	-44.71
	DF	$\sigma^2_{\eta(3)} \neq 0$	0.10	16.38	294.20	-77.71	-77.65	-64.38	54.56	50.05
			0.25	13.92	264.70	-79.66	-79.62	-71.59	-25.31	-26.95
			0.50	12.77	253.32	-80.35	-80.32	-73.94	-43.46	-44.51
			0.75	13.12	262.24	-79.80	-79.76	-71.77	-25.69	-27.36
			0.90	14.79	288.96	-78.00	-77.94	-64.82	53.02	48.39
		$\sigma^2_{\eta(3)} = 0$	0.10	-41.57	313.42	-76.48	-76.44	-71.52	-20.99	-22.92
			0.25	-43.10	283.10	-78.56	-78.53	-75.34	-50.88	-51.77
			0.50	-43.76	270.71	-79.34	-79.31	-76.71	-59.04	-59.64
			0.75	-43.66	278.27	-78.84	-78.82	-75.66	-51.40	-52.21
			0.90	-42.69	303.67	-77.06	-77.03	-72.21	-22.50	-24.15
3-Level Log-normal Model	HCR	$\sigma^2_{\eta(3)} \neq 0$	0.10	6.35	203.39	-80.22	-80.17	-70.19	43.84	39.59
			0.25	5.95	200.22	-82.14	-82.10	-73.75	-7.29	-9.80
		$\sigma^2_{\eta(3)} = 0$	0.10	-45.24	226.68	-78.35	-78.34	-74.64	-35.02	-36.30
			0.25	-46.60	223.31	-80.52	-80.51	-77.28	-48.10	-49.18
	PG	$\sigma^2_{\eta(3)} \neq 0$	0.10	6.47	208.18	-79.76	-79.71	-67.82	133.32	126.68
			0.25	6.35	205.12	-82.02	-81.97	-72.32	33.47	29.50
		$\sigma^2_{\eta(3)} = 0$	0.10	-45.10	231.66	-77.83	-77.83	-73.59	-20.05	-21.62
			0.25	-46.52	228.79	-80.42	-80.41	-76.80	-39.35	-40.63
	PS	$\sigma^2_{\eta(3)} \neq 0$	0.10	6.30	208.76	-78.38	-78.32	-64.90	237.03	229.07
			0.25	6.48	207.29	-81.31	-81.26	-70.58	77.30	72.04
		$\sigma^2_{\eta(3)} = 0$	0.10	-44.43	231.30	-76.30	-76.30	-71.65	-5.77	-7.38
			0.25	-46.08	231.07	-79.62	-79.62	-75.71	-30.96	-32.39

CHAPTER FIVE

5. Small Area Poverty Estimation under Heteroskedasticity

Multilevel models with homoskedastic nested errors are widely used for estimation of small area means, as well as for estimation of the small area distribution function for the underlying income (or expenditure) variable. However, this type of model may not be adequate if these nested errors are heteroskedastic. In particular, it is easy to show that although heteroskedasticity of the error terms in a multilevel model does not bias small area mean estimation, it can certainly bias estimation of the small area distribution function, and hence estimation of poverty measures based on this function. In this chapter we tackle an important example of this problem, which is estimation of small area distribution functions as well as poverty estimation for a finite population defined by a two-level superpopulation model with unknown heteroskedasticity at level-one. We adopt an estimation method that involves first estimating the (assumed) homoscedastic level-two variance component and then estimating the level-one variances using a non-parametric method that assumes the underlying heteroskedasticity is a smooth function of the auxiliary variables. The estimated variances from this model are then combined with the Chambers and Dunstan (1986) type smearing approach (hereinafter CD approach) to estimating small area distribution functions. Finally, a non-parametric bootstrap procedure based on this estimated distribution function is used to estimate small area quantities of interest along with their mean squared errors. The proposed methodology is compared with the well-known World Bank methodology for this case (Elbers, Lanjouw and Lanjouw, 2003). Results from simulation studies performed to compare these methods are then presented and an assessment of whether the proposed method can be considered as an alternative to World Bank poverty mapping methodology under level-one heteroskedasticity is provided.

This chapter is organized as follows: **Section 5.1** briefly describes the standard methodology of estimating small area distribution functions as well as poverty indictors with the corresponding assumptions and problems in reality; **Section 5.2** describes estimation of small area distribution function for a finite 2-level population with homoskedastic errors at both levels; **Section 5.3** describes at first the procedures of estimating level-two variance component and the level-one error variances, then the ELL procedure and the proposed CD approach considering level-one heteroskedastic random errors in the same finite population; **Section 5.4** demonstrates several realistic simulation studies to compare the mentioned methodologies; **Section 5.5** discusses the findings; and at last the concluding remarks are set out in **Section 5.6**.

5.1 Background

Standard multilevel models are developed based on the well-known homoskedastic (HM) nested-error regression model (Battese *et al.*, 1988) for estimating small area linear parameters (e.g. means or totals) of a finite population (Rao, 2003). These multilevel models can also be applied to estimate the small area distribution function and the corresponding shape measures such as medians, quartiles, and percentiles (Molina and Rao, 2010; Salvati, *et al.*, 2012). However, the standard multilevel model with strong assumption of HM errors is rare in practice, particularly in clustered, longitudinal and time series data. As for example, in a clustered data the unit-level residuals may vary with the characteristics of the corresponding units and also the cluster-level residuals with the size of the clusters. Heteroskedasticity of the residuals may also appear due to some practical consequences including model misspecification, existence of outliers, measurement errors, and asymmetry of dependent variables. If the nested errors are found heteroskedastic (HT), the developed multilevel models based on the corresponding HM assumption will be inadequate and may produce bias estimates for small area distribution function. The main task in multilevel modelling with HT nested errors is to find sources of heteroskedasticity and then modelling the heteroskedasticity. The HM variance components are also required to estimate considering other HT nested errors. The main aim of this chapter is to deal with estimation of small area distribution functions as well as poverty estimation for a finite population conforming to a 2-level superpopulation model with HM level-two errors, and HT level-one errors whose variances are assumed to follow an unknown smooth function of the auxiliary variables.

A widely used method for estimating distribution function of a finite population conforming to a linear superpopulation model is smearing-based prediction method proposed by Chambers and Dunstan (1986), hereinafter denoted as CD. The CD approach can also be applied to estimate distribution function for a finite population with unknown HT errors (Lombardía, *et al.*, 2005). In multilevel analysis, the CD approach can be easily implemented by developing predictive model at a higher level such as area-specific random effects model (Rao, 2003) and M-quantile model (Chambers and Tzavidis, 2006); and then utilizing the corresponding identically and independently distributed sample residuals in the smearing approach (Chambers and Pratesi, 2013). The CD-based approach requires an area- or cluster-specific model fitted to the sample data and the estimated sample residuals to predict the small area estimates. If a two-fold model is found as a best fit for the sample data, CD approach can also be implemented in the similar way. An important note is that the CD approach provides an analytical form of small area distribution function estimator based on the predictive model.

The basic poverty indicators defined by Foster *et al.* (1984) (hereafter, denoted by FGT) are calculated as $F_{\alpha i} = N_i^{-1} \sum_{k=1}^{N_i} \left(\frac{t - E_{ik}}{t} \right)^{\alpha} I\left(E_{ik} < t\right)$ for a small area i and poverty line t, where E_{ik} is the expenditure of k^{th} household (HH) and N_i is the number of households (HHs) in i^{th} area. The values of the parameter $\alpha = 0,1,2$ provide respectively head count rate (HCR), poverty gap (PG), and poverty severity (PS). The poverty indicator $F_{\alpha i}$ can be expressed as a function of small area distribution function as $F_{\alpha i} = \int_0^t \left(\frac{t - E_{ijk}}{t} \right)^{\alpha} dF_i\left(E_{ijk}\right)$ where $F_i\left(t\right) = N_i^{-1} \sum_{k=1}^{N_i} I\left(E_{ik} < t\right)$ is the distribution function at t for the i^{th} area. The FGT indicator $F_{\alpha i}$ provides area-specific distribution function $F_i\left(t\right)$ when $\alpha = 0$. Thus the main target of poverty estimation is to find an estimator of small area distribution function which leads to estimate the poverty indicators.

The most common method of poverty estimation is the World Bank method proposed by Elbers, Lanjouw, and Lanjouw (2002, 2003), henceforth mentioned by ELL. The ELL estimators of poverty indicators are based on simulation of area-specific distribution function. Thus the standard ELL method can be easily applied for estimating small area distribution function of a finite population following a 2-level superpopulation model with HM or HT nested errors at level-one. The method can also be implemented by developing three-level model if the third level is found to have significant contribution in the variation of target response variable.

In poverty mapping study, the ELL methodology considers spatial correlation (a 'location effect' common to all households in the same cluster has same distribution regardless of clusters) and heteroskedasticity in the HH component of the error term. Heteroskedasticity is considered at HH-level but not at cluster-level due to less number of clusters for a particular small area in the survey data (Elbers *et al.*, 2003). The level-one error variance function is developed through a parametric logistic function of the squared level-one residuals. The development of this parametric function (known as "alpha" model) is based on the assumption that the variance function is a monotone function of one or more explanatory variables. Since the method uses a logit model of squared residuals, the variance estimates are expected to be distributed as an s-shaped curve within a fixed range estimated from the survey data set. Since estimation methodology is based on parametric model, it may prone to model misspecification. Moreover, fitting an alpha model for heteroskedasticity in ELL approach is not a straightforward task. It requires fitting a logit model of squared conditional residuals considering susceptible explanatory variables or their transformations that can explain heteroskedasticity. The explanatory power of the alpha model is usually very low (less than 5%) in most of the poverty mapping studies (e.g. BBS and UNWFP, 2004; World Bank 2013). As far as we know, there are no particular study have been done to check the properties of the alpha model. A flexible non-parametric regression method can be thought of as an alternative of the parametric ELL approach for approximating the HT error variances.

The success of ELL methodology depends on the estimation of the variance components as well as the reduction of the ratio of between-cluster variation to total variation. It is noted that the better the estimation of location effect (spatial correlation), the better the precision of the ELL estimator is (Cuong *et al.*, 2010). When it is assumed that the level-one errors are HT, the level-two variance component is estimated at first and then the level-one error variances are estimated in ELL method.

Like ELL, our proposed method uses a moment approach to estimate the variance of HM level-two errors at first, and then a non-parametric smoothing technique is used to estimate the level-one error variances. The estimated residuals and their variances are combined with the CD-type smearing approach to estimate small area distribution functions. Finally, a non-parametric bootstrap procedure based on the estimated distribution function is used to estimate small area parameters along with their mean squared errors (MSE). The estimation procedure can be easily extended to a 3-level superpopulation model with level-one HT errors.

5.2 Estimation of Small Area Distributions and Poverty under Homoscedasticity

A 3-level hierarchical finite population is considered by assuming HH at level-one, cluster at level-two, and target small area at level-three. To maintain the area-homogeneity assumption of ELL methodology, nested random errors are considered only at HH and cluster. Suppose y_{ijk} indicates the value of the variable of interest Y for the k^{th} HH belonging to j^{th} cluster of i^{th} area. The superpopulation model for the finite population can be expressed as a 2-level nested-error regression model

$$y_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} + u_{ij} + \varepsilon_{ijk} \;;\; i = 1, 2, ..., D \;;\; j = 1, 2, ..., C_i ;\; k = 1, 2, ..., N_{ij} \quad (5.1)$$

where the cluster-level $\left(u_{ij}\right)$ and HH-level $\left(\varepsilon_{ijk}\right)$ random errors are assumed to be mutually independent with $u_{ij} \sim N\left(0, \sigma_{u(2)}^2\right)$ and $\varepsilon_{ijk} \sim N\left(0, \sigma_{\varepsilon(2)}^2\right)$ respectively. Since the individuals are nested within cluster, the covariance between two observations is $\sigma_{u(2)}^2$ if they are in the same cluster and zero (0) otherwise. Now consider a sample of size n has been drawn from the population following a two-stage cluster sampling design. In the sampling procedure, $C_s = \sum_{i=1}^{D} C_{s_i}$ clusters are drawn randomly at first covering all the D areas and then n_{ij} $\left(i = 1,..,D;\ j = 1,..,C_{s_i}\right)$ HHs are selected randomly from each selected clusters. Fitting a linear model via least square (LS) method, the cluster and HH specific errors are estimated as $\hat{u}_{ij} = n_{ij}^{-1}\sum_{k=1}^{n_{ij}} \hat{e}_{ijk}$ and $\hat{\varepsilon}_{ijk} = \hat{e}_{ijk} - \hat{u}_{ij}$ with $\hat{e}_{ijk} = y_{ijk} - \hat{y}_{ijk}$ respectively. The moment-based (MOM) estimators of level-specific variance components are obtained as

$$\begin{bmatrix} \hat{\sigma}^2_{\varepsilon(2)} \\ \hat{\sigma}^2_{u(2)} \end{bmatrix} = \begin{bmatrix} \dfrac{(n-1)}{(n-C_s)} & -\dfrac{(C_s-1)}{(n-C_s)} \\ -\dfrac{n-1}{n-\bar{n}_0^{(2)}}\dfrac{C_s-1}{n-C_s} & \dfrac{(n-1)}{n-\bar{n}_0^{(2)}}\dfrac{C_s-1}{n-C_s} \end{bmatrix} \begin{bmatrix} s^{(1)} \\ s^{(2)} \end{bmatrix} \tag{5.2}$$

where $\bar{n}_0^{(2)} = n^{-1}\sum_{ij\in s} n_{ij}^2$, $s^{(1)}$ and $s^{(2)}$ are HH- and cluster-level sample residual variances. The derivation of the estimators has been shown in **Appendix A.1**. Then the generalized least squares (GLS) estimates of the regression parameters and their corresponding variance-covariance matrix are obtained as $\hat{\boldsymbol{\beta}}^{gls}_{(2)} = \left(\mathbf{X}^T\hat{\mathbf{V}}_s^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\hat{\mathbf{V}}_s^{-1}\mathbf{y}\right)$ and $\hat{\mathbf{v}}\left(\hat{\boldsymbol{\beta}}^{gls}_{(2)}\right) = \left(\mathbf{X}^T\hat{\mathbf{V}}_s^{-1}\mathbf{X}\right)^{-1}$ where $\hat{\mathbf{V}}_s = diag\left\{\hat{\mathbf{V}}_c = \hat{\sigma}^2_{\varepsilon(2)}\mathbf{I}_{n_c} + \hat{\sigma}^2_{u(2)}\mathbf{1}_{n_c}\mathbf{1}_{n_c}^T\right\}$, $c = 1,...,C_s$. The estimated regression parameters $\hat{\boldsymbol{\beta}}^{gls}_{(2)}$ with $\hat{\mathbf{v}}\left(\hat{\boldsymbol{\beta}}^{gls}_{(2)}\right)$ are used as input in the estimation of small area parameters.

5.2.1 The ELL Estimator: Parametric Bootstrap (PELL)

The standard ELL estimator of small area distribution function is based on a parametric bootstrap (PB) procedure as follows. First, a set of regression parameters $\boldsymbol{\beta}^*_{(2)}$ is generated from the multivariate normal distribution $N\left(\hat{\boldsymbol{\beta}}_{(2)}^{gls}, \hat{\mathbf{v}}\left(\hat{\boldsymbol{\beta}}_{(2)}^{gls}\right)\right)$. Second, each cluster in the census is assigned a cluster-level error u_{ij}^* drawn from a suitable parametric distribution say $N\left(0, \hat{\sigma}_{u(2)}^2\right)$. Third, each HH in the census is assigned a HH-specific error ε_{ijk}^* from a suitable parametric distribution say $N\left(0, \hat{\sigma}_{\varepsilon(2)}^2\right)$. Fourth, simulated values of the response variable are generated via $y_{ijk}^* = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)}^* + u_{ij}^* + \varepsilon_{ijk}^*$ and a value of the area-specific parameter say $F_i^* = N_i^{-1} \sum_{j=1}^{C_i} \sum_{k=1}^{N_{ij}} I\left(y_{ijk}^* < t\right)$ is calculated for a specific value t. All the four steps are iterated for a large number of times say B =1000, and then mean and variance of the B area-specific estimates are considered as the ultimate ELL estimates and their mean squared errors (mse) respectively. The PB-based ELL estimator of area-specific distribution function is defined and denoted as

$$\hat{F}_i^{PELL} = B^{-1} \sum_{b=1}^{B} F_i^{*(b)} \;\&\; mse\left\{\hat{F}_i^{PELL}\right\} = B^{-1} \sum_{b=1}^{B} \left\{F_i^{*(b)} - \hat{F}_i^{PELL}\right\}^2 . \qquad (5.3)$$

The poverty indicators considering t as the poverty line is denoted by $\hat{F}_{\alpha i}^{PELL}$, α =0,1,2 and calculated as (5.3) by generating $F_{\alpha i}^* = N_i^{-1} \sum_{j=1}^{C_i} \sum_{k=1}^{N_{ij}} \left(\frac{t - y_{ijk}^*}{t}\right)^{\alpha} I\left(y_{ijk}^* < t\right)$ instead of F_i^* in the bootstrap procedure.

5.2.2 The ELL Estimator: Non-parametric Bootstrap (NPELL)

Instead of PB procedure, a non-parametric bootstrap (NPB) procedure can be followed to conduct the ELL estimation method. In NPB procedure, the level-specific errors are randomly drawn from the corresponding empirical distributions rather than the parametric distributions. Here all census clusters are assigned cluster-level error u_{ij}^*

drawn from the empirical distribution $\{\hat{u}_{ij},\ ij \in s\}$ via simple random sampling with replacement (SRSWR). Similarly, HH-level errors ε^{*}_{ijk} for each census HH are randomly drawn from the sample residuals $\{\hat{\varepsilon}_{ijk};\ ijk \in s\}$ via SRSWR. However, if the estimated raw residuals at different levels are used in NPB, the bootstrap variation (empirical) of the residuals will overestimate the estimated variance components (Carpenter *et al.*, 1999). In such case the estimated raw residuals $(\hat{\varepsilon}_{ijk},\ \hat{u}_{ij})$ are required to adjust for utilizing in the bootstrap procedure. The estimated raw residuals can be scaled respectively as $\widehat{\varepsilon}_{ijk} = \hat{\varepsilon}_{ijk}\hat{\sigma}_{\varepsilon(2)}\left\{(n-1)^{-1}\sum_{ijk\in s}\hat{e}^{2}_{jk}\right\}^{-1/2}$ and $\widehat{u}_{ij} = \hat{u}_{ij}\hat{\sigma}_{u(2)}\left\{(C_s-1)^{-1}\sum_{ij\in s}\hat{u}^{2}_{ij}\right\}^{-1/2}$ so that the empirical variations approximately close to the estimated variance components (details are shown in **Appendix A.4**). These scaled residuals instead of raw residuals are utilized in the NPB procedure. Thus the basic differences between PB and NPB procedures are in second and third steps where the level-specific errors $\left(u^{*}_{ij}\ \&\ \varepsilon^{*}_{ijk};\ i=1,...,D;\ j=1,...,C_i;\ k=1,...,N_{ij}\right)$ are assigned to population units via SRSWR from the corresponding estimated scaled sample residuals $\{\widehat{u}_{ij}, \widehat{\varepsilon}_{ijk};\ ijk \in s\}$. Similar to the PB procedure, generation of the bootstrap values y^{*}_{ijk} and calculation of area-specific estimates F^{*}_{i} are repeated B times to calculate the ultimate estimates with their MSEs similar to (5.3). The NPB-based ELL estimators of small area distribution functions and their MSEs are denoted by $\hat{F}^{NPELL}_{i}$ and $mse\{\hat{F}^{NPELL}_{i}\}$ respectively.

5.2.3 The CD Estimator: Smearing Approach (CDSM)

As an alternative of NPB-based ELL estimator, a Chambers and Dunstan (1986) type estimator has been described based on the smearing approach. For this approach, the superpopulation model (5.1) is written as

$$y_{ijk} = E\left[y_{ijk} \mid \mathbf{x}_{ijk}\right] + e_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} + u_{ij} + \varepsilon_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} + u_{ij} + \sigma_{\varepsilon(2)} \gamma_{ijk} \qquad (5.4)$$

where $\gamma_{ijk} = \sigma_{\varepsilon(2)}^{-1} \varepsilon_{ijk} \sim N(0,1)$ are normalized errors. The first task in the proposed CD approach is to estimate the cluster-specific distribution function $F_{ij}(t)$ under the superpopulation model. Since $F_{ij}(t)$ can be partitioned into known sample and unknown non-sample parts as below, the task reduces to find an estimator of non-sample part $F_{r_{ij}}(t)$ of $F_{ij}(t)$ only.

$$F_{ij}(t) = N_{ij}^{-1}\left[\sum_{s_{ij}} I\left(y_{ijk} \le t\right) + \sum_{r_{ij}} I\left(y_{ijk} \le t\right)\right] = N_{ij}^{-1} n_{ij} F_{s_{ij}}(t) + \left(1 - N_{ij}^{-1} n_{ij}\right) F_{r_{ij}}(t)$$

The expectation of the non-sample part under (5.4) can be expressed as

$$E\left\{\sum_{r_{ij}} I\left(y_{ijk} \le t\right)\right\} = E\left[\sum_{r_{ij}} I\left(\sigma_{\varepsilon(2)}\gamma_{ijk} \le t - \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} - u_{ij}\right)\right]$$
$$= \sum_{r_{ij}} P\left\{\gamma_{ijk} \le \frac{t - \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} - u_{ij}}{\sigma_{\varepsilon(2)}}\right\} = \sum_{r_{ij}} G\left\{\frac{t - \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} - u_{ij}}{\sigma_{\varepsilon(2)}}\right\}$$

where G is the distribution function of the normalized errors γ_{ijk}. It is therefore natural to approximate $I\left(y_{ijk} \le t\right)$ by an estimator of $G\left\{\sigma_{\varepsilon(2)}^{-1}\left(t - \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} - u_{ij}\right)\right\}$ to estimate $F_{r_{ij}}(t)$. Since the normalized errors $\gamma_{ijk} = \sigma_{\varepsilon(2)}^{-1}\left(t - \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} - u_{ij}\right) = \sigma_{\varepsilon(2)}^{-1} \varepsilon_{ijk} \sim G,\ \ ijk \in U$ are assumed to be independent and identically distributed, an estimator of G is the empirical distribution of the standardized residuals $\hat{\gamma}_{ijk} = \hat{\sigma}_{\varepsilon(2)}^{-1} \hat{\varepsilon}_{ijk}$. Given the values of the level-two errors $\left(u_{ij}\right)$, the CD estimator of cluster-specific distribution function can be expressed as below

$$\hat{F}_{ij}\left(t\right)=N_{ij}^{-1}\left[n_{ij}F_{s_{ij}}\left(t\right)+\left(N_{ij}-n_{ij}\right)\hat{F}_{r_{ij}}\left(t\right)\right]=N_{ij}^{-1}\left[\sum_{ijk\in s_{ij}}I\left(y_{ijk}\le t\right)+\sum_{ijk\in r_{ij}}\hat{G}\left\{\frac{t-\mathbf{x}_{ijk}^{T}\hat{\boldsymbol{\beta}}_{(2)}-u_{ij}}{\hat{\sigma}_{\varepsilon(2)}}\right\}\right]. \quad (5.5)$$

The distribution function G can be estimated using smearing approach of Chambers and Dunstan (1986) as

$$\hat{G}\left(\vartheta\right)=n^{-1}\sum_{ijk\in s}I\left(\hat{\gamma}_{ijk}\le\vartheta\right)\text{, with }\hat{\gamma}_{ijk}=\frac{y_{ijk}-\mathbf{x}_{ijk}^{T}\hat{\boldsymbol{\beta}}_{(2)}^{gls}-\tilde{u}_{ij}}{\hat{\sigma}_{\varepsilon(2)}}=\frac{\tilde{\varepsilon}_{ijk}}{\hat{\sigma}_{\varepsilon(2)}}$$

where $\tilde{u}_{ij}=n_{ij}^{-1}\sum_{k=1}^{n_{ij}}\tilde{e}_{ijk}$ and $\tilde{\varepsilon}_{ijk}=\tilde{e}_{ijk}-\tilde{u}_{ij}$ with $\tilde{e}_{ijk}=y_{ijk}-\mathbf{x}_{ijk}^{T}\hat{\boldsymbol{\beta}}_{(2)}^{gls}$ are estimated using GLS estimator of the regression parameter $\boldsymbol{\beta}_{(2)}$, and $\hat{\sigma}_{\varepsilon(2)}^{2}$ is the MOM estimator of HH-level variance component. The REML estimators of $\boldsymbol{\beta}_{(2)}$, $\sigma_{u(2)}^{2}$, $\sigma_{\varepsilon(2)}^{2}$ as well as the corresponding level-specific residuals can be used in this smearing estimator. The smearing estimator $\hat{G}$ can be rewritten as

$$\begin{aligned}\hat{G}\left\{\frac{t-\mathbf{x}_{ijk}^{T}\hat{\boldsymbol{\beta}}_{(2)}^{gls}-\tilde{u}_{ij}}{\hat{\sigma}_{\varepsilon(2)}}\right\}&=n^{-1}\sum_{c=1}^{C_s}\sum_{h=1}^{n_c}I\left(\tilde{\gamma}_{ch}\le\frac{t-\mathbf{x}_{ijk}^{T}\hat{\boldsymbol{\beta}}_{(2)}^{gls}-\tilde{u}_{ij}}{\hat{\sigma}_{\varepsilon(2)}}\right)=n^{-1}\sum_{c=1}^{C_s}\sum_{h=1}^{n_c}I\left(\mathbf{x}_{ijk}^{T}\hat{\boldsymbol{\beta}}_{(2)}^{gls}+\tilde{u}_{ij}+\tilde{\gamma}_{ch}\hat{\sigma}_{\varepsilon(2)}\le t\right)\\&=n^{-1}\sum_{c=1}^{C_s}\sum_{h=1}^{n_c}I\left\{\tilde{y}_{ijk}^{(ch)}\le t\right\}\text{ with }\tilde{y}_{ijk}^{(ch)}=\mathbf{x}_{ijk}^{T}\hat{\boldsymbol{\beta}}_{(2)}^{gls}+\tilde{u}_{ij}+\tilde{\gamma}_{ch}\hat{\sigma}_{\varepsilon(2)}.\end{aligned}$$

Then the estimate of area-specific distribution function would be obtained as

$$\hat{F}_{i}^{CD}\left(t\right)=\sum_{j}^{C_i}\hat{F}_{ij}\left(t\right)=\sum_{j}^{C_i}N_{ij}^{-1}\left[\sum_{s_{ij}}I\left(y_{ijk}\le t\right)+\sum_{r_{ij}}\left\{n^{-1}\sum_{ch\in s}I\left(\tilde{y}_{ijk}^{(ch)}\le t\right)\right\}\right]. \quad (5.6)$$

This CD estimator of area-specific distribution function can be used if and only if all the clusters of an area are covered in the sample. Since area-specific census clusters are not covered in the sample, cluster random effects for the census clusters can be generated from either the parametric distribution or the empirical distribution of estimated sample cluster random effects via SRSWR. In reality, the number of clusters per area is not large enough to use only those clusters belong to that area. Moreover, under the assumption of area-homogeneity of ELL methodology, the estimated cluster random effects can be utilized in the resampling procedure. Therefore a NPB procedure is needed to estimate the ultimate area-specific distribution in the proposed CD approach.

The basic steps of the NPB are as follows. First, develop the cluster-specific model (5.4) to the survey data to obtain $\hat{\boldsymbol{\beta}}_{(2)}^{gls}$, $\hat{\mathbf{v}}\left(\hat{\boldsymbol{\beta}}_{(2)}^{gls}\right)$, $\hat{\sigma}_{u(2)}^2$, $\hat{\sigma}_{\varepsilon(2)}^2$, and the level-specific residuals $\tilde{u}_{ij}$ and $\tilde{\varepsilon}_{jk}$. Second, the cluster-level residuals $\left\{u_{ij}^*,\ ij \in U\right\}$ are generated for all the census clusters from the estimated sample residuals $\left\{\tilde{u}_{ij},\ ij \in s\right\}$ via resampling with SRSWR. The residuals at level-one can also be resampled but only for the sample individuals to utilize in the smearing approach. Third, the bootstrap realization $\hat{F}_{ij}^*(t)$ of cluster-specific distribution function is calculated for every cluster and also aggregated them to produce a bootstrap realization of the CD estimator for each small area as $\hat{F}_i^*(t) = N_i^{-1}\sum_{j}^{C_i} N_{ij}\hat{F}_{ij}^*(t)$. The steps are repeated for B times and then estimate the ultimate parameter with MSE as (5.3). Since this CD estimator is based on a bootstrap procedure and the smearing approach, the ultimate estimator is called here as bootstrap-based CD-type smearing estimator, hereafter CDSM. Thus the proposed estimators are defined and denoted as

$$\hat{F}_i^{CDSM} = B^{-1}\sum_{b=1}^{B} F_i^{*(b)} \text{ and } mse\left\{\hat{F}_i^{CDSM}\right\} = B^{-1}\sum_{b=1}^{B}\left\{F_i^{*(b)} - \hat{F}_i^{CDSM}\right\}^2 \qquad (5.7)$$

where the term “SM” stands for smearing approach. As NPB procedure in section 5.2.2, the estimated sample residuals are required to scaling to make the estimator free from the overestimation of the estimated variance components. The scaling terms might be respectively as $\widehat{\varepsilon}_{ijk} = \tilde{\varepsilon}_{ijk}\hat{\sigma}_{\varepsilon(2)}\left\{(n-1)^{-1}\sum_{ijk\in s}\tilde{e}_{ijk}^2\right\}^{-1/2}$ and $\widehat{u}_{ij} = \tilde{u}_{ij}\hat{\sigma}_{u(2)}\left\{(C_s-1)^{-1}\sum_{ij\in s}\tilde{u}_{ij}^2\right\}^{-1/2}$ as per **Appendix A.4**.

If area-specific model instead of cluster-specific model is assumed as working model, there will be no necessity to generate area-level random effects for the sampled areas and so distribution function can be estimated directly from analytic estimator like (5.6). For details please see Salvati *et al.* (2012). However, the analytic estimator is not possible to use for the non-sampled areas and so a synthetic estimator is then needed for the non-sampled small areas.

5.2.4 The CD Estimator: Monte Carlo Simulation Approach (CDMC)

Marchetti *et al.* (2012) proposed an alternative procedure to calculate the smearing-based estimator of small area parameters following a Monte Carlo (MC) simulation approach. The authors proposed their procedure for estimating small area distribution function using M-Quantile regression model. The main reason of such approach is to speed up the calculation procedure in CD method. When the sample size is reasonably large, the smearing-based CD estimator (5.7) will take much time if all the sample residuals are used in the smearing method. The basic steps of the MC simulation approach are parallel to the ELL semi-parametric procedure. In the simulation approach, response values for the non-sample units are generated using the fitted model parameters and the estimated residuals via $y_{ijk}^* = \mathbf{x}_{ijk}^T\hat{\boldsymbol{\beta}}_{(2)}^* + u_{ij}^* + \varepsilon_{ijk}^*$ where $u_{ij}^*,\ ij \in U$ and $\varepsilon_{ijk}^*,\ ijk \in U$ are drawn from the empirical distribution functions of the corresponding model residuals $\widehat{u}_{ij},\ ij \in s$ and $\widehat{\varepsilon}_{ijk}$,

$ijk \in s$ respectively. Then combining the sample and non-sample response values, the area-specific parameters are calculated and the process is repeated for B times to calculate the ultimate estimates as (5.7). The estimators of area-specific distribution function and their MSE are denoted by $\hat{F}_i^{CDMC}$ and $mse\left\{\hat{F}_i^{CDMC}\right\}$ respectively where the term "MC" stands for MC simulation approach. Since the proportion of sampled individuals out of total population is very negligible, response values are generated for all population units to conduct the MC simulation.

5.3 Estimation of Small Area Distributions and Poverty under Heteroskedasticity

The described ELL and CD approaches will produce bias estimates of small area distribution function if the assumption of HM nested errors is violated in reality, though the small area mean estimation will not be affected much. Table 5.1 reveals this situation considering a HM and a HT 2-level population models in a simulation study, which is detailed in section 5.4. The first population is considered with HM level-one error variance as $\sigma^2_{\varepsilon(2)} = \phi_1\left(x_{ijk}\right) = 94.05$, and the second one with HT level-one error variances as $\sigma^2_{\varepsilon(2).ijk} = \phi_3\left(x_{ijk}\right) = 90 + 0.5\, x^2_{ijk}$ where $x_{ijk} \sim \chi^2\left(20\right)$. Level-two errors are considered HM $\left(\sigma^2_{u(2)} = 23.05\right)$ for both the populations. The application of PELL estimator under homoskedasticity for the HT population produces higher relative bias (RB) and lower bootstrap coverage rate (BCR). However, the small area means of the HT population are not much affected in terms of RB and BCR.

Thus both the ELL and CD approaches under the assumption of HM nested errors must be extended when the level-one error is assumed HT. In such case, level-two variance component is required to estimate at first assuming level-one errors as HT, and then estimate the level-one error variances. A parametric regression approach has been proposed in the ELL methodology based on a logistic function specification of the squared level-one residuals $\left(\hat{\varepsilon}_{ijk}\right)$. This parametric method is highly reliant on the selection of the potential variables that can explain the heteroskedasticity. Two semi-parametric alternatives of heteroskedasticity modelling have been proposed in this section at first. Similar to the ELL methodology, we assume that the level-one error variances are HT and correspond to a smooth function of the auxiliary variables.

Table 5.1: Area averages of relative bias (RB, %), relative root mean squared error (RRMSE, %) and bootstrap coverage rate (BCR, %) of nominal 95% confidence intervals for homoskedastic PELL estimator of area-specific means and distribution functions (DFs) for populations with homoskedastic and heteroskedastic level-one errors

Working Model	Parameter	True Model					
		Homoskedastic			Heteroskedastic		
		RB	RRMSE	BCR	RB	RRMSE	BCR
Homoskedastic	$q_{0.10}$	0.429	23.857	94.69	14.844	24.834	83.52
	$q_{0.25}$	0.152	15.762	94.61	-0.937	12.734	91.69
	$q_{0.50}$	-0.030	9.344	94.52	-5.442	9.067	86.05
	$q_{0.75}$	-0.069	4.864	94.59	-2.903	4.679	88.34
	$q_{0.90}$	-0.053	2.286	94.70	0.120	1.732	95.70
	Mean	0.001	0.309	94.59	0.002	0.328	94.31

5.3.1 The ELL Estimator: Parametric Bootstrap (PELL)

A new superpopulation model similar to model (5.1) is considered here as

$$y_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)}^{(ht)} + u_{ij} + \varepsilon_{ijk};\ u_{ij} \sim N\left(0, \sigma_{u(2)}^{2(ht)}\right);\ \varepsilon_{ijk} \sim N\left(0, \sigma_{\varepsilon(2),ijk}^{2}\right) \qquad (5.8)$$

where the level-one errors (also known as idiosyncratic errors) are allowed to be HT, while the level-two errors are assumed homoscedastic, independent and identically distributed from the idiosyncratic errors. In such case we have $\mathrm{Var}\left(y_{ijk}\right) = \sigma_{u(2)}^{2(ht)} + \sigma_{\varepsilon(2),ijk}^{2}$ and $\mathrm{cov}\left(y_{ijk}, y_{i'j'k'}\right) = \sigma_{u(2)}^{2(ht)}$ only if $ij = i'j'$ under the model (5.8).

Initially the regression parameters and the residuals at levels one and two are estimated using the LS method as in section 2. The level-two variance component can be estimated via MOM approach assuming HT level-one errors as below:

$$\hat{\sigma}_{u(2)}^{2(ht)} = \max\left\{ \left(\sum_{ij\in s} w_{ij}\left(1-w_{ij}\right) \right)^{-1} \left\{ \left[\sum_{ij\in s} w_{ij}\left(\hat{\bar{e}}_{ij.} - \hat{\bar{e}}_{...}\right)^2 \right] - \sum_{ij\in s} w_{ij}\left(1-w_{ij}\right)\hat{\tau}_{ij}^2 \right\}, 0 \right\} \tag{5.9}$$

where $w_{ij} = n_{ij}/n$, $\hat{\bar{e}}_{ij.} = n_{ij}^{-1}\sum_{k}^{n_{ij}} \hat{e}_{ijk}$, $\hat{\bar{e}}_{...} = n^{-1}\sum_{ijk\in s} \hat{e}_{ijk}$, $\hat{\tau}_{ij}^2 = \dfrac{1}{n_{ij}\left(n_{ij}-1\right)}\sum_{k}^{n_{ij}}\left(\hat{\varepsilon}_{ijk} - \hat{\bar{\varepsilon}}_{ij.}\right)^2$, and

$\hat{\bar{\varepsilon}}_{ij.} = n_{ij}^{-1}\sum_{k}^{n_{ij}} \hat{\varepsilon}_{ijk}$. Detail derivation of the estimator has been shown in **Appendix A.6**.

Elbers *et al.* (2002) proposed the same estimator with $w_{ij} = \sum_{k}^{n_{ij}} w_{ijk} \Big/ \sum_{ijk\in s} w_{ijk}$ where w_{ij} is the by-cluster transformed sampling weights which sum to one across clusters and w_{ijk} is the rescaled sampling weights which sum to the total sample size. The main problem of such estimators is the possibility of getting negative estimates of variance components. In case of negative value, $\sigma_{u(2)}^{2(ht)}$ can be estimated assuming HM HH-specific random errors. Interestingly, $\hat{\sigma}_{u(2)}^{2(ht)}$ in (5.9) will be exactly same as $\hat{\sigma}_{u(2)}^2$ in (5.2) when the clusters are equal sizes (proof is shown in the **Appendix A.6**).

Now the interest is estimation of HH-level residual variances. Two ways are suggested by Elbers *et al.* (2002): (1) direct estimation method assuming a HM level-one random errors and (2) logistic-type model-based computation assuming HT level-one random errors. The direct HH-level variance component $\hat{\sigma}_{\varepsilon(2)}^{2(ht)}$ can be easily obtained by taking the difference between $\hat{\sigma}_{u(2)}^{2(ht)}$ and overall mean squared error (MSE) $\hat{\sigma}_e^2$ obtained from the initial fitted model using LS method. The direct estimation method can be applied for negligible heteroskedasticity but would be unreasonable for wide heteroskedasticity. The decision can be made from the basic literature on checking heteroskedasticity via graphical and statistical diagnostics (Gujarati, 2003).

In ELL methodology, the HT residual variances are calculated based on the parametric logistic function as

$$\sigma^2_{\varepsilon(2),ijk} = \phi\left(\mathbf{x}_{ijk}, \boldsymbol{\alpha}, A, B\right) = \left[\frac{A\exp\left(\mathbf{z}^T_{ijk}\boldsymbol{\alpha}\right)+B}{1+\exp\left(\mathbf{z}^T_{ijk}\boldsymbol{\alpha}\right)}\right] \tag{5.10}$$

where $z_{ijk} = g\left(x_{ijk}\right)$, A and B are upper and lower value of $\sigma^2_{\varepsilon(2),ijk}$. This parametric form avoids both negative and extremely high values of $\sigma^2_{\varepsilon(2),ijk}$. The parameters A, B, and $\boldsymbol{\alpha}$ can be estimated following an approximate/pseudo maximum likelihood procedure. The upper and lower bound are recommended as $\hat{A} = 1.05 \times \text{maximum}\left\{\hat{\varepsilon}^2_{ijk}\right\}$ and $\hat{B} = 0$ respectively. These values of A and B help to develop a simpler logistic type link function of squared HH-level residuals with explanatory variables as:

$$\ln\left(\frac{\hat{\varepsilon}^2_{ijk}}{\hat{A}-\hat{\varepsilon}^2_{ijk}}\right) = \mathbf{z}^T_{ijk}\hat{\boldsymbol{\alpha}} + r_{ijk} \tag{5.11}$$

where r_{ijk} is a random error term. In the poverty mapping work, a slight modification may be required by adding a constant term (δ) to residual squares $\left(\hat{\varepsilon}^2_{ijk}\right)$ in model (5.11) (Haslett, *et al.*, 2010). Also trimming and winsorizing techniques may be needed for getting better estimates of the alpha parameters. Applying delta method (Oehlert, 1992), the estimator of $\sigma^2_{\varepsilon(2),ijk}$ can be expressed as (Elbers *et al.*, 2002) as:

$$\hat{\sigma}^{2.ELL}_{\varepsilon(2),ijk} \approx \left[\frac{\hat{A}D_{ijk}}{1+D_{ijk}}\right] + \frac{1}{2}\hat{v}\left(r\right)\left[\frac{\hat{A}D_{ijk}(1-D_{ijk})}{\left(1+D_{ijk}\right)^3}\right] \tag{5.12}$$

where $D_{ijk} = \exp(\mathbf{z}_{ijk}^T \hat{\boldsymbol{\alpha}})$ and $\hat{v}(r)$ is the estimated MSE obtained from (5.11). Then the estimated variance components $\hat{\sigma}_{u(2)}^{2(ht)}$ and $\hat{\sigma}_{\varepsilon(2),ijk}^{2.ELL}$ are used to obtain GLS estimates of regression parameters and their variance-covariance matrix as

$$\hat{\boldsymbol{\beta}}_{(2)}^{gls(ht)} = \left(\mathbf{X}^T \hat{\mathbf{V}}_s^{-1} \mathbf{X}\right)^{-1} \left(\mathbf{X}^T \hat{\mathbf{V}}_s^{-1} \mathbf{y}\right) \text{ and } \hat{\mathbf{v}}\left(\hat{\boldsymbol{\beta}}_{(2)}^{gls(ht)}\right) = \left(\mathbf{X}^T \hat{\mathbf{V}}_s^{-1} \mathbf{X}\right)^{-1} = \left(\sum_c \mathbf{X}_c^T \hat{\mathbf{V}}_c^{-1} \mathbf{X}_c\right)^{-1}$$

where $\hat{\mathbf{V}}_s = diag\left\{\hat{\mathbf{V}}_c = \left(\hat{\sigma}_{\varepsilon(2),c_1}^2, ..., \hat{\sigma}_{\varepsilon(2),c_{n_C}}^2\right)\mathbf{I}_{n_c} + \hat{\sigma}_{u(2)}^{2(ht)} \mathbf{1}_{n_c} \mathbf{1}_{n_c}^T;\ c = 1, ..., C_s\right\}$. The ELL method uses a simulation procedure to regenerate the conditional distribution of y_{ijk} by adding simulated values of cluster-level (u_{ij}) and HH-level (ε_{ijk}) errors to fitted values $\left(\mathbf{x}_{ijk}^T \hat{\boldsymbol{\beta}}_{(2)}^{gls(ht)}\right)$. To conduct the simulation process, $\sigma_{u(2)}^{2(ht)}$ is needed to estimate with its estimated variance $\hat{v}\left(\hat{\sigma}_{u(2)}^{2(ht)}\right)$ for generating the location-specific errors (u_{ij}). Both analytic and bootstrap variance estimators of $\hat{\sigma}_{u(2)}^{2(ht)}$ are given in Elbers *et al.* (2002). For estimating small area distribution functions and their standard errors, the ultimate inputs used in simulation are: point estimates of alpha and beta regression parameters $\left(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{(2)}^{gls(ht)}\right)$ with their estimated covariance matrices $\left(\hat{\mathbf{v}}(\hat{\boldsymbol{\alpha}}), \hat{\mathbf{v}}\left(_{(2)}^{gls(ht)}\right)\right)$, empirical and or parametric distribution of the level-specific errors. The PB procedure is conducted as the following steps.

Step 1: A set of $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}_{(2)}^{*(ht)}$ parameters are generated from multivariate normal distributions with mean vectors $\left(\hat{\boldsymbol{\alpha}},\ \hat{\boldsymbol{\beta}}_{(2)}^{gls(ht)}\right)$ and variance-covariance matrices $\left\{\hat{\mathbf{v}}(\hat{\boldsymbol{\alpha}}),\ \hat{\mathbf{v}}\left(\hat{\boldsymbol{\beta}}_{(2)}^{gls(ht)}\right)\right\}$.

Step 2: The cluster-level variance component $\left(\sigma_{u(2)}^{2(ht)}\right)$ can be remained fixed or generated from a parametric distribution such as gamma distribution with mean $\hat{\sigma}_{u(2)}^{2(ht)}$ and variance $\hat{v}\left(\hat{\sigma}_{u(2)}^{2(ht)}\right)$. The HH-level error variances $\left(\sigma_{\varepsilon(2),ijk}^{2}\right)$ are estimated from (5.12) using the generated $\boldsymbol{\alpha}$ values in Step 1. In this thesis, $\hat{\sigma}_{u(2)}^{2(ht)}$ is considered fixed.

Step 3: Level-specific errors $\left(u_{ij}^{*},\ \varepsilon_{ijk}^{*}\right)$ are then generated from their respective assumed parametric distribution using $\hat{\sigma}_{u(2)}^{2(ht)}$ and $\hat{\sigma}_{\varepsilon(2),ijk}^{2.ELL}$ respectively. Non-normality can be assumed for both errors such as student's *t*-distribution with degrees of freedom that can approximate the distribution of sample residuals estimated in the first stage.

Step 4: Generate the response values using the predicted values $\mathbf{x}_{ijk}^{T}\boldsymbol{\beta}_{(2)}^{(ht)*}$ and the generated errors via $\hat{y}_{ijk}^{*}=\mathbf{x}_{ijk}^{T}\boldsymbol{\beta}_{(2)}^{(ht)*}+u_{ij}^{*}+\varepsilon_{ijk}^{*}$ and calculate the bootstrap parameter F_{i}^{*} at t for each target sub-population.

Step 5: Repeat the steps for B times to calculate the ultimate estimates and their MSEs which are defined as (5.3) and denoted by respectively $\hat{F}_{i}^{PELL}$ and $mse\left\{\hat{F}_{i}^{PELL}\right\}$.

5.3.2 The ELL Estimator: Semi-parametric Bootstrap (SPELL)

A semi-parametric bootstrap procedure can be easily conducted using the estimated parameters obtained in the first phase of ELL methodology (Elbers *et al.*, 2002). The basic differences are in the **second** and **third** steps of PB procedure. In **second** step, instead of generating cluster-level errors $u_{ij}^{*},\ ij\in U$ from the sampling distribution, they are randomly drawn for all census clusters from the sample residuals $\hat{u}_{ij},\ ij\in s$ with replacement. In **third** step, the estimated sample residuals $\hat{\varepsilon}_{ijk},\ ijk\in s$ are standardized as $r_{ijk} = \hat{\varepsilon}_{ijk} \big/ \hat{\sigma}^{ELL}_{\varepsilon(2),ijk} - n^{-1} \sum_{ijk \in s} \hat{\varepsilon}_{ijk} \big/ \hat{\sigma}^{ELL}_{\varepsilon(2),ijk}$ from where the standardized errors r^*_{ijk}, $ijk \in U$ are randomly drawn for all census HHs via SRSWR. The standardized errors r^*_{ijk} are then rescaled as $\varepsilon^*_{ijk} = r^*_{ijk} \sigma^*_{\varepsilon(2),ijk}$ where $\sigma^{2*}_{\varepsilon(2),ijk}$ are calculated by putting the generated $\boldsymbol{\alpha}^*$ in (5.12). A conditional approach can also be followed to resample the level-one errors by drawing r^*_{ijk} for unit k only from the standardized residuals r_{ijk} s that correspond to the cluster ij from which u^*_{ij} is obtained for the unit k in step 2. Though ε_{ijk} and u_{ij} are uncorrelated, the conditional approach may capture any non-linear relationship between them (Elbers *et al.*, 2002). Using the generated level-specific errors, the bootstrap response values are simulated via $y^*_{ijk} = \mathbf{x}^T_{ijk} \boldsymbol{\beta}^*_{(2)} + u^*_{ij} + \varepsilon^*_{ijk}$ and then the target parameters of interest F^*_i are calculated by aggregating the simulated response values. The ELL estimators of area-specific distribution functions with their MSE based on the semi-parametric bootstrap procedure are denoted by $\hat{F}^{SPELL}_i$ and $mse\left\{\hat{F}^{SPELL}_i\right\}$ respectively. The LS raw residuals are recommended to scale in the NPB bootstrap procedure when level-one errors are assumed HM, however under the assumption of level-one HT errors no clear idea is given in the ELL method.

5.3.3 Non-parametric Estimation of Heteroskedastic Error Variances

The most difficult part in the ELL method is the development of heteroskedasticity (alpha) model. The estimator of the HT error variance is based on the delta method which is based on second order Taylor approximation (Oehlert, 1992) and consequently it can approximate any monotone heteroskedasticity pattern. The logistic form of the heteroskedasticity function is structured in such a way that it provides estimates within a fixed range (B, A) based on the sample data set. As a result unavailability or missing any potential explanatory variable (and hence misspecification of heteroskedasticity model) may cause understating or overstating the true variation if the sample observations are imbalanced to heteroskedasticity. An extensive exploration is needed to find out the potential explanatory variables and their transformation and then fitting an alpha model with satisfactory r-squared value is critical. As an alternative of the parametric heteroskedasticity modelling, semi-parametric modelling can be followed that does not require any recognized form of heteroskedasticity function. Since the heteroskedasticity function is assumed to be a smooth function of the explanatory variables, residual variances can be estimated for all the population units based on a non-parametric regression of sample residual squares on the sample *x*-values or their combination (say, predicted values). The main task will be finding a suitable non-parametric regression method from the sample data set. A stratified MOM approach is proposed here assuming heteroskedasticity varies with regression mean $\left(\hat{y}_{ijk}\right)$. The HH-level error variances are estimated by stratifying the sample dataset based on the marginal distribution of the predicted values obtained from the GLS regression line. A non-parametric regression approach and the stratified MOM are described in the following two sub-sections.

5.3.3.1 Non-parametric Regression Approach

Lombardia *et al.* (2005) proposed a non-parametric method to estimate error variances under a general linear model where population units are assumed uncorrelated. The method can be extended to a multilevel model to estimate the level-one error variances. In a 2-level model, the level-two errors u_{ij} are assumed independently distributed from the level-one errors ε_{ijk} and so a non-parametric variance estimator like

Nadaraya-Watson type estimator of Lombardia *et al.* (2005) can be used to estimate the level-one error variances $\left(\sigma^2_{\varepsilon(2),ijk}\right)$ as

$$\hat{\sigma}^2_{\varepsilon(2),ijk}\left(\mathbf{x}\right)=\frac{\sum_{ijk\in s}k\left\{\left(\mathbf{x}-\mathbf{x}_{ijk}\right)/b\right\}\left(\ln y_{ijk}-\mathbf{x}^T_{ijk}\hat{\beta}_{(2)}-\hat{u}_{ij}\right)^2}{\sum_{ijk\in s}k\left\{\left(\mathbf{x}-\mathbf{x}_{ijk}\right)/b\right\}}=\frac{\sum_{ijk\in s}k\left\{\left(\mathbf{x}-\mathbf{x}_{ijk}\right)/b\right\}\left(\hat{\varepsilon}^2_{ijk}\right)}{\sum_{ijk\in s}k\left\{\left(\mathbf{x}-\mathbf{x}_{ijk}\right)/b\right\}}$$

where k is a symmetric density function (say, Kernel function), b is the bandwidth parameter, $\mathbf{x}$ is a set of values of the explanatory variables. The Kernel regression is the basic and simplest non-parametric regression method, however suffers from boundary problem (Chu and Marron, 1991). At the lower and upper values of explanatory variables (wide windows), Kernel smoothing provides higher biased estimates. However, if the *x*-values are uniformly distributed (rare case in reality), there is less possibility of boundary problem in kernel regression smoothing method. The problem can be solved by choosing variable bandwidth where higher bandwidth is considered at the boundary points to estimate the parameter using more observations.

The K-nearest neighbor (KNN) regression method (Altman, 1992) can also be applied to estimate the error variances based on the sample residuals $\hat{\varepsilon}_{ijk}$ whose *x*-values are near to the *x*-values of a population HH. Thus we have to find out the conditional distribution of $\hat{\varepsilon}_{ijk}$ for a population unit with $x_{ijk}, ijk\in U$. Suppose $\mathbf{s}\left(x_{ijk}\right)$ denotes the set of sample *x*-values whose values are nearest neighbor of population $x_{ijk}, ijk\in U$, $n\left(x_{ijk}\right)$ denotes the size of the vector, and $\mathbf{w}\left(x_{ijk}\right)$ is the set of weights based on the distances of the sample *x*-values from x_{ijk}. Then the KNN estimator of $\sigma^2_{\varepsilon(2),ijk}$ for the population unit with x_{ijk} will be

$$\hat{\sigma}^2_{\varepsilon(2),ijk}(x) = n^{-1}(x_{ijk}) \sum_{ijk \in \mathbf{s}(x_{ijk})} \hat{\varepsilon}^2_{ijk} \quad \text{or} \quad \hat{\sigma}^2_{\varepsilon(2),ijk}(x) = n^{-1}(x_{ijk}) \sum_{ijk \in \mathbf{s}(x_{ijk})} \mathbf{w}(x_{ijk}) \hat{\varepsilon}^2_{ijk} \,.$$

The KNN regression approach is also suffered from the boundary effects (Rahvar and Ardakani, 2011). Local polynomial regression, hereinafter LPR, has good performance on the boundary (Hastie and Loader, 1993) and is superior to all other linear smoothers in a minimax sense (Cheng *et al.*, 1997; Avery, 2013). The LPR estimates depends on choice of weighting function (K), the size of neighborhood (b), and the order of polynomial (p) which is the order of Taylor's approximation. The LPR estimator of the error variance function can be written as

$$\hat{\sigma}^{2.LPR}_{\varepsilon(2),ijk}(\mathbf{x}) = \sum_{ijk \in s} K_b(\mathbf{x} - \mathbf{x}_{ijk}) \left(\ln y_{ijk} - \mathbf{x}^t_{ijk} \hat{\boldsymbol{\beta}}_{(2)} - \hat{u}_{ij} \right)^2 = \sum_{ijk \in s} K_b(\mathbf{x} - \mathbf{x}_{ijk}) \hat{\varepsilon}^2_{ijk} \tag{5.13}$$

where b controls the size of the neighborhood around $\mathbf{x}$, $K_b(.)$ controls the weights, where $K_b(.) \equiv b^{-1} K\left(\frac{.}{b}\right)$, and K is the kernel function. A special care is needed for extrapolation in LPR method particularly when the non-sample *x*-values are far from the sample *x*-values. In combining the survey data set with the census data set, the first assumption is that the distribution of survey explanatory variables will be similar as that of the census explanatory variables in terms of definition or measurement (Tarozzi and Deaton, 2009). Thus it can be assumed that the non-sample *x*-values can be distributed as the sample *x*-values. If frequency distribution of x is uniform in the sample data set, non-parametric method will provide the better estimates of error variances. In this study, local linear regression (LPR with first order polynomial) is used to examine the performance of $\hat{\sigma}^{2.LPR}_{\varepsilon(2),ijk}$ estimator.

5.3.3.2 Stratified Method of Moments (STR) Approach

We propose a stratification-based method of moments (MOM) approach to estimate the level-one error variances based on the conditional distribution of the level-one squared residuals of sample observations given the marginal distribution of predicted values. The scatterplot of squared residuals $\left(\hat{\varepsilon}_{ijk}^{2}\right)$ against the predicted values $\left(\mathbf{x}_{ijk}^{T}\hat{\boldsymbol{\beta}}_{(2)}\right)$ is expected to show the existence of heteroskedasticity. If heteroskedasticity exists, the scatterplot should have some specific pattern. On the basis of shape, the whole sample can be split into H strata so that the distribution of squared residuals given the predicted values looks HM within a stratum. In such situation, mean of the squared residuals $n_{h}^{-1}\sum_{ijk\in h}\hat{\varepsilon}_{ijk}^{2}$ belong to a stratum h $\left(h=1,\ldots,H\right)$ will be considered as the error variance $\hat{\sigma}_{\varepsilon(2),ijk}^{2(h)}$ for the h^{th} stratum. To make the estimates stable, the residual variances will be calculated through an iterative GLS (IGLS) method where the initial estimates of cluster variance component $\hat{\sigma}_{u(2)}^{2(ht)}$ and HH error variances are used as below.

Suppose the sample of size n is divided into H strata of size n_{h} based on the quantiles of the sample predicted values $\mathbf{x}_{ijk}^{T}\hat{\boldsymbol{\beta}}_{(2)}$, $ijk\in s$ and then the level-one error variance for the h^{th} stratum is calculated as $\hat{\sigma}_{\varepsilon(2),ijk}^{2(h)}=n_{h}^{-1}\sum_{ijk\in h}\hat{\varepsilon}_{ijk}^{2}$. This estimate is considered as the level-one error variance for all individuals belong to that stratum. Then the estimated cluster variance component $\hat{\sigma}_{u(2)}^{2(ht)}$ and the sample residual variances $\hat{\sigma}_{\varepsilon(2),ijk}^{2(h)}$ are employed to attain the GLS estimates of regression parameters, $\hat{\boldsymbol{\beta}}_{(2)}^{gls*}$. The corresponding new predicted values and HH-level conditional residuals are used again to estimate the new error variances by creating new H strata. The procedure will be iterated until the convergence of the strata error variances. After obtaining the ultimate strata with their corresponding error variances, population units are allocated to the strata according to their predicted values calculated as $\mathbf{x}_{ijk}^{T}\hat{\boldsymbol{\beta}}_{(2)}^{gls}$, $ijk \in U$ where $\hat{\boldsymbol{\beta}}_{(2)}^{gls}$ will be the GLS estimates at the final iteration. The HH-level error variances for all population units denoted as $\hat{\sigma}_{\varepsilon(2),ijk}^{2(STR)}$, $ijk \in U$ will be their corresponding estimated stratum error variance.

The advantage of this procedure is its simplicity and no parameter is required to capturing heteroskedasticity. It is expected that the procedure will not be affected due to model misspecification and outliers. A proper method of stratification can be employed to determine the number of strata. A proper scrutiny of the scatterplot may help to take decision on stratification where finer stratification is required. If it is observed that the conditional distribution of squared residuals given predicted values are skewed, a finer stratification can be considered in the non-skewed region where the most observations are. If proper stratification is made, the method might provide automatically ELL-type estimates which are based on the delta method.

5.3.4 The CD Estimator: Smearing Approach (CDSM)

The model (5.8) with HT level-one random errors can be rewritten as

$$y_{ijk} = \mathbf{x}_{ijk}^{T}\boldsymbol{\beta}_{(2)}^{(ht)} + u_{ij} + \sigma_{\varepsilon(2),ijk}\left(\mathbf{x}_{ijk}\right)\gamma_{ijk} \;;\; \gamma_{ijk} \sim N\left(0,1\right) \qquad (5.14)$$

where γ_{ijk} is the normalized error. The main difference from the procedure described in the section 5.2.3 is the estimation of level-one error variances $\sigma_{\varepsilon(2),ijk}^{2}$ and their application in the CD approach. With the HT error variances, the estimator of the distribution function G will be

$$\hat{G}\left(\vartheta\right) = n^{-1}\sum_{ijk\in s} I\left(\hat{\gamma}_{ijk} \le \vartheta\right), \text{ with } \hat{\gamma}_{ijk} = \frac{y_{ijk} - \mathbf{x}_{ijk}^{T}\hat{\boldsymbol{\beta}}_{(2)}^{gls(ht)} - \tilde{u}_{ij}}{\hat{\sigma}_{\varepsilon(2),ijk}\left(\mathbf{x}_{ijk}\right)} = \frac{\tilde{\varepsilon}_{ijk}}{\hat{\sigma}_{\varepsilon(2),ijk}\left(\mathbf{x}_{ijk}\right)}.$$

And the final CD estimator of cluster-specific distribution function is then

$$\hat{F}_{ij}(t) = N_{ij}^{-1}\left[\sum_{ijk\in s_{ij}} I\left(y_{jk} \le t\right) + \sum_{ijk\in r_{ij}} n^{-1}\sum_{c=1}^{C_s}\sum_{h=1}^{n_c} I\left\{\tilde{y}_{ijk}^{(ch)} \le t\right\}\right] \qquad (5.15)$$

with $\tilde{y}_{ijk}^{(ch)} = \mathbf{x}_{ijk}^{T}\hat{\boldsymbol{\beta}}_{(2)}^{gls(ht)} + \tilde{u}_{ij} + \tilde{\gamma}_{ch}\hat{\sigma}_{\varepsilon(2),ijk}\left(x_{ijk}\right)$. And hence the estimator of area-specific distribution function will be

$$\hat{F}_{i}^{CD}(t) = \sum_{j}^{C_i}\hat{F}_{ij}(t) = \sum_{j}^{C_i} N_{ij}^{-1}\left[\sum_{s_{ij}} I\left(y_{ijk} \le t\right) + \sum_{r_{ij}}\left\{n^{-1}\sum_{c=1}^{C_s}\sum_{h=1}^{n_c} I\left(\tilde{y}_{ijk}^{(ch)} \le t\right)\right\}\right] \qquad (5.16)$$

As per section 5.2.3 a NPB procedure is required to conduct for estimating the ultimate area-specific distribution. The steps can be summarized as follows: **Step 1**: Fit the cluster-specific model (5.14) to the survey data and then obtain $\hat{\boldsymbol{\beta}}_{(2)}^{gls(ht)}$, $\hat{\mathbf{v}}\left(\hat{\boldsymbol{\beta}}_{(2)}^{gls(ht)}\right)$, $\hat{\sigma}_{u(2)}^{2(ht)}$, and then HH-level error variances $\hat{\sigma}_{\varepsilon(2),ijk}^{2}$ for both sample and non-sample HHs using any non-parametric method; **Step 2**: calculate the level-specific residuals as $\tilde{u}_{ij} = n_{ij}^{-1}\sum_{k=1}^{n_j}\tilde{e}_{jk}$ and $\tilde{\varepsilon}_{jk} = \tilde{e}_{jk} - \tilde{u}_{ij}$ where $\tilde{e}_{jk} = y_{jk} - \mathbf{x}_{jk}^{T}\hat{\boldsymbol{\beta}}_{(2)}^{gls(ht)}$ and then obtain the standardized level-one residuals as $\tilde{\gamma}_{ijk} = \tilde{\varepsilon}_{ijk}/\hat{\sigma}_{\varepsilon(2),ijk}$; **Step 3**: randomly draw u_{ij}^{*} from $\tilde{u}_{ij}$, $ij \in s$ via SRSWR, calculate $\hat{F}_{ij}^{*}$ using the estimator (5.15) and then aggregated $\hat{F}_{ij}^{*}$ to the level of interest as (5.16). Step 3 is repeated for B times to obtain the ultimate parameters with their EMSEs denoted as $\hat{F}_{i}^{CDSM}$ and $mse\left\{\hat{F}_{i}^{CDSM}\right\}$ respectively as (5.7). As in section 5.2.3, the estimated sample residuals are also required to scaling to free from the overestimation of the estimated variance components. There is no question about scaling the cluster-level residuals as $\hat{u}_{ij} = \tilde{u}_{ij}\hat{\sigma}_{u(2)}\left\{\left(C_s - 1\right)^{-1}\sum_{ij}\tilde{u}_{ij}^{2}\right\}^{-1/2}$, however scaling the HH-level residuals in such situation is an important issue. Since there is no specific variance component for the HT level-one errors, a constant variance should be used so that the overall variation due to scaling remains same in the bootstrap procedure. A constant variance can estimated by $\hat{\sigma}^{2(ht)}_{\varepsilon(2)} = \hat{\sigma}^2_{e(2)} - \hat{\sigma}^{2(ht)}_{u(2)}$ (similar to the direct estimation approach of ELL method) where $\hat{\sigma}^2_{e(2)}$ is the overall MSE estimated via LS method at the initial stage.

5.3.5 The CD Estimator: Monte Carlo Simulation Approach (CDMC)

The MC simulation based CD approach (CDMC) is similar to the semi-parametric approach of ELL method. The basic difference is the estimation of level-one residual variances and their application in the bootstrap procedure. In ELL method, level-one residual variances are generated via the generated $\boldsymbol{\alpha}^*$ parameters in each bootstrap, while in CDMC approach the error variances are remained fixed over the bootstraps. In the CDMC approach, the level specific errors u^*_{ij} and γ^*_{ij} are generated for all census units from the corresponding GLS sample residuals $\left\{\tilde{u}_{ij},\ ij \in s\right\}$ and $\left\{\tilde{\gamma}_{ijk},\ ijk \in s\right\}$ via SRSWR. The level-one normalized errors γ^*_{ij} are multiplied by their corresponding estimated variances $\hat{\sigma}^{2.STR}_{\varepsilon(2),ijk}$, $ijk \in U$ to obtain the ultimate generated errors as $\varepsilon^*_{ijk} = \gamma^*_{ijk}\hat{\sigma}^{STR}_{\varepsilon(2),ijk}$. The generated bootstrap population values are then aggregated to obtain the area-specific target parameters. This MC simulation is repeated for B times to calculate the target parameter of interest with the MSE denoted as $\hat{F}^{CDMC}_i$ and $mse\left\{\hat{F}^{CDMC}_i\right\}$ respectively. The cluster- and HH-level residuals are also required to scale by the corresponding variance components discussed in section 5.3.4. The CD-type estimators with scaled residuals are denoted as “SCDMC” and “SCDSM”.

5.4 Numerical Evaluations

A number of model-based simulation studies have been conducted to evaluate the performances of the ELL and CD-type estimators for small area distribution functions and FGT poverty indicators. The area-specific distribution functions are calculated at 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} quantiles, while the poverty indicators are estimated at the lower two quantiles of income distribution as poverty lines. The performances of the estimators are compared by calculating relative bias (RB), relative root mean squared error (RRMSE), and bootstrap coverage rate (BCR) of nominal 95% confidence intervals. The performance measures are defined as follows:

$$RB_{i(q)} = \left(R^{-1} \sum_{r=1}^{R} F_{i(q)}^{(r)} \right)^{-1} \left\{ R^{-1} \sum_{r=1}^{R} \left(\hat{F}_{i(q)}^{(r)} - F_{i(q)}^{(r)} \right) \right\},$$

$$RMSE_{i(q)} = \left(R^{-1} \sum_{r=1}^{R} F_{i(q)}^{(r)} \right)^{-1} \sqrt{R^{-1} \sum_{r=1}^{R} \left(\hat{F}_{i(q)}^{(r)} - F_{i(q)}^{(r)} \right)^{2}}, \text{ and}$$

$$BCR_{i(q)} = \frac{1}{R} \sum_{r=1}^{R} I\left(\hat{F}_{i(q),0.025}^{(r)} \le F_{i(q)}^{(r)} \le \hat{F}_{i(q),0.975}^{(r)} \right).$$

Here R denotes the number of simulations, $F_{i(q)}^{(r)}$ denotes the true value of the distribution function at q^{th} quantile of $\mathbf{y}$ for area i in simulation r, $\hat{F}_{i(q)}^{(r)}$ denotes the corresponding estimate, $\hat{F}_{i(q),0.025}^{(r)}$ and $\hat{F}_{i(q),0.975}^{(r)}$ indicate the lower and upper limit of 95% bootstrap confidence interval of $F_{i(q)}^{(r)}$.

In the model-based simulations, two types of models are considered to generate the population values of Y. The first simulation referred as Type-I is based on a linear 2-level model and the second one referred as Type-II is based on a log-normal 2-level model. In both type of simulations, the finite population is assumed to have D=50 small domains each with C_i =10 clusters and number of HHs per cluster N_{ij} randomly drawn from a uniform distribution on [100,120] which lead to area sizes N_i within the range [1000,1200] HHs.

In **Type-I** simulations, the 2-level linear model is $y_{ijk} = 500 + 1.5x_{ijk} + u_{ij} + \varepsilon_{ijk}$ where $x_{ijk} \sim \chi^2(20)$, $u_{ij} \sim N(0, 23.5)$, $\varepsilon_{ijk} \sim N\left(0, \sigma^2_{\varepsilon(2),ijk}\right)$, $i = 1,...,D$, $j = 1,...,C_i$ and $k = 1,...,N_{ij}$. Five different scenarios have been considered based on the distribution of HH-level errors (ε_{ijk}) under Type-I simulations. In the **first** scenario, ε_{ijk} are assumed to have homoscedastic variance as $\sigma^2_{\varepsilon(2)} = \phi_1(x_{ijk}) = 94.05$ which corresponds to an intra-cluster correlation of $23.5/(23.5 + 94.05) = 0.20$. In **second** scenario the error variances are assumed to have mild heteroskedasticity as $\phi_2(x_{ijk}) = 90 + \left(0.75\, e^{0.25 * x_{ijk}}\right)^2$, while in **third** scenario strong heteroskedasticity as $\phi_3(x_{ijk}) = 90 + 0.5\, x^2_{ijk}$. In **fourth** and **fifth** scenarios, model misspecification situations are created by ignoring an explanatory variable in the sampling procedure that characterizes the heteroskedasticity of HHs in each cluster. It is assumed here that each cluster consists of two types of households (say literate and illiterate HH head; rich and poor; Muslim and Non-Muslim; slum and non-slum dwellings) with different error variances. A new heteroskedasticity function is then assumed as $\phi_4(x_{ijk}, z_{ijk}) = \phi_2(x_{ijk}) I(z_{ijk} = 0) + \phi_3(x_{ijk}) I(z_{ijk} = 1)$ where the proportion of HHs with $z_{ijk} = 0$ in a cluster is assumed to vary uniformly on [20%, 40%]. In **fourth** scenario the HHs are randomly selected without considering the grouping variable, while in **fifth** scenario the HHs are selected randomly only from the large group with $z_{ijk} = 1$ and hence the sample becomes imbalanced. In both the cases, the true populations are misspecified due to ignorance of the grouping variable in fitting the regression model.

In **Type-II** simulations, the response values are generated from a log-normal model as $\log\left(y_{ijk}\right) = 6 + 0.5x_{1ijk} - 0.55x_{2ijk} + u_{ij} + \varepsilon_{ijk}$ to see the performances of ELL and CD-type estimators for the FGT poverty indicators at the first two quantiles of the overall distribution of $\exp\left(y_{ijk}\right)$ as the poverty lines. Here two explanatory variables are assumed to follow a bivariate normal distribution with mean vector $\mu = \left(0.50 \;\; 0.75\right)^{/}$ and variance-covariance matrix $\Sigma = \left[\left(1.50 \;\; 0.10\right)^{/} \quad \left(0.10 \;\; 0.95\right)^{/}\right]$. Similar to Type-I simulation, cluster-level random errors are assumed HM with $\sigma^2_{u(2)} = 0.05$ and different HT functions are assumed for generating HH-level errors to create five different scenarios. The HT functions are $\phi_1\left(x_{ijk}\right) = 0.20$ in **first** scenario, $\phi_2\left(x_{ijk}\right) = 0.19 + 0.005\exp\left\{0.05g(x_{ijk})^2\right\}$ a monotone increasing function with $g\left(x_{ijk}\right) = 6 + 0.5x_{1ijk} - 0.55x_{2ijk}$ in **second** scenario**, and** $\phi_3\left(x_{ijk}\right) = 0.20 + 0.015std\left\{g\left(x_{ijk}\right)\right\}^2$ a non-monotone function of $g\left(x_{ijk}\right)$ in the **third** scenario where the HHs with higher and lower values of $g\left(x_{ijk}\right)$ have higher variances compared to the intermediate values. The **fourth** and **fifth** scenarios are created in the similar manner of Type-I simulation.

In each simulation process, a sample is randomly drawn via two stages: at first 2 clusters are randomly selected from each small area and then HHs are drawn randomly from the selected clusters. Sample cluster sizes n_{ij} are also randomly drawn from a uniform distribution on [10, 20]. In the simulation process population $\left(N = 55068\right)$ and sample

$(n = 1479)$ sizes are retained fixed after first generation of the population structure. The simulation process is iterated $R = 500$ times each with $B = 500$ bootstraps to estimate small area distribution functions and FGT indicators using both ELL and CD-type estimators.

5.5 Simulation Results

The estimators of HM cluster variance component $\hat{\sigma}^2_{u(2)}$ and $\hat{\sigma}^{2(ht)}_{u(2)}$ provide very close estimates in all the scenarios of Type-I and Type-II simulations (appendix **Figure A5.1**). However, the variations of their distributions significantly increase with the heteroskedasticity of HH-level errors in Type-I simulations. In case of misspecified scenarios 4 and 5 under Type-I simulation both the estimators provide bias estimates. Though the simulation results indicate either estimator can be used for cluster-level variance component, $\hat{\sigma}^2_{u(2)}$ is used only in HM cases. Under HT working model, the HH-level error variances are estimated following the ELL parametric approach, the non-parametric local linear regression (LPR) approach and the stratification method of moments (STR). The estimated HH-level error variances against the true error variances for the sample individuals are plotted in appendix **Figure A5.2** (only for Scenario 3 under both types of simulations), which shows that the ELL estimator performs better than the others for perfectly specified and monotone heteroskedasticity function but may fail in the non-monotone situations. The STR estimator provides more stable estimates than the LPR estimator and so only the STR estimates of HH-specific error variances are used in the bootstrap procedure. The results are based on five estimators – PELL, SPELL, CDMC (with naïve residuals), CDSM (with naïve residuals), and SCDMC (with scaled residuals). The SPELL is actually NPELL for the purely HM cases.

5.5.1 Type-I Simulation

The average values of RB (%), RRMSE (%), and BCR over small areas for all the estimators of distribution functions at five quantiles are shown in **Table 5.2** under the considered five scenarios. When the level-one errors are HM, no considerable difference is observed among the estimators in terms of RB, RRMSE and BCR. However, the estimator with scaled residuals SCDMC at lower quantiles and NPELL estimator at upper quantiles are providing slightly lower RB under homoskedasticity. The RB increases with the complexity of the heteroskedasticity in both perfectly and imperfectly specified scenarios. It is observed that the SCDMC estimator performs better comparatively than the other estimators in terms of RB for the all HT scenarios particularly at the lower quantiles. In the HT balanced scenarios (Scenarios 2-4), the estimators based on NPB procedure either raw or scaled residuals are likely to overestimate at lower quantiles and underestimate at upper quantiles. In the misspecified and imbalanced scenario 5 all the estimators behave in the similar trend - highly overestimate at the lower quantile and underestimate at the upper quantiles with higher RRMSE and lower BCR. However, the SCDMC and the PELL estimators perform better compared to others in terms of RB and RRMSE under Scenario 5. In terms of BCR, the estimators with raw residuals show slightly over coverage rate which is obviously for the higher bootstrap variations than the estimated residual variances.

The area-specific true bias for the distribution function estimators shown in **Figure 5.1** reveals similar results observed from the area averaged RB. The CD estimator with scaled residuals (SCDMC) has shown better performances than the others particularly at lower quantiles under all the scenarios. The SPELL estimator shows slightly higher area-specific bias than the CD estimators with raw residuals. The differences among the estimators' performances noticeably increase with the complexity of heteroskedasticity at all quantiles. In the perfectly specified scenarios, no obvious trend of bias is observed with the area-specific population size (here number of HHs) but in the misspecified cases, the observed bias shows an upward trend at lower quantiles and a downward trend at upper quantiles. It is noted that though the bootstrap procedures in CDMC and CDSM are different, they provide almost same area-specific bias but CDMC produces higher BCR than CDSM which may be due to smearing approach.

5.5.2 Type-II Simulation

The area averaged RBs for the FGT poverty estimators shown in **Table 5.3** clearly show that the RB increases with the degree of FGT indicators in case of all scenarios. Under homoskedasticity, the NPELL estimator produces the lowest RB followed by the SCDMC estimators. While under heteroskedasticity, the SCDMC estimator outperformed all the estimators particularly at the upper poverty line (25^{th} quantile) for all the FGT indicators in the balanced scenarios. In the scenarios with monotone and non-monotone HT residual variances under balanced sample (Scenarios 2-4), the CDMC and CDSM are also performing comparatively better than the traditional SPELL estimator. In the misspecified and imbalanced situation it is very tough to say which one is showing better performances that actually happen in reality. The performance of the considered estimators in terms of RB also varies with the degree of FGT indicators.

The area averaged RMSEs given in **Table 5.4** indicate that relative stability of the estimators increases with the poverty line for all poverty indicators but decreases with the order of FGT poverty indicators (from HCR to PS). No significant differences among the estimators are observed in term of RRMSE; however area averaged BCRs shown in **Table 5.5** display some slight differences. The estimators with raw residuals show slightly higher coverage than the PELL and SCDMC estimators in all the scenarios for all poverty indicators at both poverty lines as observed in case of distribution functions. Similar to Type-I simulations, the smearing estimator with naïve residuals (CDSM) shows slightly lower BCR compared to SPELL and CDMC estimators with raw residuals. The SCDMC and PELL are showing almost similar BCRs. The slightly higher bias estimate by SPELL, CDMC and CDSM estimators may another reason of such slightly higher BCRs.

The area-specific bias for the estimators of poverty indicators at the two poverty lines are shown in **Figure 5.2** and **Figure 5.3**. In the pure HM scenario 1, all the estimators are performing in the similar manner for all the poverty indicators at both poverty lines, however NPELL shows slightly lower bias. The proposed CD-based estimators with scaled residuals (SCDMC) outperforms than the other estimators for the scenarios with both monotone and non-monotone heteroskedasticity with balanced sample. The performance of SCDMC estimator is found far better than the others for the upper poverty line compared to the lower poverty line. Also the CDMC and CDSM estimators perform better than the SPELL estimator under the HT scenarios for all poverty indicators at upper poverty line. However, the performances of the estimators vary at the lower poverty line particularly for PG and PS indicators.

Both PELL and SPELL estimators show the trend of upward to downward bias with the degree of FGT indicators in the scenarios with non-monotone heteroskedasticity (Scenarios 3-5). These results may be due to failure of capturing the non-monotone behaviour of heteroskedasticity function. However, all the CD-based estimators show similar trend of bias for all the poverty indicators. The performances of the estimators in the misspecified and imbalanced scenario of Type-II simulation is not as much worse as Type-I simulation may be due to comparatively mild heteroskedasticity of Type-II simulation and the unbiased estimate of $\sigma_{u(2)}^{2(ht)}$ as well.

5.6 Concluding Remarks

This chapter develops a non-parametric estimation procedure of area-specific distribution function and poverty indicators for a finite population defined by a 2-level superpopulation model with unknown HT variances of level-one errors. The estimation procedure is based on two important steps to obtain the ultimate small area estimates: (i) the estimation of level-one error variances following a non-parametric approach and then (ii) the simulation of smearing based CD estimates of cluster-specific parameter at the target quantile. The proposed method shows some evidence that a non-parametric estimation procedure instead of modelling a complex parametric logistic function can be utilized to estimate the HH-level error variances after the estimation of HM cluster-level variance component. Moreover, the proposed stratification-based MOM (STR) does not require any parameter instead of determining the number of strata. It is also evident that the ELL method may fail in the situation where the heteroskedasticity model is very much difficult to develop based on the sample information particularly when potential explanatory variables are not available either in survey or census dataset. In such situations, non-parametric approach may be a better way to handle the situation at least to save the time to search the potential explanators of heteroskedasticity.

Simulation studies based on linear and log-normal super-population models indicate that the proposed CD-type estimators behave slightly better than the ELL-type estimators under balanced and perfectly specified cases, even much better in the situations with non-monotone HH-level heteroskedasticity. The simulation studies with misspecified but balanced scenarios also suggest that both ELL and CD-type estimators may robust to model-misspecification if the samples are found balanced to heteroskedasticity. The simulation result also confirms the necessity of scaling the raw residuals in the bootstrap procedure. The SPELL estimator may behave similar to scaled CD estimators if scaled residuals are utilized in the bootstrap procedure. Thus the CD-type estimators with scaled residuals would be a good alternative to the ELL-type estimators in either case of HM or HT level-one errors. The comparison study also confirms that scaling of the raw residuals in ELL and CD-type estimation methods might be a safe way to reduce the relative bias and increase the stability under both perfectly specified and misspecified scenarios with either balanced or imbalanced sample. Similar to the ELL methodology, the proposed bootstrap-based CD-type estimation method can be easily extended for the two-fold superpopulation model with level-one HT random errors. In **Chapter Six**, applications of two-fold ELL and CD-based estimation methods have been discussed.

Table 5.2: Area averages of relative bias (RB, %), relative root mean squared error (RRMSE, %) and bootstrap coverage rate (BCR, %) of nominal 95% confidence intervals of ELL and CD-type estimators for distribution functions (DFs) by scenarios of Type-I simulation

Hetero-skedasticity	***q***	**RB**					**RRMSE**					**BCR**				
		PELL	SPELL	CDMC	SCDMC	CDSM	PELL	SPELL	CDMC	SCDMC	CDSM	PELL	SPELL	CDMC	SCDMC	CDSM
	0.10	0.429	0.427	0.464	0.313	0.500	23.857	23.903	23.861	23.864	23.858	94.69	94.42	96.82	94.47	95.86
	0.25	0.152	0.177	0.147	0.093	0.172	15.762	15.800	15.768	15.769	15.763	94.61	94.52	96.76	94.46	96.36
None‡	0.50	-0.030	0.006	-0.052	-0.046	-0.036	9.344	9.362	9.347	9.348	9.345	94.52	94.38	96.76	94.50	96.45
	0.75	-0.069	-0.029	-0.077	-0.055	-0.068	4.864	4.872	4.865	4.866	4.866	94.59	94.46	96.84	94.58	96.35
	0.90	-0.053	-0.024	-0.052	-0.036	-0.048	2.286	2.291	2.288	2.288	2.289	94.70	94.39	96.66	94.51	95.65
	0.10	-0.583	0.799	0.573	0.273	0.587	24.126	24.127	24.058	24.062	24.053	95.03	96.94	96.67	94.32	95.85
	0.25	0.028	0.495	0.144	0.067	0.153	15.855	15.897	15.856	15.867	15.854	94.74	96.89	96.72	94.40	96.34
Mild	0.50	0.091	0.077	-0.064	-0.029	-0.060	9.382	9.401	9.373	9.376	9.370	94.65	96.85	96.85	94.57	96.41
	0.75	-0.004	-0.126	-0.100	-0.053	-0.098	4.866	4.869	4.865	4.863	4.861	94.62	96.80	96.68	94.34	96.25
	0.90	-0.035	-0.120	-0.089	-0.067	-0.088	2.274	2.272	2.272	2.270	2.270	94.79	96.78	96.57	94.34	95.58
	0.10	-2.057	3.569	3.571	2.053	3.583	20.323	20.521	20.456	20.282	20.463	94.52	98.53	98.28	93.58	97.77
	0.25	-1.368	0.317	0.248	0.106	0.256	12.988	12.958	12.919	12.939	12.923	94.36	98.59	98.50	93.87	98.20
Strong	0.50	-0.463	-0.762	-0.832	-0.486	-0.828	7.341	7.397	7.371	7.346	7.372	94.63	98.54	98.58	93.97	98.32
	0.75	0.198	-0.380	-0.494	-0.289	-0.494	3.628	3.643	3.651	3.627	3.653	94.85	98.60	98.52	93.97	98.06
	0.90	0.309	-0.001	-0.108	-0.100	-0.108	1.700	1.665	1.676	1.675	1.678	94.83	98.47	98.10	94.20	96.96
	0.10	-6.014	1.828	1.521	-0.122	1.496	21.216	20.376	20.373	20.344	20.388	94.21	98.16	98.03	93.68	97.39
	0.25	-1.246	0.106	0.689	0.252	0.672	13.493	13.417	13.437	13.437	13.440	94.76	98.22	98.25	94.44	97.89
Misspecified Balanced	0.50	-0.205	-0.547	0.081	0.448	0.075	7.834	7.852	7.821	7.839	7.820	94.54	98.09	98.20	94.21	97.90
	0.75	-0.087	-0.679	-0.540	-0.275	-0.542	3.928	3.976	3.950	3.921	3.948	94.73	98.02	97.90	93.84	97.40
	0.90	0.323	-0.241	-0.426	-0.395	-0.426	1.824	1.792	1.829	1.819	1.827	94.72	98.07	97.48	93.27	95.90
	0.10	19.170	24.719	24.810	16.648	24.818	28.205	32.222	32.261	26.523	32.255	85.66	91.08	90.20	86.34	88.01
	0.25	9.353	10.725	10.721	8.504	10.720	16.373	17.207	17.188	15.937	17.179	87.54	94.58	94.34	88.31	93.50
Misspecified Imbalanced	0.50	1.381	1.041	0.985	1.357	0.984	7.989	7.952	7.935	8.015	7.925	92.29	97.69	97.57	92.20	97.30
	0.75	-2.015	-2.600	-2.723	-1.905	-2.724	4.459	4.747	4.819	4.418	4.817	89.10	94.98	94.26	88.65	93.10
	0.90	-1.933	-2.306	-2.427	-1.863	-2.428	2.659	2.934	3.033	2.602	3.032	81.15	87.29	84.62	79.84	79.40

‡ SPELL is NPELL under homoscedasticity

Table 5.3: Area averages of relative bias (RB, %) of ELL and CD-type estimators for FGT poverty indicators by scenarios of Type-II simulation

Hetero-skedasticity	Poverty Line	Poverty Incidence (HCR)[2]					Poverty Gap (PG)					Poverty Severity (PS)				
		PELL	SPELL	CDMC	SCDMC	CDSM	PELL	SPELL	CDMC	SCDMC	CDSM	PELL	SPELL	CDMC	SCDMC	CDSM
None[‡]	0.10	0.201	0.070	0.184	0.118	0.180	0.468	0.297	0.451	0.353	0.449	0.672	0.469	0.654	0.530	0.654
	0.25	0.026	-0.048	0.011	-0.010	0.001	0.148	0.040	0.133	0.084	0.122	0.269	0.134	0.253	0.184	0.242
Monotone	0.10	0.061	0.620	0.365	0.172	0.361	-0.169	0.734	0.699	0.364	0.693	-0.422	0.754	0.983	0.534	0.974
	0.25	0.137	0.286	0.064	0.027	0.056	0.075	0.480	0.267	0.132	0.255	-0.029	0.583	0.440	0.223	0.426
Non-monotone	0.10	1.426	2.019	0.790	0.615	0.788	-0.277	0.603	0.994	0.900	0.992	-2.262	-1.166	0.697	0.687	0.696
	0.25	1.110	1.319	0.185	0.068	0.190	1.053	1.497	0.483	0.342	0.490	0.518	1.146	0.651	0.525	0.660
Misspecified Balanced	0.10	1.208	1.879	0.787	0.632	0.787	-0.072	0.923	1.039	0.926	1.040	-1.555	-0.312	0.947	0.883	0.947
	0.25	0.936	1.189	0.250	0.166	0.266	0.913	1.425	0.536	0.418	0.558	0.523	1.238	0.718	0.599	0.745
Misspecified Imbalanced	0.10	1.133	1.806	0.560	0.379	0.553	0.200	1.215	1.722	1.633	1.709	-0.958	0.321	2.447	2.452	2.429
	0.25	0.720	0.950	-0.256	-0.376	-0.249	0.806	1.309	0.265	0.122	0.268	0.565	1.284	0.816	0.691	0.817

Table 5.4: Area averages of relative root mean squared error (RRMSE, %) of ELL and CD-type estimators for FGT poverty indicators by scenarios of Type-II simulation

Hetero-skedasticity	Poverty Line	Poverty Incidence (HCR)					Poverty Gap (PG)					Poverty Severity (PS)				
		PELL	SPELL	CDMC	SCDMC	CDSM	PELL	SPELL	CDMC	SCDMC	CDSM	PELL	SPELL	CDMC	SCDMC	CDSM
None[‡]	0.10	16.077	16.066	16.077	16.063	16.078	18.853	18.826	18.846	18.828	18.848	21.619	21.586	21.611	21.590	21.612
	0.25	10.882	10.884	10.885	10.877	10.884	13.443	13.437	13.444	13.431	13.441	15.545	15.530	15.542	15.527	15.539
Monotone	0.10	16.209	16.185	16.223	16.216	16.208	19.143	19.079	19.073	19.059	19.059	22.104	22.006	21.964	21.956	21.950
	0.25	10.904	10.911	10.914	10.911	10.913	13.509	13.500	13.500	13.491	13.500	15.684	15.654	15.648	15.635	15.647
Non-monotone	0.10	16.067	16.111	16.010	15.999	16.002	18.624	18.573	18.564	18.550	18.553	21.434	21.260	21.217	21.214	21.203
	0.25	10.994	11.027	10.956	10.957	10.948	13.430	13.465	13.385	13.374	13.372	15.424	15.430	15.384	15.369	15.369
Misspecified Balanced	0.10	16.162	16.244	16.098	16.091	16.101	18.782	18.829	18.738	18.737	18.741	21.551	21.533	21.476	21.496	21.478
	0.25	11.009	11.052	10.978	10.976	10.979	13.515	13.584	13.470	13.464	13.473	15.551	15.618	15.513	15.506	15.515
Misspecified Imbalanced	0.10	16.069	16.121	16.018	16.012	16.015	18.752	18.767	18.757	18.748	18.754	21.557	21.501	21.629	21.645	21.625
	0.25	10.967	10.992	10.970	10.974	10.969	13.450	13.488	13.421	13.416	13.422	15.493	15.525	15.467	15.456	15.469

‡ SPELL is NPELL under homoscedasticity

Table 5.5: Area averages of bootstrap coverage rate (BCR, %) of nominal 95% confidence intervals of ELL and CD-type estimators for FGT poverty indicators by scenarios of Type-II simulation

Hetero-skedasticity	Poverty Line	Poverty Incidence (HCR)					Poverty Gap (PG)					Poverty Severity (PS)				
		PELL	SPELL	CDMC	SCDMC	CDSM	PELL	SPELL	CDMC	SCDMC	CDSM	PELL	SPELL	CDMC	SCDMC	CDSM
None‡	0.10	94.93	94.85	96.84	94.79	95.31	94.81	94.67	96.60	94.71	95.59	94.88	94.62	96.49	94.64	95.01
	0.25	95.00	94.90	96.82	94.96	96.15	94.99	94.86	96.80	94.83	96.50	94.96	94.70	96.71	94.75	96.29
Monotone	0.10	94.91	96.84	96.84	94.72	95.38	94.88	96.88	96.66	94.62	95.49	94.91	96.78	96.51	94.45	94.93
	0.25	95.01	97.02	96.95	94.70	96.21	94.88	97.00	96.95	94.68	96.57	94.87	96.95	96.85	94.62	96.32
Non-monotone	0.10	95.08	96.81	96.61	94.57	95.19	95.10	96.98	96.58	94.32	95.26	95.00	96.76	96.38	94.41	94.59
	0.25	94.76	96.66	96.88	94.81	96.26	94.86	96.90	96.87	94.60	96.34	95.02	97.00	96.76	94.44	96.05
Misspecified Balanced	0.10	94.81	96.74	96.66	94.36	95.09	94.90	96.64	96.40	94.25	95.11	94.82	96.54	96.24	94.10	94.43
	0.25	94.51	96.49	96.64	94.51	95.91	94.53	96.61	96.73	94.31	96.24	94.74	96.62	96.64	94.18	96.00
Misspecified Imbalanced	0.10	95.01	96.68	96.39	94.35	94.91	95.06	96.73	96.40	94.18	95.13	94.99	96.78	96.26	94.17	94.26
	0.25	94.67	96.61	96.72	94.54	95.98	94.75	96.77	96.68	94.37	96.18	94.77	96.76	96.56	94.34	95.96

‡ SPELL is NPELL under homoscedasticity

Figure 5.1: Area-specific bias (smooth line) of ELL and CD-type estimators of distribution function (DFs) at different percentiles by scenarios of Type-I simulation

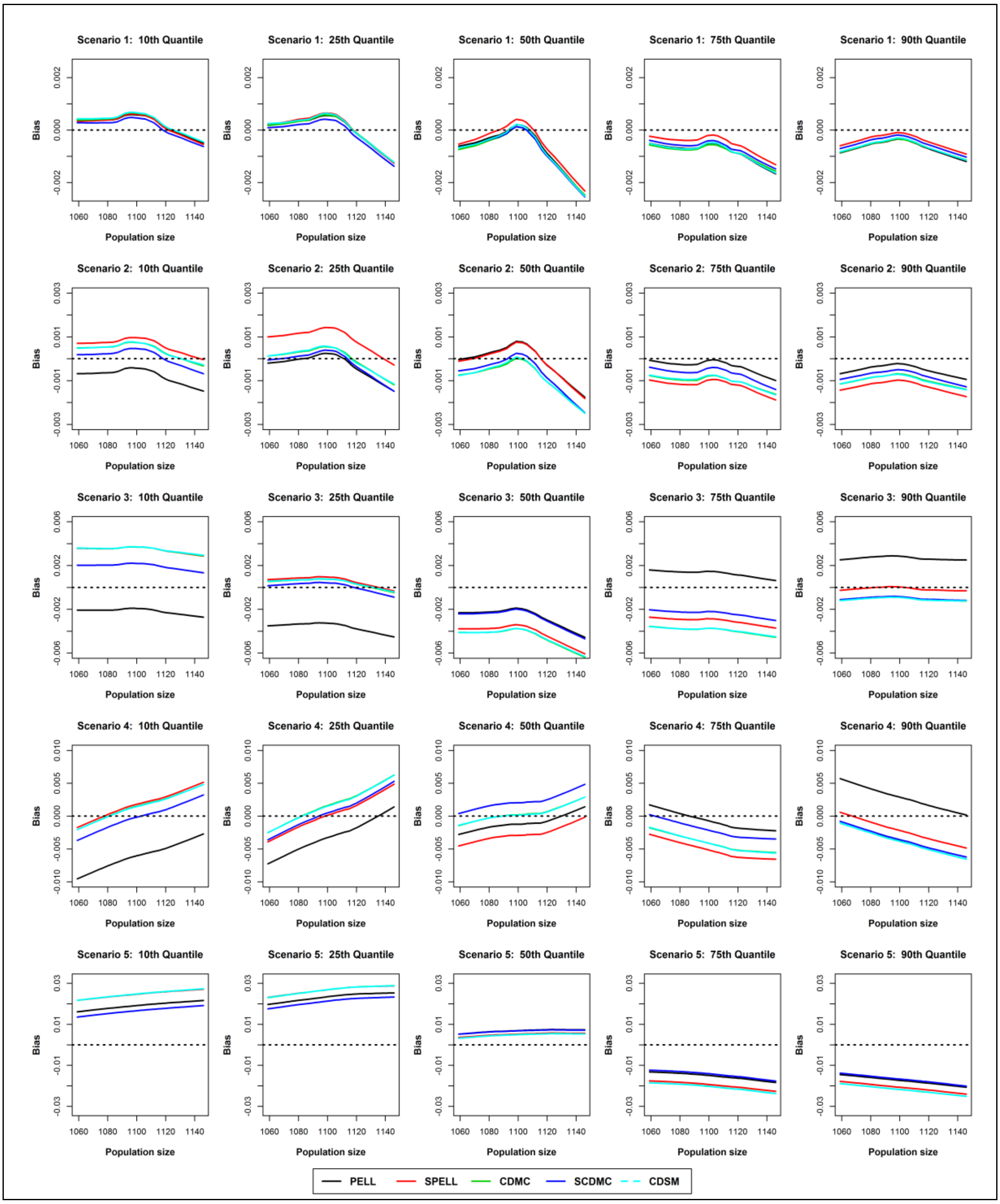


Figure 5.2: Area-specific bias (smooth line) of ELL and CD-type estimators of FGT poverty indicators at lower poverty line corresponds to the 10th percentile by scenarios of Type-II simulation

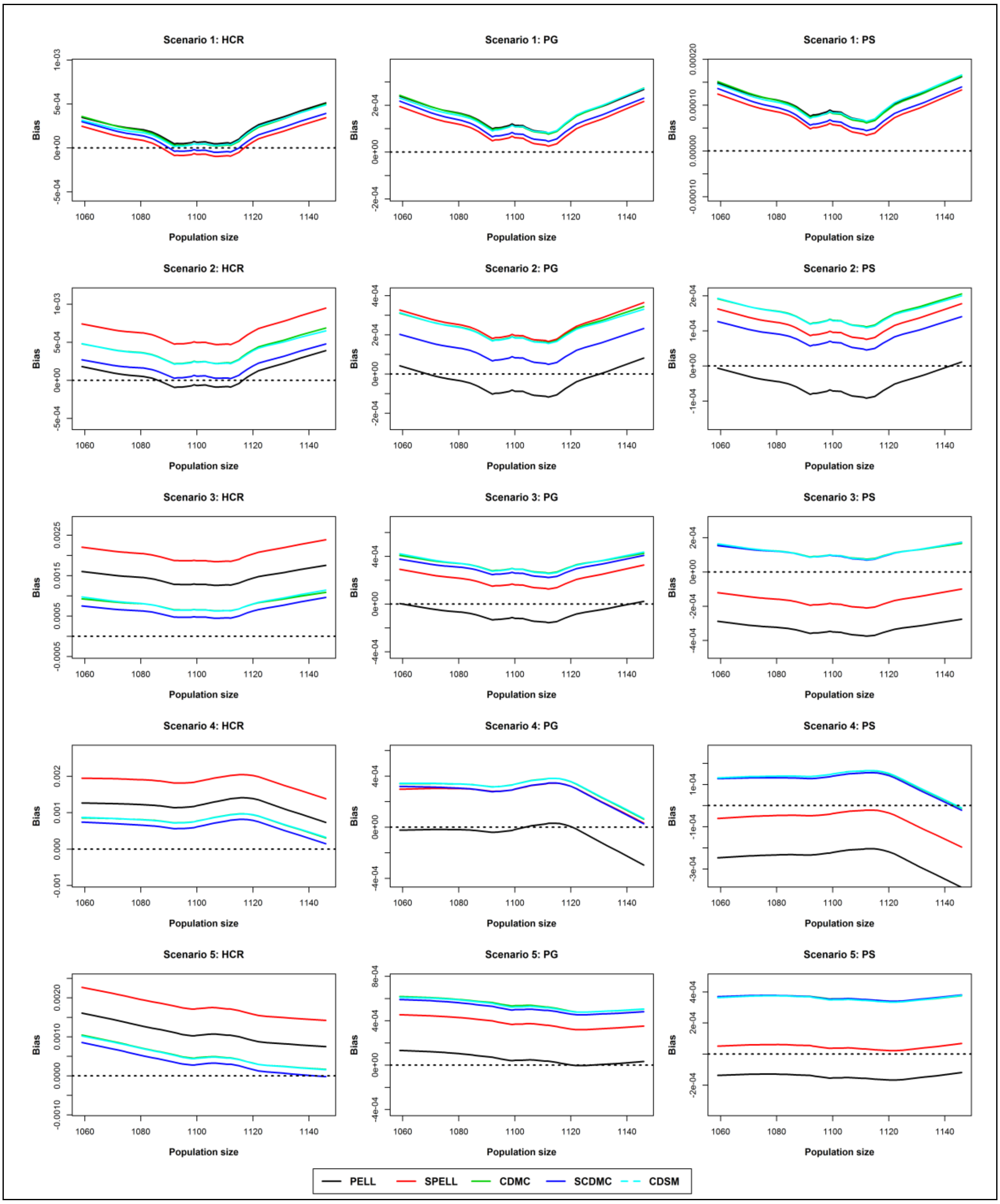


Figure 5.3: Area-specific bias (smooth line) of ELL and CD-type estimators of FGT poverty indicators at upper poverty line corresponds to the 25th percentile by scenarios of Type-II simulation

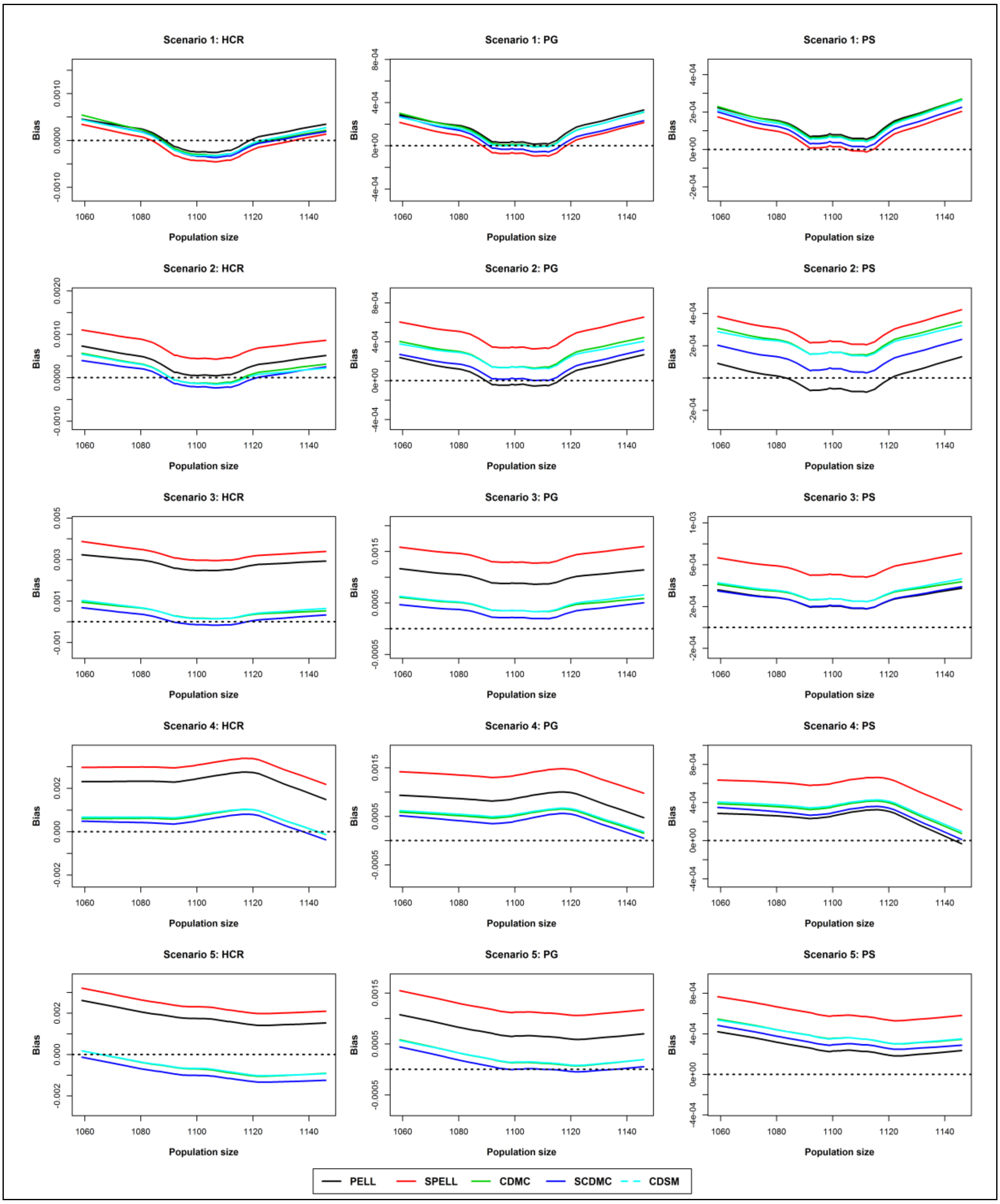


Figure A5.1: Distribution of estimated homoskedastic level-two variance component under the scenarios of Type-I and Type-II simulations

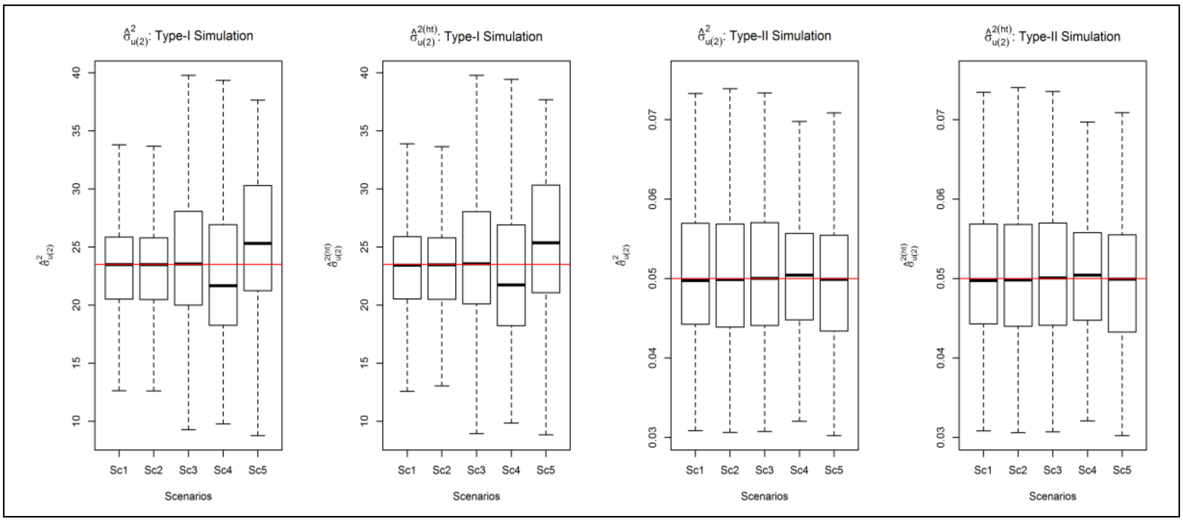
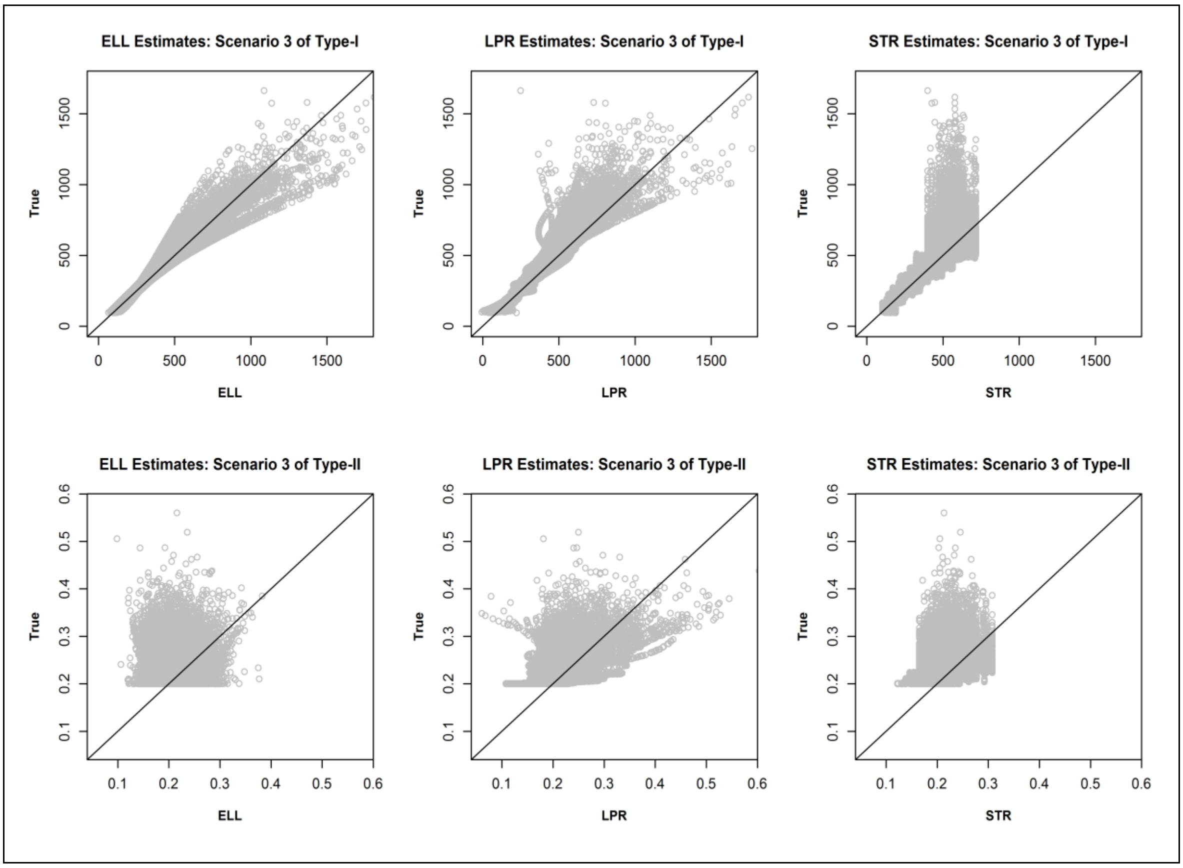


Figure A5.2: Heteroskedastic level-one error variances estimated by ELL, local linear regression (LPR), and stratified method of moments (STR) estimators against true variances under Scenario 3 of Type-I and Type-II simulations



CHAPTER SIX

6. Extensions to ELL Method and Application to Bangladesh Poverty Mapping

The ELL method (Elbers, Lanjouw and Lanjouw, 2003) has been criticized because of the risk of underestimating the mean squared error (MSE) of the ELL-based poverty estimates in situations where its area homogeneity assumption is violated. Its use of a complex parametric approach to modeling heteroskedasticity of household-level errors represents another area of concern. A robust ELL-based MSE estimation methodology that assumes unit-level homoskedasticity, and a semi-parametric approach to modelling unit-level heteroskedasticity have been proposed in Chapter Four and Chapter Five respectively. A non-parametric poverty estimation method based on the same assumption of area homogeneity as the ELL approach has also been developed in Chapter Five based on the smearing approach of Chambers and Dunstan (1986), hereafter the CD method. These ideas are combined in this Chapter, allowing us to extend the ELL and CD methods in order to resolve the issue of underestimating the MSE poverty estimates under heteroskedasticity. In order to evaluate the applicability and flexibility of these extensions in a realistic data scenario, we apply them to data used in a recent Bangladesh poverty mapping study.

The chapter is organized as follows: **Section 6.1** briefly summarizes the Bangladesh poverty mapping study and points out some practical issues relating to the available data sources; **Section 6.2** describes the ELL and CD estimation methods with their modifications for capturing the potential area variability under the assumption of both homoskedastic (HM) and heteroskedastic (HT) household-level errors; **Section 6.3** demonstrates how the proposed methodologies are implemented in the Bangladesh datasets; **Section 6.4** explains the findings, and **Section 6.5** concludes the chapter.

6.1 Background

6.1.1 Poverty Mapping in Bangladesh

Bangladesh is the tenth most densely populated country (1108 per square km) in the world and one of the poorest where poverty head count rate (HCR) at $1.90 a day is near about 45% (World Bank, 2015). To monitor the poverty situation, the poverty rates have been estimating at national and divisional levels using Household Income and Expenditure Survey since 1983-84 based on the food consumption expenditure. To unveil the actual variation in poverty incidence (HCR) at local administrative units, Bangladesh Bureau of Statistics (BBS) in conjunction with United Nation World Food Program (UNWFP) conducted the first poverty mapping study using the Bangladesh 2001 Population and Housing Census (hereafter referred as 2001 Census) and the Bangladesh 2000 Household Income and Expenditure Survey (hereafter referred as 2000 HIES) datasets (BBS and UNWFP, 2004). Recently poverty map is updated using the 2005 HIES data (WB, BBS and WFP, 2009).

Though the national level poverty incidence was about 40.0 percent in 2005 (BBS, 2011), sub-district level poverty incidence varied from about zero to 55.0 percent (WB, BBS and UNWFP, 2009). The ELL method of World Bank has been implemented to obtain the poverty estimates at sub-district level. Both the poverty maps show that the areas close to the capital city Dhaka have lower poverty rates but the actual size of poor population is large. Comparatively, the sub-districts in the Chittagong Hill Tracks (south-eastern part of Bangladesh) have high poverty incidence but the size of population is relatively small. On the other hand, the sub-districts in the northern part (known as seasonal food insecure area) have large population size as well as higher poverty rates (WB, BBS, and WFP, 2009).

6.1.2 Data Sources for Bangladesh Poverty Mapping

In the first poverty mapping study of Bangladesh (BBS and UNWFP, 2004), hereafter referred as BBS-2004 study, the BBS considered only 5% enumeration areas (EAs) from each sub-districts of 2001 Census by systematic sampling instead of using a mammoth full census data. The 5% 2001 Census covers 5 divisions, 64 districts (*Zila*), 507 sub-districts (*Upzila*), 12908 enumeration areas (EAs), 1258240 HHs, and 6,156,000 individuals. The target small domains are the sub-districts which are not considered in the survey sampling design. The structures of the full 2001 Census and the 5% 2001 Census are discussed in **Chapter Three**.

The 2000 HIES is used as the survey dataset which covers 295 out of 507 sub-districts. The sample is drawn following a standard two-stage stratified sampling design where 442 EAs (clusters) are drawn from 16 strata at the first stage and 7428 HHs are drawn from the selected EAs (10-20 HHs per PSU) at the second stage. It is observed that about two-third sampled sub-districts (222 of 295) had single cluster and hence sub-district specific sample sizes are very small. The sampling design and the structure of 2000 HIES data are detailed in **Chapter Three**. The 2001 Census and 2000 HIES datasets are utilized in this Chapter to examine the proposed estimators.

6.1.3 Issues Related to Model Selection

In ELL methodology, the main task is to reduce the overall variation particularly at cluster and area levels by incorporating more explanatory variables at different hierarchies of the population in the model specification. In the BBS-2004 study, 27 HH specific and 3 sub-district specific explanatory variables are considered to fit a two-level random effects model considering HH and cluster as levels one and two respectively. The between-area variation was ignored may be due to about 75% of the sampled sub-districts have single sampled cluster. In this study, an exploration has been made to examine the variations due to cluster and sub-district hierarchies by fitting both 2-level (2L) and 3-level (3L) models considering sub-district as level-three. It is observed that about 20% variations $\left(\sigma_\eta^2 + \sigma_u^2\right)/\sigma_e^2$ in the HH expenditure are due to cluster and sub-district level variations (the row for Set-1 in **Table 6.1**). The contribution of sub-district level variation is found negligible (about 5%) but statistically significant (p-value < 0.0001). Negligible between-area variation and also lack of sufficient survey data to fit an appropriate 3-level model enforce someone to implement the standard 2-level model-based ELL method by ignoring the sub-district level random effects.

Table 6.1: Estimated variance components by method of moments estimator under 2-level (2L) and 3-level (3L) homoskedastic models for different datasets

Data Set	Model	DF	$\hat{\sigma}^2_\varepsilon$	$\hat{\sigma}^2_u$	$\hat{\sigma}^2_\eta$	% $\hat{\sigma}^2_u/\hat{\sigma}^2_e$	% $\hat{\sigma}^2_\eta/\hat{\sigma}^2_e$	MR	CR	p-value of LRT
Set-1:	2L	33	0.1132	0.0253	-	18.28	-	59.84	67.18	
All	3L	34	0.1132	0.0192	0.0062	13.82	4.46	59.97	67.29	<0.00005
Set-2:	2L	33	0.1091	0.0267	-	19.67	-	64.17	71.21	-
Multiple Clusters	3L	34	0.1091	0.0186	0.0082	13.69	6.03	64.50	71.50	0.00010
Set-3: Rural Clusters	2L	27	0.1121	0.0222	-	16.52	-	47.83	56.45	-
Set-4:	2L	27	0.1143	0.0282	-	19.77	-	63.07	70.37	-
Urban Clusters	3L	28	0.1143	0.0215	0.0066	15.12	4.68	63.34	70.60	0.00065

Note: MR – Marginal R-squared, CR – Conditional R-squared

Scrutiny of the 2000 HIES data shown in **Table 6.2** suggests that the sampled sub-districts with single cluster are mainly from rural parts (197 out of 222 rural clusters), while the sub-districts with multiple clusters mainly have urban clusters (183 out of 220 urban and SMA clusters). To examine whether area variability exits in the urban parts of Bangladesh, an investigation has been done by creating a new sample dataset (hereinafter, data set-2) considering the sampled sub-districts with multiple clusters so that both 2-level and 3-level random effects models can be fitted and tested the significance of between-area variability. The row for Set-2 in **Table 6.1** confirms the statistical significance of the sub-district level random effects in the urban parts of Bangladesh.

Both survey and census datasets suggest that a significant number of sub-districts have both urban and rural parts (**Table 6.2**) and so it is considerable to divide each dataset into rural and urban sub-sets, and then check the area variability by residential areas. In the rural survey dataset (hereafter set-3), only 35 out of 232 sub-districts have multiple clusters, while the number is 71 out of 96 sub-districts in the urban survey dataset (hereafter set-4). The row for Set-4 in **Table 6.1** also confirms the significance of area variability in the urban survey dataset. These investigations stand out some questions regarding application of naïve ELL methodology based on a 2-level model which might not capture the existing area variability in 2000 HIES data.

Table 6.2: Distribution of household (HH), cluster, and sub-districts (area) by type of residence in 2000 HIES and 2001 Census

Type of Residence	2001 Census			2000 HIES							
				Overall			Single Cluster		Multiple Cluster		
	HH	Cluster	Area	HH	Cluster	Area	HH	Area	HH	Cluster	Area
Urban	244849	2506	263	2775	208	96	489	25	2286	183	71
Rural	1013241	10403	455	4653	234	232	3929	197	724	37	35
Total	1258090	12908	507	7428	442	295	4418	222	3010	220	73

Note: 1 Census EA has both rural (22) and urban (68) HHs

6.1.4 Alternatives to ELL Methodology

The ELL method is based on the assumption of cluster-heterogeneity rather than area-heterogeneity. If the assumption is violated, the method produces approximately unbiased poverty estimates but with underestimated mean squared errors (MSEs) that produce poor coverage rate of the real poverty estimates (Tarozzi and Deaton, 2009). The estimated MSEs help to prioritize the small areas according to their corresponding poverty estimates. A small area with underestimated MSE will be less prioritized compared to an area with the same poverty estimates with correct MSE. The estimated MSE depends on several issues including (i) fitting regression model, (ii) unexplained variation in the response variable after accounted for the explanatory variables, (iii) estimated variation at higher levels (cluster/area), and (iv) population size of a small area. In this chapter, an investigation has been made on the estimated MSEs of the ELL poverty estimates under the violation of area homogeneity assumption using the Bangladesh datasets.

In the implementation of ELL methodology to real dataset, HH specific random errors are usually considered heteroskedastic (HT) but the cluster random effects are remained homoskedastic (HM) due to availability of a few clusters per sampled area in the survey data set. The ignorance of heteroskedasticity will lead to biased estimates of distribution functions and hence FGT poverty incidences (shown in Chapter Five). In such situation, at first the HM cluster-specific variance component is estimated considering heteroskedasticity at HH-level, and then a heteroskedasticity model known as “alpha” model is developed to estimate the HH-level error variances using the potential explanatory variables. A suitable non-parametric approach proposed in **Chapter Five** can also be applied instead of fitting an alpha model to approximate the HH-level error variances.

In either case of HM or HT level-one errors, a 2-level nested-error regression model is fitted in ELL methodology ignoring area-specific random effects. If area-level variation remains small but significant after incorporating area-level contextual variables in the regression model, the ELL method provides smaller estimated mean squared error (MSE) than the expected one and hence shows under coverage of the true poverty estimates. The modified ELL (MELL) methodology proposed in **Chapter Four** shows better performances in such practical situation. The MELL method has been developed on the assumption of HM level one errors. The proposed MELL method is required to modify to account heteroskedasticity of HH-level errors.

In several poverty mapping studies, the optimistic and the conservative ELL methodologies are implemented (Elbers, *et al.*, 2008; World Bank, 2013) considering heteroskedasticity at HH-level, however, the methodology is based on the assumption of HM random effects for both 2-level and 3-level random effects models. No instruction has been made with respect to heteroskedasticity for the optimistic and conservative methods. An attempt has been made to implement the MELL methodology under heteroskedasticity at HH-level in this chapter.

The CD-based poverty estimation method developed in **Chapter Five** has also been implemented to the Bangladesh data set as an alternative of the ELL methodology. Since the CD-type estimators based on 2-level model also violates the area homogeneity assumption of ELL methodology, the CD estimators are also modified and then implemented to show how the proposed idea of modification performs in both the ELL and CD methods. With these aims, the proposed methodologies developed in **Chapter Four** and **Chapter Five** are modified by considering heteroskedasticity at HH-level.

6.2 The ELL Methodology and Its Extensions

Suppose $y_{ijk} = \log\left(E_{ijk}\right)$ and m_{ijk} indicate the logarithm of per capita household expenditure $\left(E_{ijk}\right)$ and the number of family members of k^{th} household (HH) belonging to the j^{th} cluster in the i^{th} area. Then the area-specific FGT poverty indicators are calculated as $F_{\alpha i} = M_i^{-1} \sum_{j=1}^{C_i} \sum_{k=1}^{N_{ij}} m_{ijk} \left(\frac{t - E_{ijk}}{t}\right)^{\alpha} I\left(E_{ijk} < t\right)$; $\alpha = 0,1,2$ where $M_i = \sum_j^{C_i} \sum_k^{N_{ij}} m_{ijk} = \sum_j^{C_i} M_{ij}$ and C_i are respectively total number of individuals and clusters in i^{th} area; $M_{ij} = \sum_k^{N_{ij}} m_{ijk}$ and N_{ij} are respectively total number of individuals (population) and households (HHs) in j^{th} cluster of i^{th} area. When HH-specific weights are ignored or equal, the FGT indicator becomes $F_{\alpha i} = N_i^{-1} \sum_{j=1}^{C_i} \sum_{k=1}^{N_{ij}} \left(\frac{t - E_{ijk}}{t}\right)^{\alpha} I\left(E_{ijk} < t\right)$ with $N_i = \sum_j^{C_i} N_{ij}$.

In standard ELL methodology a 2-level nested error regression model is considered assuming HHs at level-one and clusters at level-two as

$$
\begin{aligned}
& y_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} + u_{ij} + \varepsilon_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} + u_{ij} + \sigma_{\varepsilon(2)} \gamma_{ijk} \\
& u_{ij} \sim N\left(0, \sigma_{u(2)}^2\right),\ \varepsilon_{ijk} \sim N\left(0, \sigma_{\varepsilon(2)}^2\right),\ \gamma_{ijk} \sim N\left(0,1\right) \\
& i = 1,2,...,D;\ j = 1,2,...,C_i;\ k = 1,2,....,N_{ij}
\end{aligned}
\tag{6.1}
$$

where, u_{ij} and ε_{ijk} are identically and independently distributed cluster-specific and HH-specific HM random errors. Here the sub-script (l) is used to indicate any parameter under a perfectly specified l-level model. If the HH-specific random errors are assumed to be HT, the model (6.1) can be expressed as below

$$\begin{aligned} y_{ijk} &= \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)}^{(ht)} + u_{ij} + \varepsilon_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)}^{(ht)} + u_{ij} + \sigma_{\varepsilon(2),ijk} \gamma_{ijk} \\ u_{ij} &\sim N\left(0, \sigma_{u(2)}^{2(ht)}\right);\ \varepsilon_{ijk} \sim N\left(0, \sigma_{\varepsilon(2),ijk}^2\right)\ \&\ \gamma_{ijk} \sim N(0,1) \end{aligned} \tag{6.2}$$

where the super-script (ht) stands for heteroskedasticity. Now if an area-specific random effect is assumed as an additional one in the above two models, the corresponding 3-level models can be expressed as below.

$$\begin{aligned} y_{ijk} &= \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(3)} + \eta_i + u_{ij} + \varepsilon_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(3)} + \eta_i + u_{ij} + \sigma_{\varepsilon(3)} \gamma_{ijk} \\ \eta_i &\sim N\left(0, \sigma_{\eta(3)}^2\right), u_{ij} \sim N\left(0, \sigma_{u(3)}^2\right),\ \varepsilon_{ijk} \sim N\left(0, \sigma_{\varepsilon(3)}^2\right), \gamma_{ijk} \sim N(0,1) \end{aligned} \tag{6.3}$$

$$\begin{aligned} y_{ijk} &= \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(3)}^{(ht)} + \eta_i + u_{ij} + \varepsilon_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(3)}^{(ht)} + \eta_i + u_{ij} + \sigma_{\varepsilon(3),ijk} \gamma_{ijk} \\ \eta_i &\sim N\left(0, \sigma_{\eta(3)}^{2(ht)}\right), u_{ij} \sim N\left(0, \sigma_{u(3)}^{2(ht)}\right), \varepsilon_{ijk} \sim N\left(0, \sigma_{\varepsilon(3),ijk}\right),\ \gamma_{ijk} \sim N(0,1) \end{aligned} \tag{6.4}$$

Now if the model (6.3) is true but the corresponding 2-level model (6.1) is assumed to conduct the standard 2-level ELL, the estimated MSE will underestimate the true mean squared error (MSE). The similar problem will be appeared if the HT 2-level model (6.2) is used instead of HT 3-level model (6.4) in the implementation of ELL methodology. Since the area homogeneity assumption of ELL method is violated in both the situations, it is necessary to prevent the sub-sequent consequence of underestimated MSE by capturing the level-three variability. The MELL methodology proposed in **Chapter Four** is adapted for accounting the HH-level heteroskedasticity in this chapter.

Estimation of variance components is one of the most difficult tasks when the HH random errors are assumed HT. The estimation methods of HM variance components and the HT error variances are discussed at first and then the ELL- and CD- based methods are discussed under both 2-level and 3-level working models considering HT level-one errors. At last 2-level model-based ELL- and CD-type estimators are modified to account for the ignored area variability of a 3-level true model.

6.2.1 Variance Component Estimation under Heteroskedasticity

The variance components are estimated via method of moment (MOM) approach in this chapter. The MOM approaches for 2-level and 3-level homoskedastic models are explained in **Appendix A.1** and **Appendix A.2** respectively. Under the 2-level HT model (6.2), a moment-based estimator of $\sigma_{u(2)}^{2(ht)}$ can be obtained under the assumption of known HH-level error variances as

$$\hat{\sigma}_{u(2)}^{2(ht)} = \left(\sum_{ij\in s} w_{ij}\left(1-w_{ij}\right)\right)^{-1}\left\{\left[\sum_{ij\in s} w_{ij}\left(\hat{\bar{e}}_{ij.}-\hat{\bar{e}}_{...}\right)^2\right]-\sum_{ij\in s} w_{ij}\left(1-w_{ij}\right)\hat{\tau}_{ij}^2\right\}$$

where $w_{ij}=n_{ij}/n$, $\hat{\bar{e}}_{ij.}=n_{ij}^{-1}\sum_{k}^{n_{ij}}\hat{e}_{ijk}$, $\hat{\bar{e}}_{...}=n^{-1}\sum_{ijk\in s}\hat{e}_{ijk}$, $\hat{\tau}_{ij}^2=\dfrac{1}{n_{ij}\left(n_{ij}-1\right)}\sum_{k}^{n_{ij}}\left(\hat{\varepsilon}_{ijk}-\hat{\bar{\varepsilon}}_{ij.}\right)^2$ and $\hat{\bar{\varepsilon}}_{ij.}=n_{ij}^{-1}\sum_{k}^{n_{ij}}\hat{\varepsilon}_{ijk}$. In the similar manner, the moment-based estimators of $\sigma_{u(3)}^{2(ht)}$ and $\sigma_{\eta(3)}^{2(ht)}$ under (6.3) can be obtained as

$$\hat{\sigma}_{u(3)}^{2(ht)} = \frac{1}{\sum_{ij} w_{ij}-\sum_{i\in s} w_i^{-1}\sum_{j\in s} w_{ij}^2}\left[\begin{array}{l}\sum_{ij\in s} w_{ij}\left(\hat{\bar{e}}_{ij.}-\hat{\bar{e}}_{...}\right)^2-\sum_{i\in s} w_i\left(\hat{\bar{e}}_{i..}-\hat{\bar{e}}_{...}\right)^2\\ -\sum_{ij\in s} w_{ij}\left(1-w_{ij}\right)\hat{\tau}_{ij}^2+\sum_{i\in s} w_i\left(1-w_i\right)\hat{\tau}_i^2\end{array}\right] \text{ and}$$

$$\hat{\sigma}_{\eta(3)}^{2(ht)} = \frac{1}{\sum_{i\in s} w_i\left(1-w_i\right)}\frac{1}{\sum_{ij} w_{ij}-\sum_{i\in s} w_i^{-1}\sum_{j\in s} w_{ij}^2}\times$$
$$\left[\begin{array}{l}\sum_{ij\in s} w_{ij}\left(1-w_{ij}\right)\left\{\sum_{i\in s} w_i\left(\hat{\bar{e}}_{i..}-\hat{\bar{e}}_{...}\right)^2-\sum_{i} w_i\left(1-w_i\right)\hat{\tau}_i^2\right\}\\ -\sum_{i\in s}\left(w_i^{-1}-1\right)\sum_{j\in s} w_{ij}^2\left\{\sum_{ij\in s} w_{ij}\left(\hat{\bar{e}}_{ij.}-\hat{\bar{e}}_{...}\right)^2-\sum_{ij} w_{ij}\left(1-w_{ij}\right)\hat{\tau}_{ij}^2\right\}\end{array}\right]$$

where $w_i=n_i/n$ and $\hat{\bar{e}}_{i..}=n_i^{-1}\sum_{j\in s} n_{ij}\hat{\bar{e}}_{ij.}$. The derivations of these estimators are shown in **Appendices A.6** and **A.7** respectively. The negative value of the estimator will be treated as zero.

The HH-level error variances are estimated by fitting a heteroskedasticity model known as "alpha model" in ELL methodology. Moment-based estimates of HH-level random errors are utilized to develop a logistic-type regression model to estimate the alpha parameters. These estimated alpha parameters are used to obtain the ultimate error variances using the estimator

$$\hat{\sigma}^{2(ELL)}_{\varepsilon(2),ijk} \approx \left[\frac{\hat{A}D_{ijk}}{1+D_{ijk}}\right] + \frac{1}{2}\hat{v}(r)\left[\frac{\hat{A}D_{ijk}(1-D_{ijk})}{\left(1+D_{ijk}\right)^3}\right]$$

where $\hat{A} = 1.05 \times \text{maximum}\left\{\hat{\varepsilon}^2_{ijk}\right\}$, $D_{ijk} = \exp(z^T_{ijk}\hat{\alpha})$, and $z_{ijk} = g\left(x_{ijk}\right)$. The estimated alpha parameters $\hat{\alpha}$ and the estimated mean squared error $\hat{v}(r)$ are obtained from the fitted alpha model. The procedure is detailed in Chapter 5 under 2-level HT working model.

A stratification-based moment method (STR) has also been proposed in **Chapter Five** to estimate HH-level error variances in a flexible way via iterative generalized least square (IGLS) method. The HM variance components and the HT level-one error variances can be obtained simultaneously via the IGLS method so that the HH-level error variances are consistent. The STR estimates $\left(\hat{\sigma}^{2(STR)}_{\varepsilon(2),ijk}\right)$ are used only in the Chambers and Dunstan (CD) approach. Under the 3-level model (6.4), the estimates of $\sigma^2_{\varepsilon(3),ijk}$ can be estimated by the ELL and STR estimators using the corresponding estimated HH-level residuals.

6.2.2 The ELL Method

After fitting the regression model and the corresponding parameters, the second stage of ELL method is to conduct either a parametric bootstrap (PB) or a non-parametric bootstrap (NPB) procedure to obtain the ultimate area-specific poverty estimates and their corresponding estimated mean square errors (ESMEs). In either case of PB or NPB procedure, the basic steps are: (1) generate regression parameters $\boldsymbol{\beta}^*$ from a suitable sampling distribution say multivariate normal distribution $N\left(\hat{\boldsymbol{\beta}}_{gls}, \hat{\mathbf{v}}\left(\hat{\boldsymbol{\beta}}_{gls}\right)\right)$; (2) generate level-specific random errors using an appropriate parametric distribution or by resampling via SRSWR from the estimated level-specific sample residuals; (3) generate bootstrap income values y_{ijk}^* using the generated regression parameters and the level-specific random errors. The generated income values are used to estimate the area-specific parameter say $F_{\alpha i}^* = N_i^{-1}\sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}} I\left[\exp\left(y_{ijk}^*\right) < t\right]$ for a specific poverty line t. These steps are iterated for a large number of times say B =500 and then the mean and variance of the B estimates are considered as the ultimate estimates and their MSEs respectively.

6.2.3 The Bootstrap-based CD Method

The smearing-based CD approach has been developed in **Chapter Five** based on cluster-specific 2-level model. The method can be extended for a 3-level model by developing area-specific 3-level model. Under the 3-level HM model (6.3), unweighted CD-type estimator of cluster-specific distribution function can be expressed as

$$\begin{aligned}\hat{F}_{ij}(t) &= N_{ij}^{-1}\left[n_{ij}F_{s_{ij}}(t) + \left(N_{ij} - n_{ij}\right)\hat{F}_{r_{ij}}(t)\right] \\ &= N_{ij}^{-1}\left[\sum_{ijk\in s_{ij}} I\left(y_{ijk} \le t\right) + \sum_{ijk\in r_{ij}} \hat{G}\left\{\frac{t - \mathbf{x}_{ijk}^T\hat{\boldsymbol{\beta}}_{(3)}^{gls} - \eta_i - u_{ij}}{\hat{\sigma}_\varepsilon}\right\}\right]\end{aligned}$$

given the values η_i and u_{ij}. The smearing estimator $\hat{G}$ can be obtained from area-specific or whole sample information as

$$\hat{G}(\vartheta) = n^{-1}\sum_{jk\in s_i} I\left(\hat{\gamma}_{ijk} \le \vartheta\right) \text{ or } \hat{G}(\vartheta) = n^{-1}\sum_{ijk\in s} I\left(\hat{\gamma}_{ijk} \le \vartheta\right),$$

where $\hat{\gamma}_{ijk} = \dfrac{y_{ijk} - \mathbf{x}_{ijk}^T \hat{\boldsymbol{\beta}}_{(3)}^{gls} - \tilde{\eta}_i - \tilde{u}_{ij}}{\hat{\sigma}_\varepsilon} = \dfrac{\tilde{\varepsilon}_{ijk}}{\hat{\sigma}_\varepsilon}$, $\tilde{\eta}_i = n_{ij}^{-1} \sum_{k=1}^{n_{ij}} \tilde{e}_{ijk}$, $\tilde{u}_{ij} = n_{ij}^{-1} \sum_{k=1}^{n_{ij}} \tilde{e}_{ijk}$ and $\tilde{\varepsilon}_{ijk} = \tilde{e}_{ijk} - \tilde{\eta}_i - \tilde{u}_{ij}$ with $\tilde{e}_{ijk} = y_{ijk} - \mathbf{x}_{ijk}^T \hat{\boldsymbol{\beta}}_{(3)}^{gls}$. In reality, calculation based on area-specific estimator $\hat{G}$ might be critical due to availability of a few clusters per sampled area. The smearing estimator $\hat{G}$ can be rewritten as

$$\hat{G}\left\{\frac{t - \mathbf{x}_{ijk}^T \hat{\boldsymbol{\beta}}_{(3)}^{gls} - \tilde{\eta}_i - \tilde{u}_{ij}}{\hat{\sigma}_\varepsilon}\right\} = n^{-1} \sum_{c=1}^{C_s} \sum_{h=1}^{n_c} I\left(\tilde{\gamma}_{ch} \le \frac{t - \mathbf{x}_{ijk}^T \hat{\boldsymbol{\beta}}_{(3)}^{gls} - \tilde{\eta}_i - \tilde{u}_{ij}}{\hat{\sigma}_\varepsilon}\right)$$

$$= n^{-1} \sum_{c=1}^{C_s} \sum_{h=1}^{n_c} I\left(\mathbf{x}_{ijk}^T \hat{\boldsymbol{\beta}}_{(3)}^{gls} + \tilde{\eta}_i + \tilde{u}_{ij} + \tilde{\gamma}_{ch} \hat{\sigma}_\varepsilon \le t\right) = n^{-1} \sum_{c=1}^{C_s} \sum_{h=1}^{n_c} I\left\{\tilde{y}_{ijk}^{(ch)} \le t\right\}$$

with $\tilde{y}_{ijk}^{(ch)} = \mathbf{x}_{ijk}^T \hat{\boldsymbol{\beta}}_{(3)}^{gls} + \tilde{\eta}_i + \tilde{u}_{ij} + \tilde{\gamma}_{ch} \hat{\sigma}_\varepsilon$. Under the CD approach, the estimate of area-specific distribution function would be obtained as

$$\hat{F}_i^{CD}(t) = \sum_j^{C_i} \hat{F}_{ij}(t) = \sum_j^{C_i} N_{ij}^{-1} \left[\sum_{s_{ij}} I\left(y_{ijk} \le t\right) + \sum_{r_{ij}} \left\{n^{-1} \sum_{ch \in s} I\left(\tilde{y}_{ijk}^{(ch)} \le t\right)\right\}\right].$$

Since all clusters of an area are not available in the sample dataset and also a few sampled clusters are available per sampled small area, the analytic estimator cannot be used to obtain the ultimate area-specific estimate. A NPB procedure is needed to estimate the ultimate area-specific distribution in the proposed CD based approach.

The basic steps of the NPB are as follows: First, develop the area-specific 3-level model (6.3) to the survey data and obtain $\hat{\boldsymbol{\beta}}_{(3)}^{gls}$, $\hat{\mathbf{v}}\left(\hat{\boldsymbol{\beta}}_{(3)}^{gls}\right)$, $\hat{\sigma}_{\eta(3)}^2$, $\hat{\sigma}_{u(3)}^2$, $\hat{\sigma}_{\varepsilon(3)}^2$, and the level-specific residuals $\tilde{\eta}_{ij}$, $\tilde{u}_{ij}$ and $\tilde{\varepsilon}_{jk}$. Second, the level specific errors η_i^*, u_{ij}^*, and ε_{ijk}^* are generated for all the census units from the corresponding estimated sample residuals via resampling with SRSWR. Third, the area-specific bootstrap realization $\hat{F}_i^*(t) = N_i^{-1} \sum_j^{C_i} N_{ij} \hat{F}_{ij}^*(t)$ is calculated from the bootstrap income values. The steps are repeated for B times to calculate the ultimate area-specific parameters with their MSE. The Monte Carlo simulation approach of Marchetti *et al.* (2012) can be easily implemented in the similar way of the ELL semi-parametric approach. The basic difference is that ELL approach is based on the least squares (LS) raw residuals instead of the generalized least squares (GLS) raw residuals.

In case of HT model, the normalized errors $\tilde{\gamma}_{ijk}$ are calculated using the HH-specific error variances as $\tilde{\gamma}_{ijk} = \hat{\sigma}^{-1}_{\varepsilon(3),ijk}\tilde{\varepsilon}_{ijk}$. The residual variances $\sigma^2_{\varepsilon(3),ijk}$ can be estimated by ELL parametric approach or STR non-parametric approach. In this study, the CD based estimators are based on STR estimates. The CD method with smearing approach requires extensive computation time particularly when if all the sample residuals are used in the smearing method. In such situation the smearing method can be conducted by dividing the large population into several sub-populations. Dunstan and Chambers (1989) extended the CD approach to obtain efficient estimates using the "summary information" of the auxiliary variable instead of "full information". This approach can be though to reduce the computational burden of the "full information" CD method.

6.2.4 Modification of the ELL and CD Methods

Both the ELL- and CD- based approach will produce underestimated MSE if the area variability is ignored. Both the approaches based on 2-level working model can be modified to capture the potential area variability following the modified ELL approach which is already developed in **Chapter Four**. In this sub-section, the proposed idea is illustrated considering HH-specific weight and HT random errors. The basis of the modified ELL (MELL) methodology is the bias correction of variance estimator of weighted area mean $\bar{Y}_i = M_i^{-1}\sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}} m_{ijk}y_{ijk}$ under an incorrect 2-level model to make it unbiased under the corresponding 3-level model. Under the 3-level model (6.3), the variance of $\bar{Y}_i$ and its plug-in estimator can be expressed

$$\mathrm{Var}_{(3)}\left(\bar{Y}_i\right)=\sigma^2_{\eta(3)}+\sigma^2_{u(3)}\bar{m}^{(2)}_{Ui}+\sigma^2_{\varepsilon(3)}\bar{m}^{(3)}_{Ui} \text{ and } \hat{\mathrm{V}}_{(3)}\left(\bar{Y}_i\right)=\hat{\sigma}^2_{\eta(3)}+\hat{\sigma}^2_{u(3)}\bar{m}^{(2)}_{Ui}+\hat{\sigma}^2_{\varepsilon(3)}\bar{m}^{(3)}_{Ui}$$

where $\bar{m}^{(2)}_{Ui}=M_i^{-2}\sum_{j=1}^{C_i}M_{ij}^2<1$ and $\bar{m}^{(3)}_{Ui}=M_i^{-2}\sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}}m^2_{ijk}$. Under an incorrect 2-level model (6.1), the variance of $\bar{Y}_i$ and its plug-in estimator can be written as

$$\mathrm{Var}_{(2)}\left(\bar{Y}_i\right)=\sigma^2_{u(2)}\bar{m}^{(2)}_{Ui}+\sigma^2_{\varepsilon(2)}\bar{m}^{(3)}_{Ui} \text{ and } \hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)=\hat{\sigma}^2_{u(2)}\bar{m}^{(2)}_{Ui}+\hat{\sigma}^2_{\varepsilon(2)}\bar{m}^{(3)}_{Ui}.$$

The expectations of the variance estimator $\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)$ under the true 3-level model becomes $E_3\left[\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)\right]=\left\{R\sigma^2_{\eta(3)}+\sigma^2_{u(3)}\right\}\bar{m}^{(2)}_{Ui}+\sigma^2_{\varepsilon(3)}\bar{m}^{(3)}_{Ui}$ which always underestimates the true variance $\mathrm{Var}_{(3)}\left(\bar{Y}_i\right)$ since $\bar{m}^{(2)}_{Ui}<1$ and $R=\dfrac{n-\bar{n}_0^{(3)}}{n-\bar{n}_0^{(2)}}<1$.

Unlike homoskedastic working model, an unbiased plug-in estimator of $\mathrm{Var}\left(\bar{Y}_i\right)$ is difficult to obtain under the multilevel models with HT level-one random errors due to estimation of heteroskedasticity. However, a plug-in consistent estimator of $\mathrm{Var}\left(\bar{Y}_i\right)$ can be obtained if a consistent estimator of HT error variances is available. Under the HT population models (6.4) and (6.2), the variance estimators can be expressed as respectively

$$\mathrm{Var}_{(3)}\left(\bar{Y}_i\right)=\sigma^{2(ht)}_{\eta(3)}+\sigma^{2(ht)}_{u(3)}\bar{m}^{(2)}_{Ui}+\xi^{(3)}_{ijk} \text{ and } \mathrm{Var}_{(2)}\left(\bar{Y}_i\right)=\sigma^{2(ht)}_{u(2)}\bar{m}^{(2)}_{Ui}+\xi^{(2)}_{ijk}$$

where $\xi^{(3)}_{ijk}=M_i^{-2}\sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}}m^2_{ijk}\sigma^2_{\varepsilon(3),ijk}$ and $\xi^{(2)}_{ijk}=M_i^{-2}\sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}}m^2_{ijk}\sigma^2_{\varepsilon(2),ijk}$.

Under the HT population models, it can be shown that $\hat{\sigma}^{2(ht)}_{\eta(3)}$, $\hat{\sigma}^{2(ht)}_{u(3)}$, and $\hat{\sigma}^{2(ht)}_{u(2)}$ are unbiased and consistent estimators under the assumption of known HH-level error variances (please see **Appendices A.6** and **A.7**). Now suppose the HH error variance estimator $\hat{\sigma}^2_{\varepsilon(2),ijk}$ and $\hat{\sigma}^2_{\varepsilon(3),ijk}$ are consistent estimators of $\sigma^2_{\varepsilon(2),ijk}$ and $\sigma^2_{\varepsilon(3),ijk}$ respectively. Then the consistent plug-in estimators of $\mathrm{Var}_{(3)}\left(\bar{Y}_i\right)$ and $\mathrm{Var}_{(2)}\left(\bar{Y}_i\right)$ can be considered as

$$\hat{\mathrm{V}}_{(3)}\left(\bar{Y}_i\right) = \hat{\sigma}^2_{\eta(3)} + \hat{\sigma}^2_{u(3)}\bar{m}^{(2)}_{Ui} + \hat{\xi}^{(3)}_{ijk} \text{ and } \hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right) = \hat{\sigma}^2_{u(2)}\bar{m}^{(2)}_{Ui} + \hat{\xi}^{(2)}_{ijk}$$

where $\hat{\xi}^{(3)}_{ijk} = M_i^{-2}\sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}} m^2_{ijk}\hat{\sigma}^2_{\varepsilon(3),ijk}$ and $\hat{\xi}^{(2)}_{ijk} = M_i^{-2}\sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}} m^2_{ijk}\hat{\sigma}^2_{\varepsilon(2),ijk}$. Under the assumption of known $\sigma^2_{\varepsilon(2),ijk}$, it can be shown that $E_3\left[\hat{\sigma}^{2(ht)}_{u(2)}\right] = \sigma^{2(ht)}_{u(3)} + \frac{n-\bar{n}^{(3)}_0}{n-\bar{n}^{(2)}_0}\sigma^{2(ht)}_{\eta(3)}$. Then the expectations of $\hat{\mathrm{V}}_{(3)}\left(\bar{Y}_i\right)$ and $\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)$ under the HT 3-level model (6.4) can be expressed as

$$E_3\left[\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)\right] \approx \left\{R\sigma^{2(ht)}_{\eta(3)} + \sigma^{2(ht)}_{u(3)}\right\}\bar{m}^{(2)}_{Ui} + \xi^{(3)}_{ijk} \;<\; \sigma^{2(ht)}_{\eta(3)} + \sigma^{2(ht)}_{u(3)}\bar{m}^{(2)}_{Ui} + \xi^{(3)}_{ijk} \approx E_3\left[\hat{\mathrm{V}}_{(3)}\left(\bar{Y}_i\right)\right]$$

under the assumption of $E_3\left[\hat{\xi}^{(3)}_{ijk}\right] \approx E_3\left[\hat{\xi}^{(2)}_{ijk}\right] \approx \xi^{(3)}_{ijk}$.

Thus the estimator $\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)$ might underestimate the true variance $\mathrm{Var}_{(3)}\left(\bar{Y}_i\right)$ in both the HM and HT cases. An area-specific adjustment or robustification can be done to make $\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)$ an unbiased or approximately unbiased estimator of $\mathrm{Var}_{(3)}\left(\bar{Y}_i\right)$ as

$$\hat{\mathrm{V}}^M_{(2)}\left(\bar{Y}_i\right) = \left\{\left(1/\bar{m}^{(2)}_{Ui}\right)\hat{\sigma}^2_{\eta(3)} + \hat{\sigma}^2_{u(3)}\right\}\bar{m}^{(2)}_{Ui} + \hat{\sigma}^2_{\varepsilon(2)}\bar{m}^{(3)}_{Ui} \ \&$$

$$\hat{\mathrm{V}}^M_{(2)}\left(\bar{Y}_i\right) = \left\{\left(1/\bar{m}^{(2)}_{Ui}\right)\hat{\sigma}^{2(ht)}_{\eta(3)} + \hat{\sigma}^{2(ht)}_{u(3)}\right\}\bar{m}^{(2)}_{Ui} + \hat{\sigma}^{2(ht)}_{\varepsilon(2)}\bar{m}^{(3)}_{Ui}$$

which are approximately unbiased under the true 3-level model. The variance estimator $\hat{\mathrm{V}}_{(2)}^{M}\left(\bar{Y}_i\right)$ would be robust under model misspecification, since $\hat{\sigma}_{\eta(3)}^{2}$ might be very small (close to zero) under a true 2-level model, and hence the first term of $\hat{\mathrm{V}}_{(2)}^{M}\left(\bar{Y}_i\right)$ might be negligible.

In the modified ELL and CD methods, the cluster-level variance component is adjusted to capture the potential area-level variability. Under HM and HT cases, the adjustment will be same without the estimated cluster-specific and area-specific variances components (details are given in **Appendix A.8**). The adjusted cluster-level variance component under HM working model can be calculated as below:

$$\hat{\hat{\sigma}}_u^2 = k_1\hat{\sigma}_{u(2)}^2 \text{ with } k_1 = \hat{\sigma}_{u(2)}^{-2}\left[\hat{\sigma}_{\eta(3)}^2 D_s^{-1}\sum_{i=1}^{D_s}\left(1/\bar{m}_{si}^{(2)}\right)+\hat{\sigma}_{u(3)}^2\right],$$

$$\hat{\hat{\sigma}}_u^2 = k_2\hat{\sigma}_{u(2)}^2 \text{ with } k_2 = \hat{\sigma}_{u(2)}^{-2}\left[\hat{\sigma}_{\eta(3)}^2 D^{-1}\sum_{i=1}^{D}\left(1/\bar{m}_{Ui}^{(2)}\right)+\hat{\sigma}_{u(3)}^2\right], \text{ and}$$

$$\hat{\hat{\sigma}}_u^2 = k_3^{(h)}\hat{\sigma}_{u(2)}^2 \text{ with } k_3^{(h)} = \hat{\sigma}_{u(2)}^{-2}\left[\hat{\sigma}_{\eta(3)}^2 D_{(h)}^{-1}\sum_{i=1}^{D_{(h)}}\left(1/\bar{m}_{Ui}^{(2)}\right)+\hat{\sigma}_{u(3)}^2\right]; \quad h=1,\ldots,H$$

where $\bar{m}_{si}^{(2)} = m_i^{-2}\sum_{jk\in s} m_{ijk}^2$, D_s is number sampled areas, m_i is total members observed in i^{th} sampled area, m_{ij} is total members observed in j^{th} sampled cluster of i^{th} sampled area.

The modified ELL can be implemented via both PB and NPB procedures. In PB procedure, the cluster level residuals are generated from a suitable parametric distribution with the adjusted cluster-variance component, say $N\left(0,\hat{\sigma}_u^{2(M)}\right)$ where $\hat{\sigma}_u^{2(M)}$ is the adjusted cluster-specific variance component. In the NPB procedure, the cluster-level scaled residuals $\left(\tilde{u}_{ij}\right)$ are required to rescale in such way so that the ratio of bootstrap variations $\operatorname{var}\left(\breve{u}_{ij}\right) / \operatorname{var}\left(\tilde{u}_{ij}\right)$ approximates the ratio of corresponding estimated variance components $\hat{\sigma}_u^{2(M)} / \hat{\sigma}_{u(2)}^2$, where scaled residuals are $\breve{u}_{ij(2)} = \tilde{u}_{ij(2)} \hat{\sigma}_u^{(M)} \left\{ \left(C_s - 1\right) \sum_{ij \in s} \tilde{u}_{ij(2)}^2 \right\}^{-1/2}$ and $\tilde{u}_{ij(2)} = \hat{u}_{ij(2)} \hat{\sigma}_{u(2)} \left\{ \left(C_s - 1\right) \sum_{ij \in s} \hat{u}_{ij(2)}^2 \right\}^{-1/2}$. The procedure will be similar for all the proposed modifications. For the stratification based MELL, the generation of cluster residuals or resampling the sample cluster-specific residuals will be conducted based on the stratum-specific adjusted cluster-variance component, say $\hat{\sigma}_{u.h}^{(M)} = k_3^{(h)} \hat{\sigma}_{u(2)}^2$, $h = 1,..,H$. The bootstrap procedures are also explained in next section.

6.3 Implementation of ELL Method and Its Alternatives

6.3.1 First Stage Regression

The first task in the ELL methodology is to fit the regression model utilizing the survey dataset. It is recommended to incorporate considerably a large number of explanatory variables at different hierarchies in order to capture the potential between-cluster and between-area variabilities. Overfitting the regression model is the main problem of including more explanatory variables. Also inclusion of more explanatory variables and contextual variables may not guarantee the reduction of between-area variability sufficiently. To avoid such overfitting problem, log of HH per capita monthly consumption expenditure is regressed on those 30 explanatory variables that are used in BBS-2004 study. The list of the explanatory variables is included in the appendix **Table A3.1**. For the first two datasets (Set-1 and Set-2), all these explanatory variables are used to develop both 2-level and 3-level models. In case of rural (Set-3) and urban (Set-4) datasets, the explanatory variable “rural” that reperesents type of clusters and the corresponding two-way interections are not included in fitting the regression models.

In the ELL methodology, the level-specific errors and the variance components are estimated using moment-based estimators after fitting a linear model via LS method. The estimated variance-components are then utilized to obtaine GLS regresion parameters and their variance-covariance matrix. Marginal and conditional R-squared values are calculated following Nakagawa and Schielzeth (2013) to compare the multilevel models. Both 2-level and 3-level regression models are developed, and the corresponding estimated values of parameters and level-specific random errors are stored for the bootstrap procedure. The multilevel models are developed considering both HM and HT variances of HH-specific random errors only for the first two-data sets.

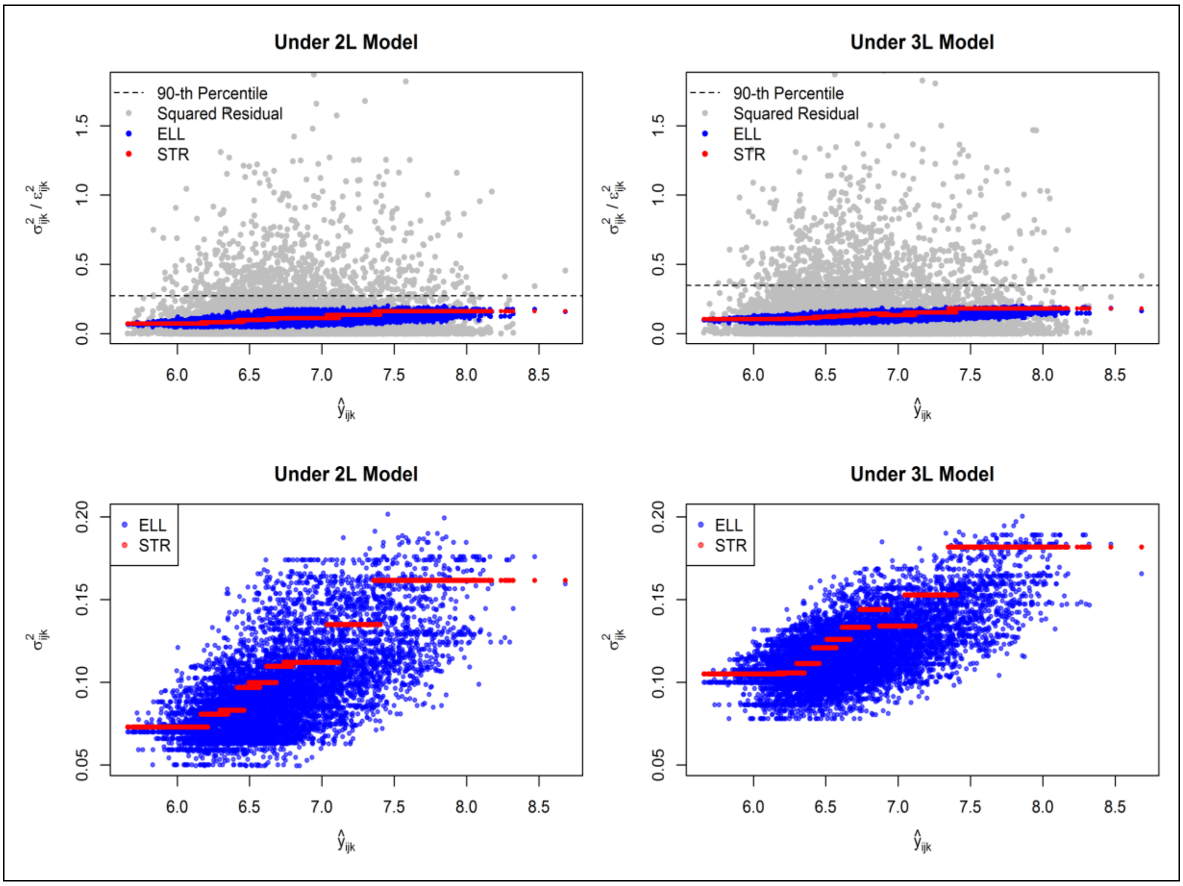
6.3.2 Heteroskedasticity Modelling

To examine the heteroskedasticity of the HH-level random errors, the squared LS residuals obtained from the Set-1 are plotted against the predicted values (**Figure 6.1**). The plot suggests a negligible monotone heteroskedasticity due to less information at the extreme tails. The "alpha" model is fitted with the explanatory variables that are used in the BBS-2004 study to avoid the extensive exploration of potential explanators of heteroskedasticity. The estimated "alpha" parameters are shown in appendix **Table A6.1**. The exploration of heteroskedasticity explanators is not rquired for the proposed STR estimation approach which depends only on the GLS residuals and the predicted values.

Both the ELL and STR estimation methods have been utilized to model the HH-level error variances. The **Figure 6.1** shows how the STR method goes through the center of the ELL estimates. The ELL estimates are used in ELL methodologies and the STR estimates in the CD based methods. The HT error variances estimated by the ELL parametric approach under 2-level model ($\hat{\sigma}^{2.ELL}_{\varepsilon(2),ijk}$) are fluctuating more than those under 3-level model ($\hat{\sigma}^{2.ELL}_{\varepsilon(3),ijk}$) which may be due to ignoring the existing area variability in the survey dataset. The HT variances estimated by either ELL or CD method are found slightly higher with less variation when estimated under a 3-level model than those under a 2-level model.

The estimated area-level and cluster-level homoskedastic variance components under the assumption of heteroskedastic HH-level errors are shown in appendix **Table A6.2**. The variance components estimated by MOM and stratified MOM approaches are found close to those estimated under the assumption of homoskedastic HH-level errors. These results suggest a negligible influence of level-one heteroskedasticity in the estimation of higher level variance components.

Figure 6.1: Unit-level heteroskedastic error variances estimated by ELL and stratified method of moments (STR) estimators under 2-level and 3-level models



6.3.3 Bootstrapping

In this study both PB and NPB bootstrap procedures have been implemented to calculate the FGT poverty estimates and their MSEs. Under the HT model, the NPB procedure is mentioned as semi-parametric bootstrap (SPB) since the HH-level error variances are generated assuming parametric distribution of the alpha parameters. In both types of bootstrap procedure, the regression parameters are generated from multivariate normal distribution with mean vector $\boldsymbol{\beta}_{(.)}^{gls}$ and variance-covariance matrix $\hat{\mathbf{v}}\left(\boldsymbol{\beta}_{(.)}^{gls}\right)$. In PB procedure, level specific random errors are generated following normal distribution with zero mean and the corresponding estimated variances as variance. In NPB and SPB procedures, the LS raw residuals are used under the ELL methods, and the GLS raw residuals under the CD based approaches to generate level-specific random errors via simple random sampling with replacement (SRSWR). Under a 2-level model, the residuals can be drawn either from sample raw residuals or scaled residuals if the bootstrap variation is same as the estimated variance component. Under a 3-level model, the moment-based cluster-specific residuals and are-specific residuals are same for the 75% sub-districts and so all the residuals are required to scale. In the study scaled residuals are utilized in all bootstrapping under both 2-level and 3-level models.

Under heteroskedasticity at HH-level, there is no recommended way to scale the HH-level residuals when they are put in bootstrapping. Similar to Chapter Five, the HH-specific HT residuals are scaled by HH-level variance component estimated as $\hat{\sigma}_e^2 - \hat{\sigma}_{u(2)}^{2(ht)}$ under a 2-level model where $\hat{\sigma}_e^2$ is the MSE of initial single-level linear model fitted by LS method. The mean of estimated HH-level residual variances (say, $n^{-1}\sum\hat{\sigma}_{\varepsilon(2),ijk}^{2.ELL}$) or the HM variance component $\hat{\sigma}_{\varepsilon(2)}^2$ can be used for the scaling purpose.

Resampling of the level-specific residuals via SRSWR can be implemented either unconditionally or conditionally according to ELL (2002) method. Under the unconditional sampling method, the level-specific errors are assigned to census units from the full set of sample residuals. Under the conditional sampling method, the level-specific residuals are drawn following a nested approach. Suppose a cluster-specific sample residual is randomly assigned to a census cluster. Then census HHs nested within a census cluster are assigned HH-level random errors from the sub-set of sample HHs nested within the selected sample cluster.

Under a 3-level working model, conditional approach can be used if a sufficient number of sampled areas have multiple clusters. In the study, conditional approach has been followed under a 2-level model and unconditional approach under a 3-level model. It is observed that both unconditional and unconditional approaches behave similar under a 2-level model but conditional approach produce unstable results under a 3-level model. The ELL methodologies have been performed via PB, NPB, and SPB procedures, while only NPB procedure is followed in CD based methods.

Similar to NPB-based ELL, the Monte Carlo simulation based CD method (CDMC) is implemented following conditional approach under 2-level model and unconditional approach under 3-level model. In the smearing based CD method (CDSM), only cluster- and area-specific random errors are drawn via SRSWR from the sample residuals and the HH-specific residuals are remained same in each bootstrap.

Since ELL and CD based methods can be implemented in different ways based on (i) bootstrap procedure (PB/NPB/SPB), (ii) skedasticity of HH-level random errors (HM/HT), and (ii) the assumed working model (2L/3L); the estimators are denoted differently. As for example “PELL.HM.2L” stands for PB based ELL estimator under a 2-level HM working model. A list of the estimators is given in appendix **Table A6.3**.

6.3.4 Application of Modified ELL and CD Methods

Under the assumption of HM and HT level-one errors, the stratification-based adjustment (Adjustment 3) has been applied to implement the modified version of ELL and CD methods. Sub-districts are grouped into six strata according to 20th, 35th, 50th, 65th, and 80th percentile of the distribution of their population size. Cluster-level variance component $\hat{\sigma}^2_{u(2)}$ is adjusted for each stratum by $\hat{\sigma}^{2(h)}_{u(2)} = k_3^{(h)}\hat{\sigma}^2_{u(2)}$, $h = 1,...,6$. The same stratification has been used under both HM and HT working models. Thus stratum specific bootstrap procedures are conducted based on the stratum-specific adjusted cluster-level variance component in the ELL and CD methods.

6.3.5 Mixed ELL Methodology

It is observed that between-area variability exists in urban sub-districts but doesn't in rural sub-districts. Since a sampled area may comprise both urban and rural clusters, implementation of either a cluster-level 2-level model or an area-level 3-level model may mislead the overall variation. Moreover, the socio-economic characteristics vary with the place of residence. To avoid such problem, an attempt has been made to implement separate multilevel models for rural and urban clusters in a single ELL bootstrap procedure. To conduct the mixed ELL bootstrap, the HHs belong to urban clusters are assigned three level-specific errors estimated under an area-specific 3-level model and the HHs belong to rural clusters are assigned two level-specific errors estimated under a cluster-specific 2-level model. Thus the bootstrap income values are calculated based on the location of cluster in a small area as

$$y^*_{ijk} = \begin{cases} \mathbf{x}^T_{ijk}\boldsymbol{\beta}^*_{(2)} + u^*_{ij(2)} + \varepsilon^*_{ij(2)} & \text{if HH in Rural Cluster} \\ \mathbf{x}^T_{ijk}\boldsymbol{\beta}^*_{(3)} + \eta^*_{i(3)} + u^*_{ij(3)} + \varepsilon^*_{ij(3)} & \text{if UU in Urban Cluster} \end{cases}$$

where the sub-script (l) stands for l-level model. The procedure would be complex but may be reasonable to capture the actual variability observed in the survey data set. This procedure is implemented based only on ELL method following PB procedure under the assumption of HM level-one errors and hence the estimator is denoted as MIX.PELL.

6.4 Results and Discussion

In the 2000 HIES, lower and upper poverty lines were developed to measure the FGT poverty indices. Poverty lines were set up according to the 16 strata which were created for the survey sampling design. Since the 2000 HIES was based on 1991 Bangladesh Population and Housing Census, poverty lines were rearranged according to 2001 Census in the poverty mapping study of BBS-2004. The poverty lines used in the BBS study are not exactly maintained in this Chapter for some sub-districts due to lack of information. The poverty lines shown in appendix **Table A3.5** are utilized in this study.

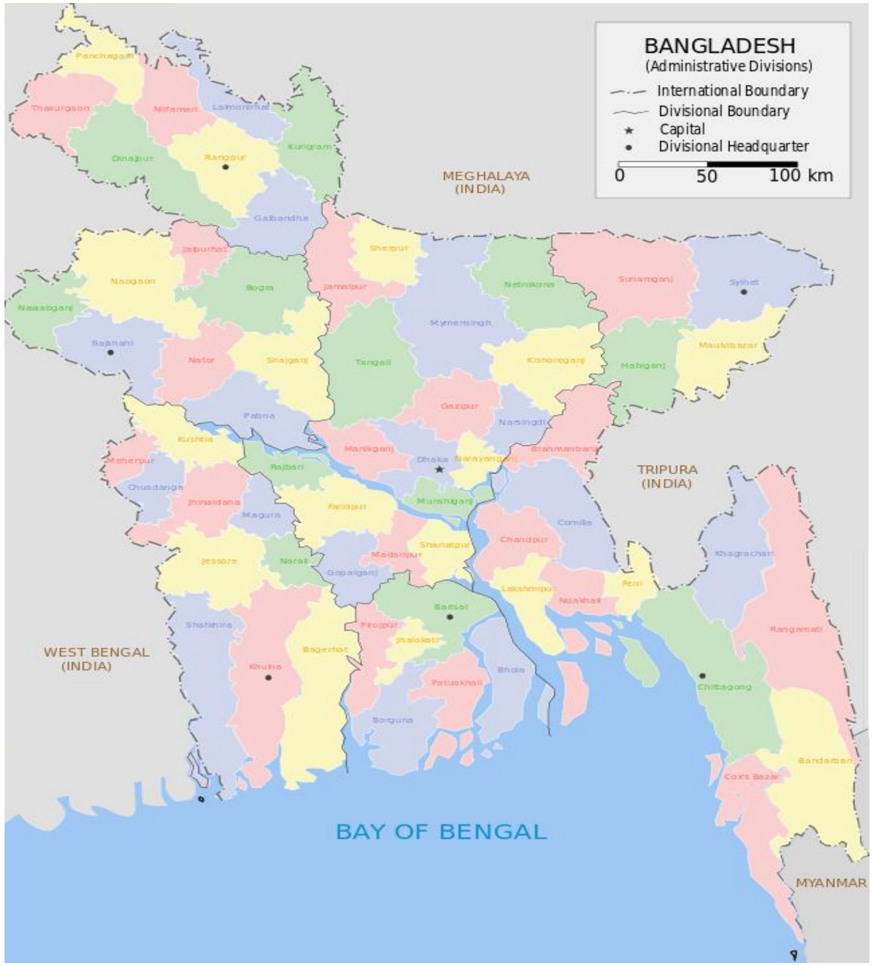
In BBS-2004 study, sub-district level poverty incidences were calculated using the ELL method with SPB via conditional approach under a 2-level working model. In this thesis the similar approach with scaled residuals is followed to create a poverty map of Bangladesh. Figure 6.2 shows an administrative map and a poverty map where sub-district level poverty incidences are calculated at the lower poverty line (LPOVLN). The map is comparable to the poverty map reported in BBS and UNWFP (2004, page 35). The sub-districts at the north-western (Rangpur division) and mid-western (Rajshahi division) regions have higher poverty incidence compared to the central (Dhaka division) region. Also the hilly areas of the south-eastern (Bandarban district) region, the wetland (called *haor*) areas of the mid-northern (*Mymensingh and Netrokona* districts) region and some sub-districts at the coastal region (Bhola district) are vulnerable to poverty incidence.

Consistency of the estimated FGT poverty indices is checked by comparing summary statistics of the estimated poverty measures and their MSEs with those of BBS and UNWFP (2004, page 34). **Table 6.3** shows that means of the estimated FGT measures are slightly lower with slightly higher standard deviation (SD) compared to those of BBS results, while means and SDs of the estimated MSEs are marginally smaller. The reasons of these differences may be (i) different procedure of fitting the regression model, (ii) different way of heteroskedasticity modelling, (iii) different bootstrap procedure, (iv) slightly different poverty lines for some sub-districts, and (v) the number of resampling in bootstrap procedure (500 instead of 100 bootstrap). Also two covariates created be us behave differently in the fitted regression model.

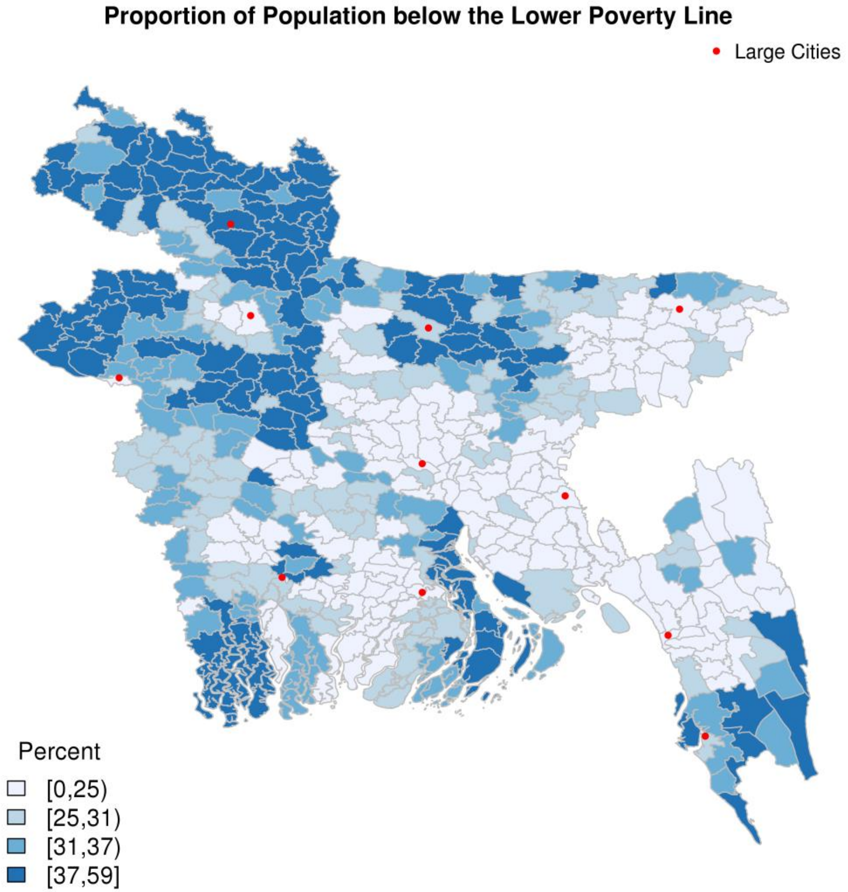
Summary statistics of the estimated MSEs by the 3-level model-based ELL (SPELL.HT.3L), and the modified ELL (MSPELL.HT) estimators are shown in **Table 6.3** along with SPELL.HT.2L estimator. The SPELL.HT.2L estimator shows underestimated MSEs in comparison to the SPELL.HT.3L estimator, and the MSPELL.HT estimator overcomes this underestimation under area variability. The SPELL.HT.2L estimator might provide better accuracy if really area variability is absent.

In the BBS-2004 poverty mapping study, MSEs were associated with the FGT estimates and number of households. For examining the relationship, sub-district specific population size, estimated HCR at upper poverty line (UPOVLN), and estimated MSE (SPELL.HT.2L and MSPELL.HT) are plotted in **Figure 6.3**. First three maps of Figure 6.3 suggest that sub-districts with large population and closer to the capital (Dhaka) or port cities (Chittagong) have lower HCRs with lower MSEs. Conversely, sub-districts with smaller population in coastal and hilly regions have higher HCRs with higher MSEs. Some sub-districts in the north-western part with large population (Rangpur and Rajshahi divisions) have relatively higher HCRs with lower MSEs.

Figure 6.2: Bangladesh maps of administrative units and sub-district level poverty incidences at lower poverty line



Source: https://en.wikipedia.org/wiki/Administrative_geography_of_Bangladesh



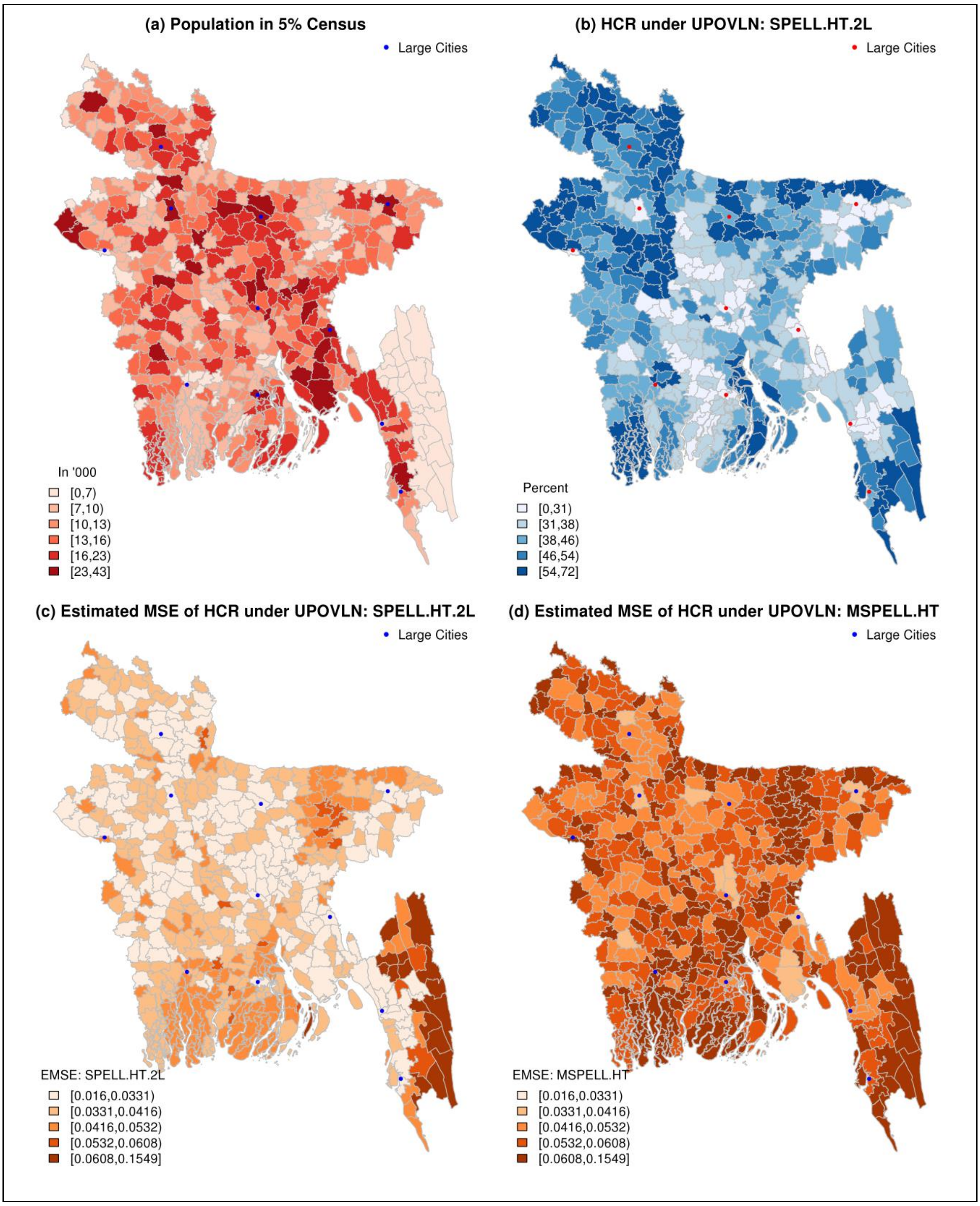
Source: Author's calculation

Table 6.3: Comparison of summary statistics of the estimated HCR, PG, and PS at lower (LPOVLN) and upper (UPOVLN) poverty lines with their estimated MSEs by different estimators assuming unit-level heteroskedasticity (HT) with those of BBS-2004 study

Parameter	Poverty line	Estimated by	Min.	Q1	Median	Q3	Max.	Mean	SD
HCR	LPOVLN	BBS-2004	0.0014	-	-	-	0.5531	0.2930	0.1063
		SPELL.HT.2L	0.0089	0.1930	0.2797	0.3686	0.5898	0.2810	0.1174
	UPOVLN	BBS-2004	0.0049	-	-	-	0.7081	0.4212	0.1238
		SPELL.HT.2L	0.0453	0.3302	0.4224	0.5115	0.7239	0.4164	0.1280
PG	LPOVLN	BBS-2004	0.0001	-	-	-	0.1638	0.0679	0.0297
		SPELL.HT.2L	0.0012	0.0364	0.0585	0.0840	0.1732	0.0622	0.0325
	UPOVLN	BBS-2004	0.0006	-	-	-	0.2441	0.1142	0.0421
		SPELL.HT.2L	0.0073	0.0745	0.1056	0.1389	0.2462	0.1082	0.0443
PS	LPOVLN	BBS-2004	0.0000	-	-	-	0.0650	0.0228	0.0112
		SPELL.HT.2L	0.0003	0.0104	0.0177	0.0275	0.0676	0.0201	0.0121
	UPOVLN	BBS-2004	0.0001	-	-	-	0.1091	0.0429	0.0182
		SPELL.HT.2L	0.0019	0.0240	0.0371	0.0512	0.1079	0.0392	0.0190
Estimated MSE of HCR	LPOVLN	BBS-2004	0.0025	-	-	-	0.1145	0.0388	0.0146
		SPELL.HT.2L	0.0047	0.0268	0.0330	0.0395	0.1047	0.0340	0.0122
		SPELL.HT.3L	0.0070	0.0542	0.0642	0.0712	0.1152	0.0617	0.0159
		MSPELL.HT	0.0146	0.0478	0.0545	0.0622	0.1533	0.0553	0.0146
	UPOVLN	BBS-2004	0.0062	-	-	-	0.1081	0.0416	0.0143
		SPELL.HT.2L	0.0127	0.0302	0.0351	0.0418	0.1027	0.0371	0.0115
		SPELL.HT.3L	0.0226	0.0630	0.0680	0.0737	0.1147	0.0680	0.0118
		MSPELL.HT	0.0304	0.0523	0.0580	0.0644	0.1549	0.0596	0.0135
Estimated MSE of PG	LPOVLN	BBS-2004	0.0003	-	-	-	0.0436	0.0127	0.0054
		SPELL.HT.2L	0.0008	0.0076	0.0103	0.0130	0.0313	0.0106	0.0045
		SPELL.HT.3L	0.0012	0.0150	0.0201	0.0247	0.0401	0.0196	0.0071
		MSPELL.HT	0.0037	0.0183	0.0218	0.0253	0.0549	0.0216	0.0061
	UPOVLN	BBS-2004	0.0009	-	-	-	0.0538	0.0169	0.0066
		SPELL.HT.2L	0.0027	0.0112	0.0139	0.0168	0.0434	0.0144	0.0051
		SPELL.HT.3L	0.0045	0.0224	0.0271	0.0320	0.0538	0.0269	0.0073
		MSPELL.HT	0.0085	0.0234	0.0268	0.0301	0.0730	0.0270	0.0066
Estimated MSE of PS	LPOVLN	BBS-2004	0.0001	-	-	-	0.0203	0.0055	0.0026
		SPELL.HT.2L	0.0003	0.0029	0.0041	0.0056	0.0125	0.0044	0.0021
		SPELL.HT.3L	0.0003	0.0054	0.0079	0.0106	0.0210	0.0081	0.0035
		MSPELL.HT	0.0013	0.0088	0.0110	0.0133	0.0250	0.0110	0.0035
	UPOVLN	BBS-2004	0.0002	-	-	-	0.0300	0.0082	0.0036
		SPELL.HT.2L	0.0009	0.0048	0.0064	0.0083	0.0209	0.0068	0.0028
		SPELL.HT.3L	0.0014	0.0094	0.0125	0.0156	0.0268	0.0126	0.0044
		MSPELL.HT	0.0033	0.0123	0.0148	0.0174	0.0386	0.0149	0.0041

The maps (c) and (d) in **Figure 6.3** show that the SPELL.HT.2L MSE estimator shows more accuracy of HCRs compared to the MSPELL.HT estimator. These performances of SPELL.HT.2L are due to the population sizes of sub-districts. The higher the population size, the lower the MSE if there is no other potential source of variability. Ignorance of area variability is the main reason of such lower MSEs by the SPELL.HT.2L estimator. If area variability exits in reality, this accuracy will mislead the policy makers.

Figure 6.3: Bangladesh maps of sub-district specific population, estimated poverty incidence at upper poverty line (UPOVLN) and their estimated MSEs (EMSE) by SPELL.HT.2L and MSPELL.HT estimators



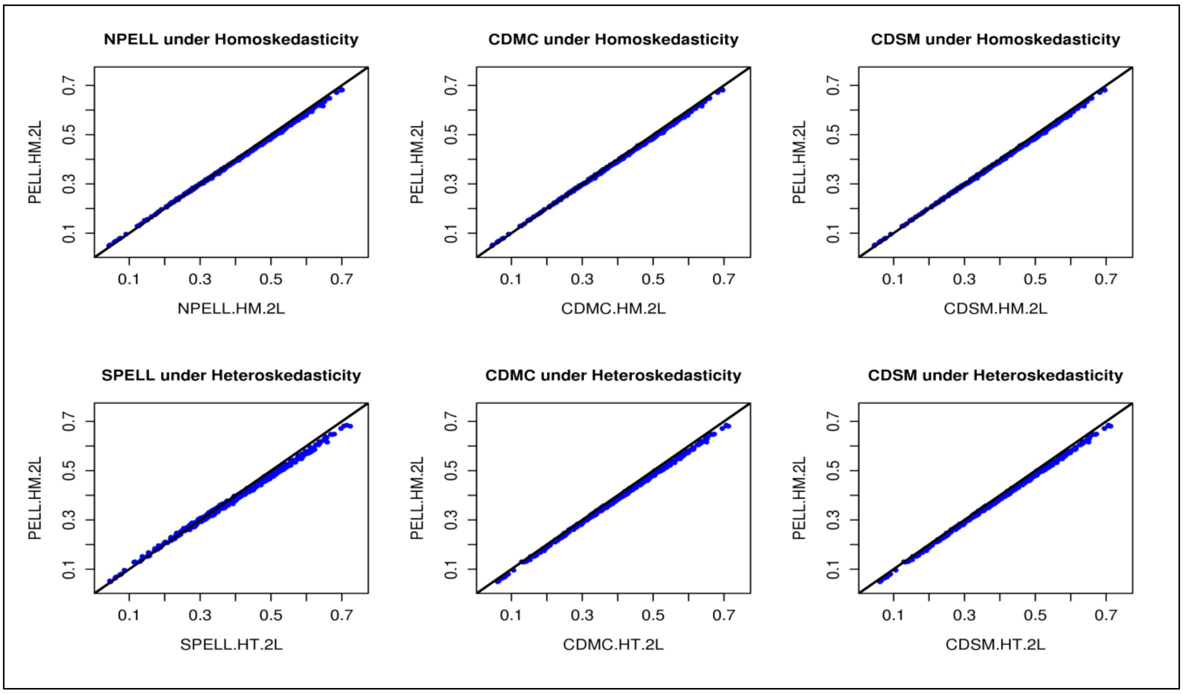
Under the assumption of area variability, it can be said that the SPELL.HT.2L estimator provides underestimated MSEs particularly for the large cities (e.g., blue points) due to ignorance of area variability and the MSPELL.HT estimator adjusts them by considering the potential area variability. The SPELL.HT.2L and MSPELL.HT estimators show similar MSEs only for the sub-districts with significantly smaller sub-districts particularly in the south-eastern regions. Apparently variation in the estimated MSEs is more explicit in Map (c) compared to Map (d). The reason may be the tendency of providing lower MSE for the sub-districts with a modest population size (more than 13,000) and highest MSE for the smaller sub-districts by the SPELL.HT.2L estimator, while MSPELL.HT estimator provides comparatively higher weight to the smaller sub-districts and comparatively lower weight to the larger sub-districts for accounting the between-area variability. With respective to the presence of negligible area variability, SPELL.HT.2L is highly optimistic and MSPELL.HT is reasonably conservative. More detail comparisons are given in the next sub-sections. The proposed FGT estimators, MSE estimators, and mixed ELL approach are compared with the traditional 2-level model-based ELL estimators considering only UPOVLN in the following sub-sections.

6.4.1 Comparison of Poverty Estimators

The FGT poverty indicators at UPOVLN are estimated assuming both HM and HT HH-level errors under 2-level and 3-level working models. The HCRs estimated by different ELL- and CD-type estimators are plotted against the HCRs estimated by the standard 2-level model-based ELL estimator with PB procedure (PELL.HM.2L).

Figure 6.4 shows that the NPB-based ELL- and CD-type estimators under a 2-level working model (NPELL.HM.2L, CDMC.HM.2L, and CDSM.HM.2L) provide almost same results as PELL.HM.2L estimator under homoskedasticity. The HT estimators (SPELL.HT.2L, CDMC.HT.2L, and CDSM.HT.2L) provide slightly overestimated HCRs compared to the PELL.HM.2L estimator except the SPELL.HT.2L estimator which shows slight underestimation for the areas with lower HCRs.

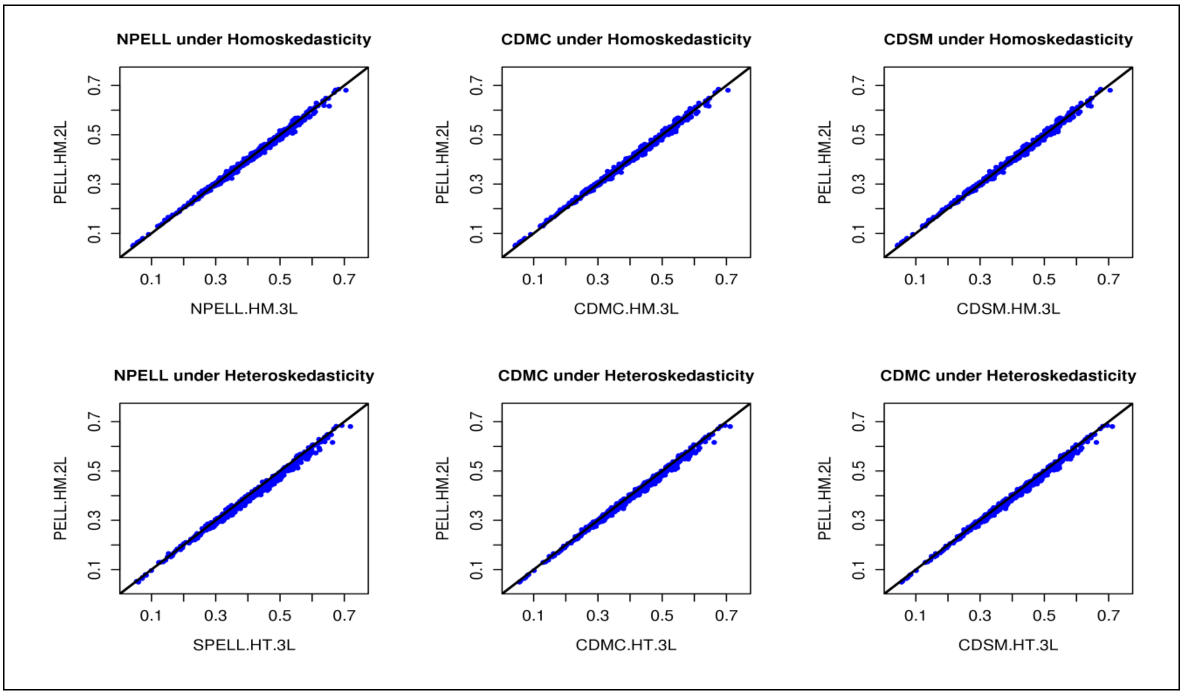
Figure 6.4: Estimated HCR at upper poverty line (UPOVLN) under 2-level working model with homoskedastic (HM) and heteroskedastic (HT) level-one errors by the ELL and CD-type estimators via conditional bootstrap



The estimators based on 3-level working model perform similar to the 2-level model-based estimators with some fluctuations (**Figure 6.5**), which suggest that 3-level model-based estimators also provide unbiased poverty estimates as 2-level model-based estimators. The insignificant variation among the results of ELL- and CD-type estimators under both 2-level and 3-level working models suggest applicability of the proposed CD method as well as the non-parametric estimation method (STR) of heteroskedasticity. Also minor differences among the FGT estimates calculated by the HM and HT estimators recommend that the observed heteroskedasticity does not influence much the estimated poverty measures under homoskedasticity.

To illustrate complexity of utilizing a 3-level model in the Bangladesh datasets, FGT measures are estimated via both conditional and unconditional bootstrap procedure. Appendix **Figure A6.1** shows that unconditional bootstrap with scaled residuals provides stable HCRs and estimated MSEs compared to conditional bootstrap with scaled residuals. One of the main reasons of such difference might be due to a large number of sampled areas (about 75%) with single sampled cluster.

Figure 6.5: Estimated HCR at upper poverty line (UPOVLN) under 3-level working model with homoskedastic (HM) and heteroskedastic (HT) level-one errors by the ELL and CD-type estimators via unconditional bootstrap



6.4.2 Comparison of MSE Estimators

The naïve ELL- and CD-type estimators based on 2-level model have higher possibility to produce underestimated MSE if the between-area variability is not captured by the explanatory variables included in the regression model. Fitting a 3-level model with the available data is also difficult with respect to estimation of variance components and bootstrapping. The modified ELL- and CD-type estimators are expected to provide compromised results compared to the naïve 2-level and 3-level model-based estimators.

Estimated MSEs calculated by the ELL- and CD-type estimators and their modified versions under the assumption of both HM and HT level-one errors are plotted against area-specific population sizes in **Figure 6.6**. The figure shows declining trend of estimated MSEs with population size proportionally. As expected the 2-level model-based ELL- and CD-type MSE estimators are showing lower estimates than those of the 3-level model-based estimators. The modified MSE estimators overcome the underestimation problem of 2-level model-based estimators. In full dataset (Set-1), the modified estimators provide MSE close to the naïve 3-level estimator for the areas with small population, but shows lower MSE for the areas with large population. Differences between the estimated MSEs calculated by the 2-level and 3-level model-based estimators increase with the population size. The proposed modification for both ELL- and CD-type estimators performs similarly in all types of bootstrap procedure under both the HM and HT working models. Performances slightly vary for PG and PS measures. Appendix **Figure A6.2** shows that the modified estimators provide estimated MSEs very nearly to the 3-level model-based estimators for PG and slightly higher for PS.

Since both the ELL- and CD-type estimators perform similarly, only the parametric and non-parametric ELL approaches under homoskedasticity have been implemented for the data set-2. When area variability is obvious in the data Set-2, the modified estimators based on PB procedure perform almost in similar way as the data Set-1 (please see Figure 6.7 and appendix **Figure A6.3)**. The modified ELL estimators show underestimated MSE only for HCR but perform much better than the naïve 2-level ELL estimators (**Figure 6.7**). This underestimation problem of the modified estimators disappears for the PG and PS indicators (**Appendix Figure 6.3**).

Figure 6.7 also shows no significant difference between conditional and unconditional NPB approaches under the 2-level working model (similar to PELL estimator) but interestingly the conditional bootstrap under the 3-level working model behaves differently from unconditional bootstrap though small areas have multiple clusters in the survey dataset. Under the conditional approach, the naïve 3-level estimator is providing MSE estimates closer to those of the naïve 2-level estimators (Figure 6.7 and appendix

Figure A6.3). Thus in either case of conditional or unconditional bootstrap procedures, the modified 2-level ELL- or CD-type estimators provide better estimated MSEs compared to the naïve 2-level estimators in both Set-1 and Set-2 datasets.

Figure 6.6: Estimated MSE (EMSE) of HCR at upper poverty line (UPOVLN) under 2-level and 3-level models with homoskedastic (HM) and heteroskedastic (HT) level-one errors by ELL and CD-type estimators with their modified versions

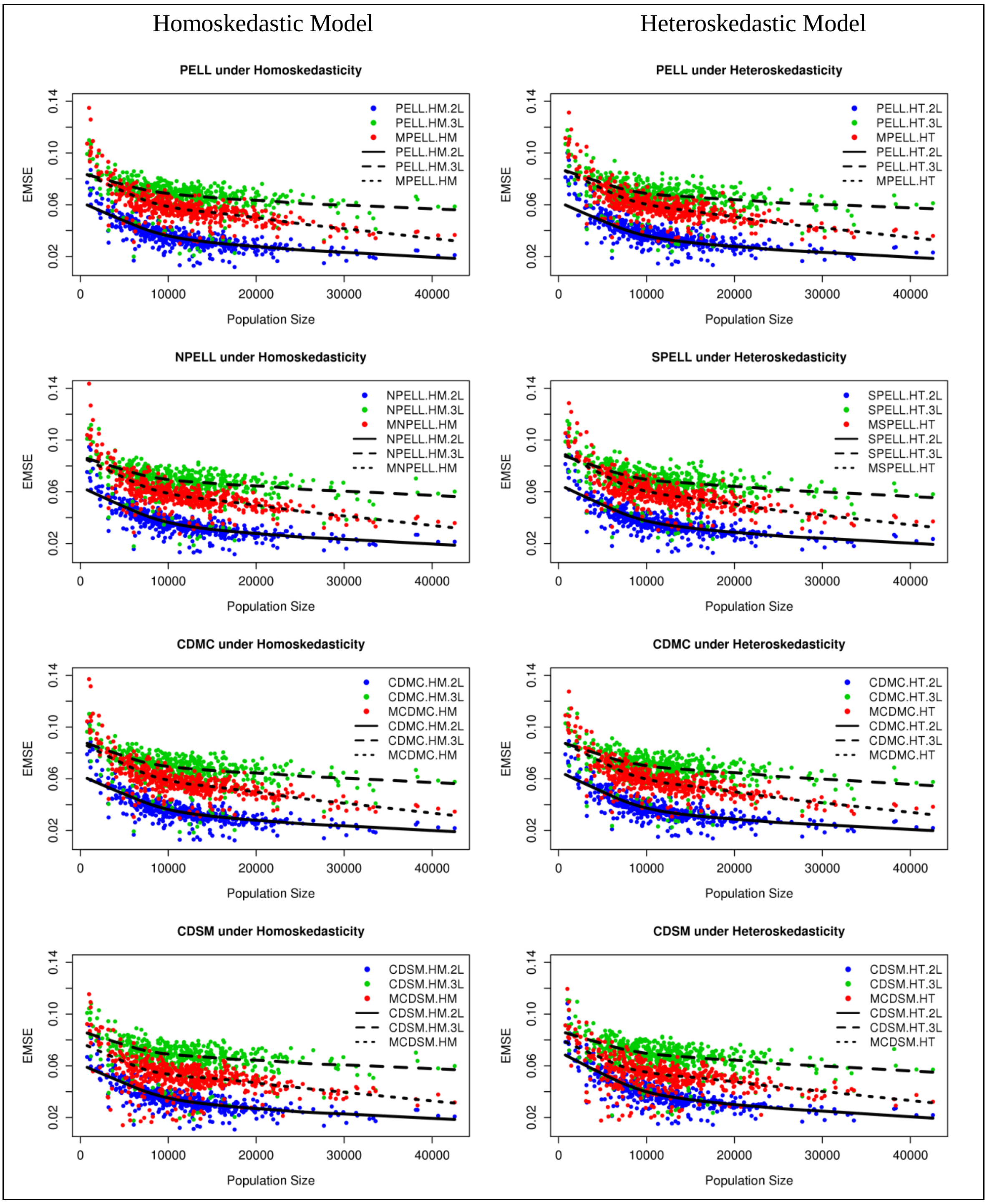
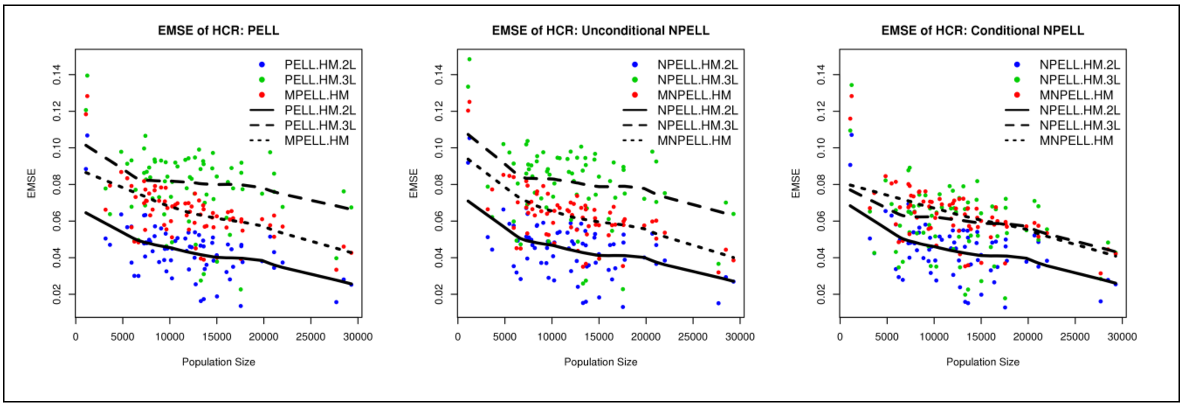


Figure 6.7: Estimated MSE (EMSE) of HCR at upper poverty line (UPOVLN) under 2-level and 3-level models with homoskedastic (HM) level-one errors by PELL and NPELL estimators with their modified versions for sampled sub-districts with multiple clusters

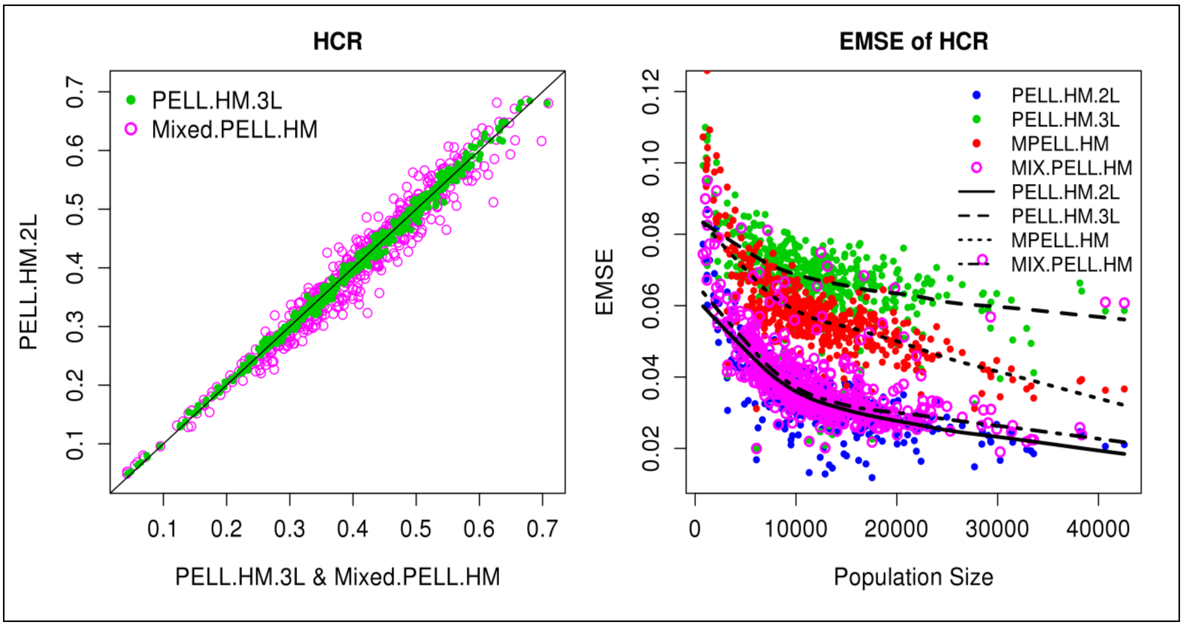


6.4.3 Comparison of Mixed ELL Approach

For Set-3 and Set-4 datasets, only the parametric ELL approach has been utilized to conduct the mixed ELL approach (MIX.ELL) under homoskedasticity. **Figure 6.8** shows the performance of MIX.ELL estimator for both FGT estimates and their estimated MSEs. The estimated HCR calculated by the naïve 3-level estimator (PELL.HM.3L) and MIX.ELL estimator are plotted against the estimated HCR by the naïve 2-level ELL estimator (PELL.HM.2L). The MIX.ELL estimator shows more variation in the estimated HCR than those of ELL.3L estimator. The variation may be due to the standard errors of the regression parameters estimated separately for 2-level and 3-level models. Also sample sizes vary for the two data sets. The fitted regression models for different datasets are shown in appendix **Table A6.4**.

In case of MSE estimation, the MIX.ELL estimator shows similar trend of the PELL.HM.2L estimator with some exceptions. The behaviour might be due to more sub-districts with rural clusters where 2-level model is used in the bootstrap procedure. For the sub-districts with more urban clusters, the naïve 2-level estimator is still underestimating compared to the MIX.ELL estimator. The MIX.ELL estimator provides higher estimated MSEs for a significant number of sub-districts with smaller and larger population (pink circles). For the sub-districts with large population, the PELL.HM.2L estimator is significantly underestimating in comparison to the MIX.ELL estimator due to ignorance of area variability in urban parts. In comparison to the estimated MSE by MIX.ELL estimator, the naïve 2-level estimator seems better for the rural sub-districts but the problem of underestimation remains for the sub-districts with urban clusters where the actual population size is large but the poverty rate is low. For PG and PS, performances of the MIX.ELL estimator shown in appendix **Figure A6.4** are found similar to those of HCR.

Figure 6.8: Estimated HCR at upper poverty line (UPOVLN) with estimated MSE (EMSE) under 2-level and 3-level homoskedastic (HM) models by PELL, modified PELL, and MIX.ELL estimators



6.5 Concluding Remarks

In this chapter we have shown that the proposed CD-based estimators perform similar to the ELL-type estimators in a real poverty mapping dataset of Bangladesh. Also the proposed non-parametric approach of modelling heteroskedasticity via simple stratification-based moment approach provides a flexible way of estimating HH-level error variances. All possible versions of ELL method have been implemented and compared with the CD-based methods. These comparisons support the applicability of the CD-based estimation methods as an alternative of the well-known ELL methodology. In other way, the CD-based method (CDMC) can be considered as a non-parametric (instead of semi-parametric) version of ELL method which depends on parametric modelling of heteroskedasticity. Though the smearing-based CD-type estimators (CDSM) require extensive time to implement for a big data, it may produce more stable results than the ELL methodology does. The MC simulation based CD approach (CDMC) which is parallel to the ELL methodology overcomes this computational problem and can approximate the CDSM results.

It has been shown that both the ELL- and CD-type estimators based on 2-level working model must fail to capture the true MSE if the working regression model violates the area homogeneity assumption. In the Bangladesh dataset, the area homogeneity assumption is violated particularly in the urban small areas (sub-districts). Since ELL- or CD-type estimators based on 3-level model are not suited for the considered dataset perfectly (most sampled sub-districts have single sampled clusters), the 2-level model-based ELL- and CD-type estimators are used to obtain stable FGT estimates with underestimated MSE. We have shown that the proposed modified version of both ELL- and CD-type estimators overcome this underestimation problem of 2-level model-based naïve ELL- and CD-type estimators. Under the assumptions of both HM and HT level-one errors, the proposed modification idea suits properly. The properties of the modified version have also been examined under the situation where area variability is obvious (data set-2) and negligible (data set-1). In both situations, the modified estimators performed better than the naive 2-level ELL- and CD-type estimators. The modified estimators show slightly underestimated MSE compared to 3-level estimators only for HCR but better than the naïve 2-level estimators.

The application of mixed ELL method (simultaneous consideration of 2-level and 3-level models) to Bangladesh dataset suggests the appropriateness of naïve 2-level model since most of the small areas mainly have rural clusters where area variability does not exist. However, the small areas with urban clusters where population density is higher compared to the rural small areas, the naïve ELL method underestimates the MSE compared to both the mixed model and the 3-level model. The modified ELL method helps to overcome this MSE error problem for the small areas with urban infrastructure. These findings support the applicability of the proposed modified version of ELL- and CD-type estimators particularly in such situation where the area-heterogeneity problem cannot be solved by including contextual variables in the model specification.

The 2-level model-based naïve MSE estimators provide more accuracy of the FGT estimates than the modified 2-level and naïve 3-level MSE estimators under the assumption of area homogeneity. However, such accuracy of 2-level naïve estimators will mislead the policy makers in decision making if area variability exists in reality. The results of this chapter show how the naïve 2-level model-based estimators are highly optimistic when the area homogeneity assumption is violated particularly in the urban areas of Bangladesh. The modified MSE estimators may be conservative when capturing the area variability but reduce the problem of MSE underestimation. In Chapter Four, we also showed how the naïve 2-level model-based estimators become worse in such situations via simulation study. The cost due to having underestimated MSE (showing incorrectly higher precision) may be much higher than the premium for getting conservative precision (slightly overestimated MSE). Thus proper care and appropriate investigation should be taken for considering a naïve 2-level model-based MSE estimator to obtain FGT estimates with correct higher precision.

Figure A6.1: Estimated HCR at upper poverty line (UPOVLN) and estimated MSE (EMSE) under 2-level and 3-level homoskedastic (HM) models by NPELL estimators via conditional and unconditional bootstraps

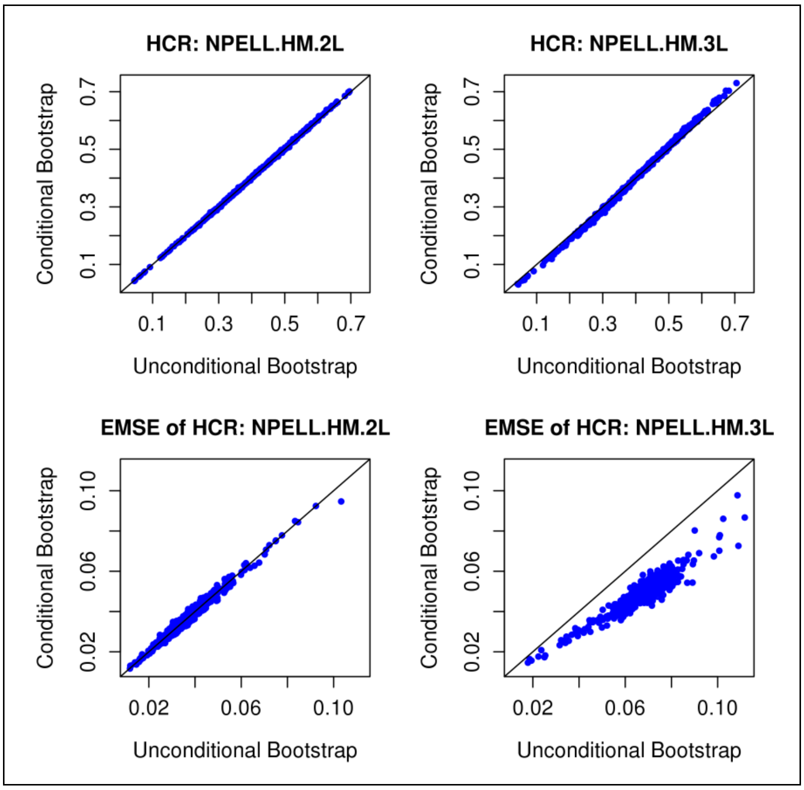


Figure A6.2: Estimated MSE (EMSE) of estimated PG and PS at upper poverty line (UPOVLN) under 2-level and 3-level models with homoskedastic (HM) and heteroskedastic (HT) level-one errors by PELL and modified PELL estimators

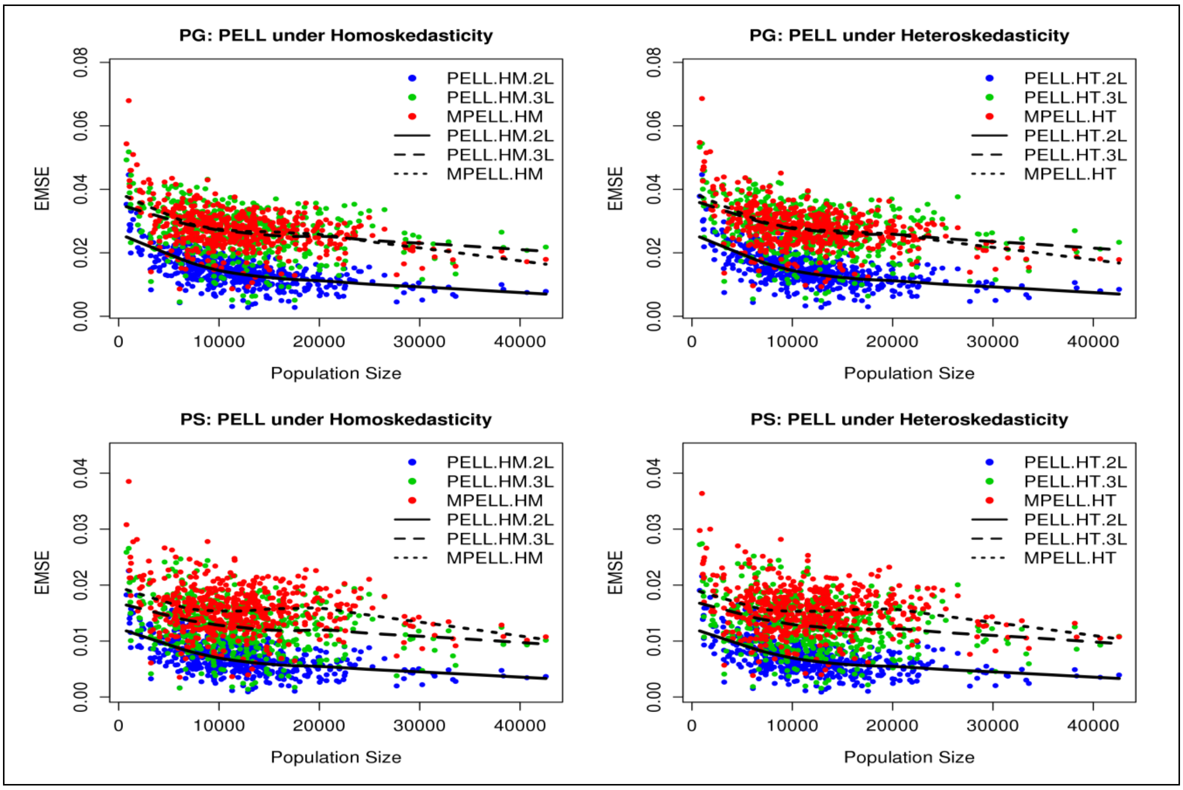


Figure A6.3: Estimated MSE (EMSE) of estimated PG and PS at upper poverty line (UPOVLN) under 2-level and 3-level homoskedastic models by PELL and NPELL with their modified estimators for sampled sub-districts with multiple clusters

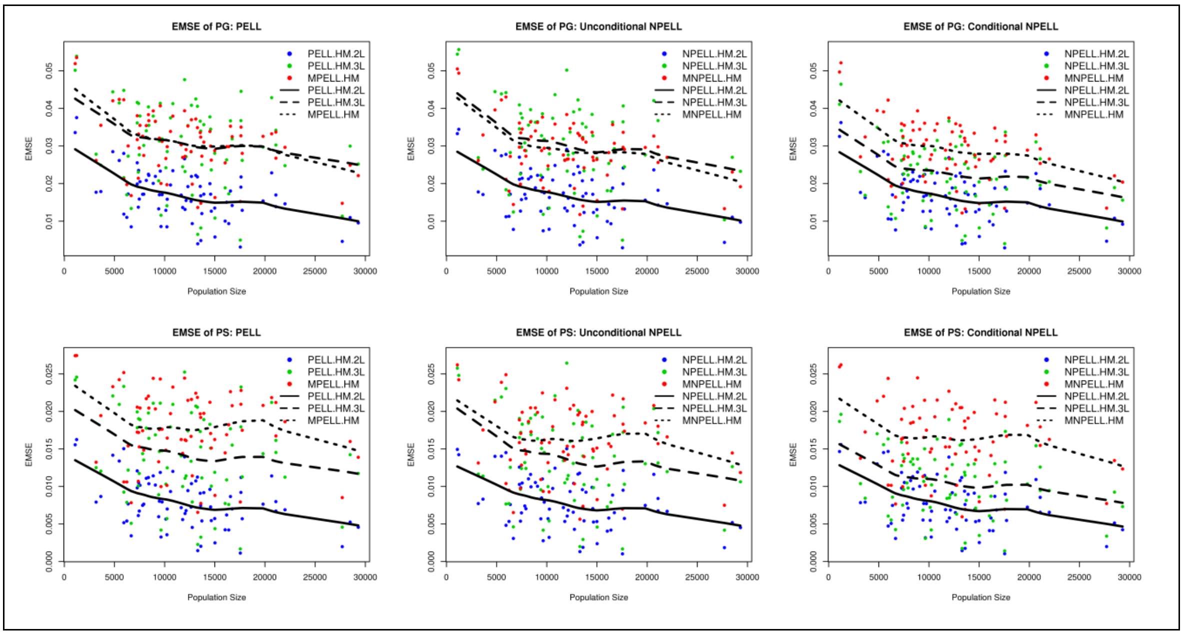


Figure A6.4: Estimated PG and PS at upper poverty line (UPOVLN) with their estimated MSE (EMSE) under 2-level and 3-level homoskedastic (HM) models by PELL, modified PELL, and MIX.ELL estimators

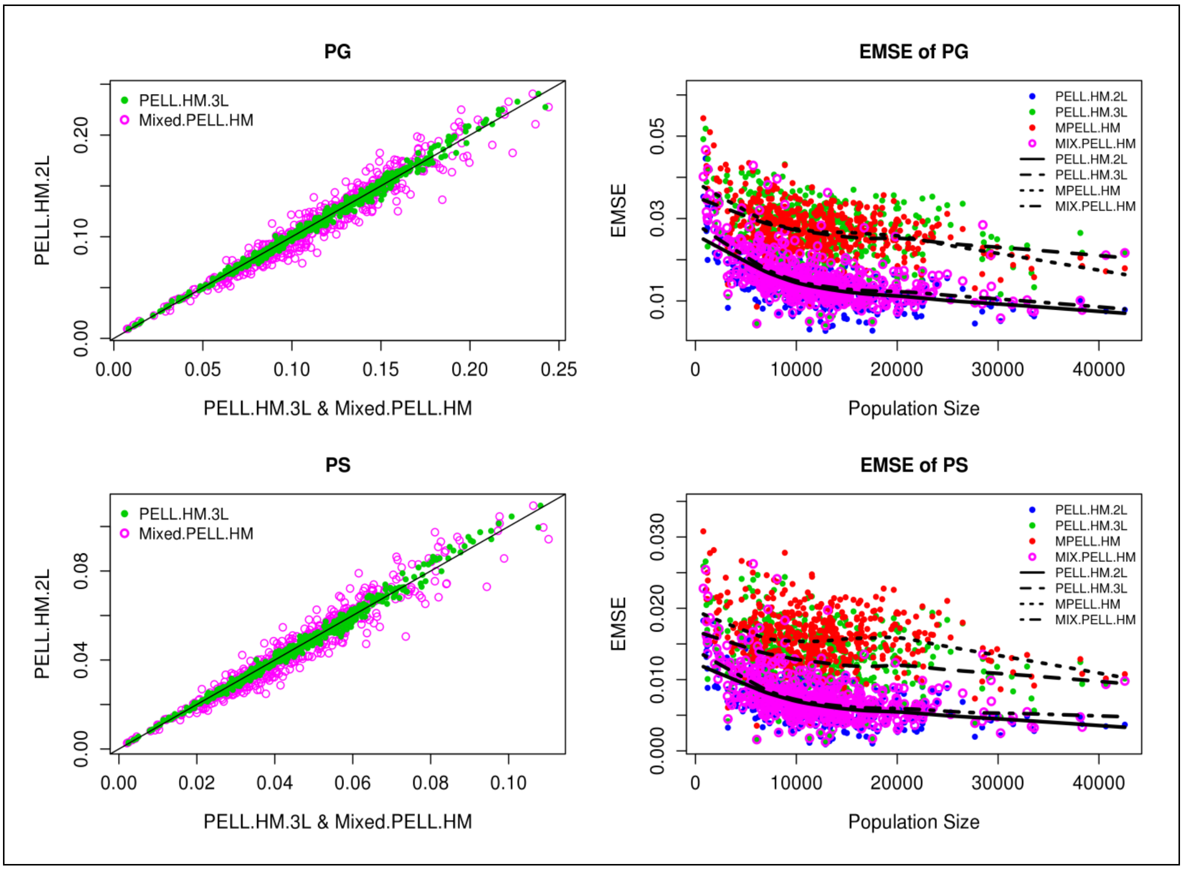


Table A6.1: Heteroskedasticity model using ELL approach under 2-level and 3-level heteroskedastic (HT) working models

Variables	2-level Model				3-level Model			
	α	*se*	*t*	*p*	α	*se*	*t*	*p*
Intercept	-4.77	0.14	-33.71	0.00	-4.77	0.14	-34.67	0.00
hhhprmed	-0.29	0.08	-3.56	0.00	-0.14	0.08	-1.74	0.08
ibuild_4	0.15	0.11	1.39	0.16	0.15	0.11	1.39	0.16
idiv_1	-0.28	0.10	-2.71	0.01	-0.33	0.10	-3.30	0.00
idiv_2	-0.04	0.08	-0.50	0.62	-0.11	0.07	-1.54	0.12
idiv_4	-0.11	0.09	-1.21	0.23	-0.22	0.09	-2.56	0.01
idiv_5	0.06	0.07	0.82	0.41	-0.08	0.07	-1.04	0.30
ilattr_1	-0.01	0.08	-0.15	0.88	0.03	0.08	0.40	0.69
literatep	0.26	0.10	2.58	0.01	0.12	0.10	1.22	0.22
ownaglnd	-0.08	0.06	-1.29	0.20	-0.06	0.06	-1.00	0.32
workst2p	-0.31	0.07	-4.48	0.00	-0.14	0.07	-2.16	0.03
% R^2	1.39				0.566			

Table A6.2: Variance components under 2-level and 3-level models with homoskedastic (HM) and heteroskedastic (HT) level-one errors by method of moments (MOM) and stratified MOM for different datasets

Skedasticity at HH-level	Cluster	Model	DF	$\hat{\sigma}^2_\varepsilon$	$\hat{\sigma}^2_u$	$\hat{\sigma}^2_\eta$	$\hat{\sigma}^2_u/\hat{\sigma}^2_e$	$\hat{\sigma}^2_\eta/\hat{\sigma}^2_e$
HM: MOM	Single	2L	33	0.1153	0.0238	-	16.00	-
	Multiple	2L	33	0.1091	0.0267	-	19.67	-
		3L	34	0.1091	0.0186	0.0082	13.69	6.03
	All	2L	33	0.1132	0.0253	-	18.28	-
		3L	34	0.1132	0.0192	0.0062	13.82	4.46
HT: MOM*	Single	2L	33	0.1162	0.0220	-	15.90	-
	Multiple	2L	33	0.1123	0.0268	-	19.25	-
		3L	34	0.1122	0.0187	0.0082	13.65	5.95
	All	2L	33	0.1137	0.0254	-	18.26	-
		3L	34	0.1176	0.0195	0.0059	14.01	4.25
HT: Stratified MOM (IGLS)*	Single	2L	33	0.1155	0.0225	-	16.31	-
	Multiple	2L	33	0.1123	0.0288	-	20.41	-
		3L	34	0.1122	0.0201	0.0085	15.19	6.04
	All	2L	33	0.1137	0.0258	-	18.50	-
		3L	34	0.1176	0.0200	0.0058	14.53	4.04

*Under heteroskedasticity, $\hat{\sigma}^2_\varepsilon = \hat{\sigma}^2_e - \hat{\sigma}^2_\eta - \hat{\sigma}^2_\eta$

Table A6.3: ELL and CD-type estimators of FGT poverty indicators and their MSE based on different bootstrap procedures under 2-level (2L) and 3-level (3L) models with homoskedastic (HM) and heteroskedastic (HT) level-one errors

Estimator	**Description**	**Parameter**
PELL.HM.2L	PB-based ELL estimator under 2L HM model	FGT & MSE
PELL.HM.3L	PB-based ELL estimator under 3L HM model	FGT & MSE
PELL.HT.2L	PB-based ELL estimator under 2L HT model	FGT & MSE
PELL.HT.3L	PB-based ELL estimator under 3L HT model	FGT & MSE
NPELL.HM.2L	NPB-based ELL estimator under 2L HM model	FGT & MSE
NPELL.HM.3L	NPB-based ELL estimator under 3L HM model	FGT & MSE
SPELL.HT.2L	SPB-based ELL estimator under 2L HT model	FGT & MSE
SPELL.HT.3L	SPB-based ELL estimator under 3L HT model	FGT & MSE
CDMC.HM.2L	MC & NPB-based CD estimator under 2L HM model	FGT & MSE
CDMC.HM.3L	MC & NPB-based CD estimator under 3L HM model	FGT & MSE
CDMC.HT.2L	MC & NPB-based CD estimator under 2L HT model	FGT & MSE
CDMC.HT.3L	MC & NPB-based CD estimator under 3L HT model	FGT & MSE
CDSM.HM.2L	SM & NPB-based CD estimator under 2L HM model	FGT & MSE
CDSM.HM.3L	SM & NPB-based CD estimator under 3L HM model	FGT & MSE
CDSM.HT.2L	SM & NPB-based CD estimator under 2L HT model	FGT & MSE
CDSM.HT.3L	SM & NPB-based CD estimator under 3L HT model	FGT & MSE
MPELL.HM	Modified PB-based ELL estimator under 2L HM model	MSE
MPELL.HT	Modified PB-based ELL estimator under 2L HT model	MSE
MNPELL.HM	Modified NPB-based ELL estimator under 2L HM model	MSE
MSPELL.HT	Modified SPB-based ELL estimator under 2L HT model	MSE
MCDMC.HM	Modified CDMC estimator under 2L HM model	MSE
MCDMC.HT	Modified CDMC estimator under 2L HT model	MSE
MCDSM.HM	Modified CDSM estimator under 2L HM model	MSE
MCDSM.HT	Modified CDSM estimator under 2L HT model	MSE

Table A6.4: Fitted regression models for Set-1, Set-2, Set-3, and Set-4 datasets under 2-level and 3-level homoskedastic (HM) models via method of moments

Variables	Set-1				Set-2				Set-3		Set-4	
	2-level		3-level		2-level		3-level		2-level		3-level	
	β	se	β	se	β	se	β	se	β	se	β	se
Intercept	6.95*	0.14	6.89*	0.15	7.02*	0.21	6.90*	0.25	6.83*	0.21	6.91*	0.24
electric	0.22*	0.02	0.22*	0.02	0.22*	0.02	0.22*	0.02	0.21*	0.02	0.24*	0.03
ilattr_1	0.11*	0.02	0.11*	0.02	0.11*	0.02	0.11*	0.02	0.11*	0.02	0.11*	0.02
ilattr_3	-0.13*	0.02	-0.13*	0.02	-0.06	0.04	-0.06	0.04	-0.14*	0.02	-0.06	0.05
iwater_1	0.15*	0.03	0.16*	0.03	0.13*	0.03	0.14*	0.04	0.19	0.17	0.15*	0.04
ibuild_3	0.11*	0.02	0.11*	0.02	0.12*	0.02	0.12*	0.02	0.25*	0.03	0.12*	0.02
rural	-0.09	0.05	-0.08	0.05	-0.15	0.09	-0.14	0.09	-	-	-	-
owner	0.12*	0.02	0.12*	0.02	0.13*	0.02	0.13*	0.02	0.09*	0.03	0.14*	0.02
ibuild_4	0.32*	0.03	0.32*	0.03	0.36*	0.04	0.35*	0.04	0.46**	0.21	0.34*	0.04
workst2p	-0.23*	0.01	-0.23*	0.01	-0.18*	0.02	-0.18*	0.02	-0.28*	0.02	-0.17*	0.02
workst3p	-0.14*	0.02	-0.15*	0.02	-0.14*	0.03	-0.14*	0.03	0.01	0.02	-0.11*	0.03
iincom_3	-0.07*	0.01	-0.07*	0.01	-0.06*	0.02	-0.05*	0.02	-0.08*	0.02	-0.06*	0.02
num_hh	-0.08*	0.00	-0.08*	0.00	-0.08*	0.00	-0.08*	0.00	-0.07*	0.00	-0.08*	0.00
num_hh2	0.01*	0.00	0.01*	0.00	0.01*	0.00	0.01*	0.00	0.01*	0.00	0.01*	0.00
hhhprmed	-0.16*	0.02	-0.16*	0.02	-0.14*	0.02	-0.13*	0.02	-0.09*	0.02	-0.16*	0.02
literatep	0.39*	0.02	0.39*	0.02	0.39*	0.03	0.39*	0.03	0.40*	0.02	0.38*	0.03
child5p	-0.53*	0.03	-0.53*	0.03	-0.52*	0.05	-0.52*	0.05	-0.52*	0.04	-0.52*	0.06
mhhsize	0.09**	0.03	0.10*	0.03	0.07	0.05	0.10	0.06	0.09	0.04	0.10	0.06
depratio	-0.38*	0.09	-0.38*	0.09	-0.39*	0.14	-0.41**	0.16	-0.23	0.12	-0.47*	0.16
paginc	0.07	0.07	0.03	0.07	0.12	0.11	0.07	0.12	-0.06	0.09	0.14	0.12
idiv_1	-0.04	0.04	-0.03	0.04	-0.08	0.06	-0.05	0.07	-0.04	0.04	-0.03	0.07
idiv_2	0.15*	0.03	0.15*	0.03	0.13**	0.05	0.11**	0.05	0.15*	0.05	0.13**	0.05
idiv_4	-0.17*	0.03	-0.16*	0.03	-0.25*	0.04	-0.24*	0.05	-0.08	0.04	-0.26*	0.05
idiv_5	-0.17*	0.03	-0.14*	0.03	-0.29*	0.05	-0.27*	0.06	-0.08**	0.03	-0.25*	0.05
femalep*rural‡	-0.19*	0.03	-0.19*	0.03	-0.10	0.08	-0.10	0.08	-0.19*	0.03	-0.03	0.04
ibuild_3*rural	0.15*	0.03	0.15*	0.04	0.21*	0.07	0.21*	0.07	-	-	-	-
owner*ibuild_4	0.10**	0.04	0.10*	0.04	0.10**	0.04	0.11**	0.04	-0.08	0.22	0.10**	0.04
rural*workst3p	0.19*	0.03	0.20*	0.03	0.19*	0.05	0.19*	0.05	-	-	-	-
rural*num_hh	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	-	-	-	-
rural*num_hh2	0.00**	0.00	0.00**	0.00	0.00*	0.00	0.00*	0.00	-	-	-	-
rural*hhhprmed	0.06**	0.03	0.06**	0.03	0.07	0.05	0.07	0.05	-	-	-	-
n	7428				3010				4653		2775	

* p<0.001, ** p<0.05, ‡ femalep for Set-3 and Set-4

CHAPTER SEVEN

7. Conclusion

This thesis describes and evaluates an MSE estimation method for the well-known World Bank's ELL methodology for poverty estimation that is robust to the violation of this methodology's area-homogeneity assumption. In addition, a flexible poverty estimation method that can account for household-level heteroskedasticity and based on the Chambers and Dunstan (1986) smearing approach to estimation of a finite population distribution function (referred to earlier as the CD-based method) was developed within the framework of the ELL methodology. This chapter concludes the thesis with a summary of the robust MSE estimation method, the CD-based method, the limitations of the proposed methods, and the results. We then discuss some promising directions for further research in poverty estimation.

7.1 Thesis Summary

One of the aims of the thesis is to examine how the ELL, EBP and MQ methods work in a realistic data set relevant to developing countries. The simulation study reported in **Chapter Three** was therefore based on the structure of actual census and survey datasets for Bangladesh. On the basis of the comparisons reported in this study we concluded that the standard ELL methodology based on 2-level working model performs better than the EBP and MQ methods in a realistic developing country scenario where between-cluster variability is higher than between-area variability. However, Gaussian-type confidence intervals based on MSE estimates obtained via ELL still undercover compared to the alternative methods. A 3-level model-based ELL method can be used to overcome this issue, but then has to overcome the problem that most sampled areas have only a single cluster, so fitting a 3-level model becomes difficult. As a consequence, a 2-level model is typically used instead of a 3-level model.

In **Chapter Three** we also developed guidelines for selecting a good multilevel model for use in small area estimation (SAE) in the context of poverty estimation. These included: (1) identification of the highest level in the data hierarchy where between-unit variation is maximized, (ii) If between-cluster variability is higher than the between-area variability, then one should fit a cluster based 2-level model and follow the standard ELL method (or EBP/MQ), reducing the between-area variability as much as possible by including area-level contextual variables, (iii) if between-area variability is higher than between-cluster variability, then one should fit an area-level 2-level model and follow EBP method (or ELL/MQ) after reducing cluster-level variability by including cluster-specific contextual variables in the model specification. However, this approach can be problematic, since fitting such an area-level model implies the availability of cluster-specific contextual variables from other sources, which may be impractical (e.g. spatial coordinates of a cluster are typically much less accessible than those of an area).

The results set out in **Chapter Three** indicate that the standard ELL method underestimates the true MSE of the poverty estimates that it produces, mainly because it ignores between-area variability in its model specification. These findings suggest the development of a modified version of 2-level model-based ELL method that accounts for the impact of between-area variability on the ELL-based MSE estimates. A modified ELL methodology that is robust to the presence of between-area variability is therefore developed in **Chapter Four**. The basic idea underpinning this modified methodology is an estimator of the MSE of area-specific means that is robust to ignoring level three in a 3-level model implemented under a simple stratification of the target small areas according to their population size. This modified MSE estimator shows improved performance for area-specific means, distribution functions, and poverty estimates compared with the standard 2-level model-based ELL method when there is between-area variability in the data. We also show that this modified method performs better than the optimistic and conservative version of ELL methods that have been proposed in the literature. Our findings in **Chapter Three** and the analysis set out in **Chapter Four** also provide evidence that both the EBP and MQ methods could also underestimate MSE if significant between-cluster variability is ignored in model specification.

In **Chapter Five** we consider the issue of dealing with level-one heteroskedasticity under an ELL approach to poverty estimation. Here we propose a semi-parametric approach to heteroskedasticity modelling based on stratified MOM estimation (STR) and show that the method is flexible and performs similarly to the ELL "alpha model" approach in this situation. We also demonstrate that the STR method can work better than the ELL approach when the heteroskedasticity function is non-monotone. If the HH-level error variances are estimated by the STR method and then used in the ELL bootstrap procedure, the approach can be considered as a purely non-parametric (rather than semi-parametric) extension of the ELL method.

Chapter Five also contains the next major contribution of the thesis, a small area poverty estimation method based on the smearing approach of CD (Chambers and Dunstan, 1986). This CD-based method is developed under both homoskedastic and heteroskedastic level-one errors for a 2-level working model similar to the one employed by the ELL method, and the proposed STR method is used to estimate heteroskedastic level-one error variances. Numerical experiments reported in Chapter Five show that the proposed method performs similarly or better than the ELL method.

In **Chapter Six**, the 2-level model-based ELL and CD-type MSE estimators are modified to account for between-area variability as well as heteroskedasticity at level-one, and then used in a poverty mapping exercise based on the Bangladesh datasets. These empirical results provide support for the applicability and the flexibility of the proposed CD-based method and the modified versions of both the ELL and CD-type estimators. They also show how the 2-level model-based ELL and CD-type estimators underestimate MSEs by ignoring the presence of between-area variability for those small areas with significant urban components and a large population size. Overall, the results in Chapter Five and Chapter Six confirm the viability of the proposed CD approach as a robust alternative to ELL methodology.

7.2 Future Research

This thesis has focused on the MSE estimation problem when ELL methodology is used for poverty estimation, and has developed an approach to overcoming this problem. However there are still theoretical and practical research questions that relate to the ELL and CD-based poverty estimation methods and their modifications. Some ideas for future research are discussed below.

Idea 1: Identification of guidelines for choosing an appropriate SAE method

The results from the numerical experiments reported in Chapter Three and based on the Bangladesh census and survey datasets suggest guidelines for choosing an appropriate model for poverty estimation. Depending on the hierarchical structure of a survey dataset, the general procedure is to find an appropriate 2-level model (either cluster or area level) and then apply one or more of the ELL, EBP, and MQ methods. Note that although fitting the MQ model does not require prior identification of the data hierarchy, the M-quantile coefficients and corresponding regression coefficients can be estimated at that hierarchy level (cluster or area) for which maximum variation is observed in a fitted multilevel model.

If a 3-level model is a better representation of a survey dataset compared with a 2-level model then 3-level model-based ELL, EBP and MQ methods should be better. In such cases the ELL method will be the easiest to implement due its flexibility. Model-based EBP and MQ methods based on a 3-level model require further research. Note that Diallo (2014) has developed a two-fold EBP method under skewed normal random errors.

The structure of the survey and census datasets should be an important consideration when selecting an SAE methodology for poverty estimation. In developing countries, the proportions of target small domains in the household level income and expenditure surveys tend to be small (20 - 50%) (Elbers and Van der Weide, 2014). In such situations, the EBP and MQ method may work better for the sampled areas if between-area variability dominates spatial variation in the data. In developed countries, the survey may include most of the target small areas and in this case the EBP and MQ methods perform consistently provided any intermediate level between the unit and target small area does not significantly contribute to the local variation of response variable. This implies that it is important to develop appropriate diagnostics for identifying significant between-cluster variability if the EBP or MQ methods are to be considered for poverty estimation.

Idea 2: Further development of poverty estimation based on a 3-level model

One of the key differences between fitting a single-level model and fitting a multilevel model is estimation of variance components. Nationally representative household surveys are usually designed with multiple purposes in mind and consequently tend to ignore lower-level administrative units. Estimating variance components defined by these lower-level administrative units (such as cluster and small areas) is then difficult, particularly in terms of efficiency. In this context, Haslett (2013) notes previous research by Münnich and Burgard (2012) and Haslett (2012) that consider estimation of variance components when a survey is not designed to produce small area estimates.

Unavailablity of multiple clusters in most of the sampled areas makes it difficult to separate area random effects from cluster random effects. In the Bangladesh poverty mapping study (BBS and UNWFP, 2004) more than 75% of the sampled areas had a single sample cluster, which creates issues in obtaining consistent and efficient estimates of higher-level variance components. Research is necessary to determine the optimum proportion of sampled areas with multiple clusters for which both the cluster and area-specific variance components estimators will be unbiased, stable, and consistent. This research should contribute significantly towards resolving the problem of whether to consider either a 2-level or a 3-level model for the ELL and EBP methods.

Reduction of between-area variation is vital when applying ELL methodology. The relevant question here is "how much of total variation due to between-area variability can be considered as negligible?" when a suitable test (e.g. a likelihood ratio test or LRT) shows that the area-level variance component is significant. Because of large overall sample size, it can be the case that a negligble (perhaps less than 0.5%) proportion of between-area variation could be identified as significant. Datta *et al.* (2011) have proposed a preliminary test estimator for the presence of a random effect at a specified level in a data hierarchy for an area-level model with a modest number of small areas. This estimator can be used to check the significance of cluster and area effects separately, but what is needed is the capacity to check the significance of the area effect given a significant cluster effect and vice versa. Thus a standard test or criterion is still required for identifying whether or not area-level random effects are redundant in a three-level model specification.

Idea 3: EBP and MQ methods under cluster-heterogeneity

In Chapter Three it was observed that both the EBP and the MQ methods underestimate the true simulation MSE when there is high between-cluster variability and negligible between-area variability. In Chapter Four, we show that if the 2nd level of a 3-level model is ignored then the 1st and 3rd variance components are biased, complementing similar results reported in Tranmer and Steel (2001a). This bias is one of the main reasons for underestimation of MSE by the EBP method. In contrast, the MQ method then behaves similarly to the 2-level model-based ELL method since the area-specific MQ coefficient is actually a cluster-specific MQ coefficient for the 75% of target areas containing a single cluster. This suggests that both the EBP and MQ methods should also be modified in order to deal with this underestimation of MSE issue, perhaps following a similar approach to the one used to develop the modified ELL methodology described in Chapter Four.

Idea 4: Practical issues in the use of unit-level SAE

The application of unit-level SAE methodology can be a real practical problem when the census dataset is huge, e.g. as in countries like China and India. In the Bangladesh study, a 5% Census data set was used. If a full census dataset was available and used, the computational effort would be very high for unit-level SAE methods. Using additional (or adequate) computational resources might be one way of handling this type of "big data”. In the usual case of limited resources, an alternative approach could be based on a partitioning of the whole country into several strata. Area-level SAE methods are another alternative; however, given the focus of poverty, and, by implication, small area income distributions, indicator-specific area-level models are then needed.

The use of sample census data (e.g. 5% Census data) raises the question of how big a proportion of a real population is required in order to make reliable inferences about the whole country. One possibility is that the poverty estimates based on sample Census data be scaled up to the whole census via an appropriate calibration method. Specification of how this calibration should work remains an open research question.

The ELL method is also used in nutrition mapping where anthropometric data on children, rather than HH incomes, are used (BBS and UNWFP, 2004; Fujii, 2010). However, in nutrition studies most unit-level models have low goodness of fit due to a lack of proper explanatory variables. In such studies, observed maximum variation is at child level, as one would expect, with these variations most likely due to human genetic variation. This lack of fit can be improved by inclusion of child-level demographic variables that are available in the survey dataset but are not available in the census dataset. That is, a better multilevel model can be developed using the survey dataset alone, but it is not possible to use this model in an ELL-type analysis due to lack of these child-level demographic variables in the census dataset. In such a situation, an area-level SAE approach may be better, particularly if relevant area-specific demographic and genetic variables can be estimated from other sources (e.g. contemporaneous genetic surveys).

Idea 5: Non-normality of random errors

The methodologies developed in Chapter Four and Chapter Five are based on an assumption of nested Gaussian errors. This assumption also held for the numerical experiments that were reported there. Some evidence for the validity of this assumption is the fact that for these simulations, which were based on the Bangladesh data sets; there were no significant differences in the MSE estimation performances of the ELL and CD-based methods that used parametric (i.e. based on a Gaussian assumption) as well as non-parametric bootstrap procedures. However, the question still remains about the performance of these approaches when errors are non-normally distributed.

Idea 6: Spatial correlation of random errors and distances among small areas

The proposed modified ELL method described in Chapter Four is based on the population sizes of the small areas and ignores their geospatial positions. Small areas close to the capital/port/metropolitan cities are highly interconnected in terms of communication facilities and employment opportunities, and can be expected to have similar poverty characteristics. This type of correlation is ignored when implemented the modified ELL procedure. In such cases, both the population size and the distance from neighboring small areas could be considered for the stratification used in the modified ELL methodology.

More generally, the ideas set out in the thesis have been developed based on the assumption of spatially uncorrelated random errors. Though such spatial correlation can be reduced by incorporating contextual variables in the regression model (Elbers *et al.*, 2008), there is still the possibility of spatial lag dependence (value of dependent variable in one area is affected by values of the dependent variable in contiguous areas) and spatial error dependence (the error term in one area is correlated with the error terms in nearby areas). Olivia *et al.* (2009) have examined such situation in a real data set relating to Shaanxi, China using exact measures of distance between each household. They show that ignoring the spatial error structure and the spatial lags at the modeling stage may lead to over-stated precision of local-level estimates of poverty. The main constraint in modeling such spatial effects is the availability of spatial data at household and cluster levels in the nationally representative household survey. However, the growing use of GPS (Global Positioning System) in recent household surveys (Gibson and McKenzie, 2007) implies that a spatial version of ELL methodology will soon be developed.

Idea 7: Heteroskedasticity at higher levels of the population hierarchy

The heteroskedasticity pattern in the Bangladesh dataset was monotone and negligible, so the ELL and CD-based methods led to very similar results. This raises the question of how different would be the results generated by these two methods if in fact the heteroskedasticity that is observed is very prevalent and non-monotone. Also, both approaches assume heteroskedastic random errors at unit-level only. But there may be heteroskedasticity at higher (cluster/group/area) levels (Gordon, 2012). If the variances of cluster-level (or area-level) random errors vary with the cluster size (or area size), then the ELL method needs to be modified in order to account for such heteroskedasticity. Note that this issue of general heteroskedasticity also arises for the EBP method proposed by Van der Weide (2014).

Idea 8: Application of SAE methods to inequality estimation

The ELL method is usually used to measure the FGT class of poverty measures - HCR $\left(F_{0i}\right)$, PG $\left(F_{1i}\right)$, and PS $\left(F_{2i}\right)$. Poverty severity F_{2i} is a combined measure of poverty and income/expenditure inequality. The Sen index (Sen, 1976) combines HCR and PG with the Gini index $\left(G_i\right)$ as $F_{S_i} = F_{0i}G_i^p + F_{1i}\left(1-G_i^p\right)$ where G_i^p is the Gini coefficient of inequality among the poor people belonging to area i. The Gini coefficient (Gini, 1912) is calculated as $G_i = 2\operatorname{cov}\left(\mathbf{y}_i, \mathbf{R}_i\right)/\bar{y}_i$ where $\mathbf{R}_i$ is a vector of standardized ranks of individuals or households in the income distribution (here 0 denotes poorest and 1 denotes richest). Several authors have modified the Sen index to use with a desired poverty inequality. A widely used modification is the Sen-Shorrocks-Thon index (Shorrocks, 1995), which is calculated as $F_{SST_i} = F_{0i}F_{1i}^{p}\left(1-\hat{G}_i^{p}\right)$ where F_{1i}^{p} is the PG calculated over poor people only, and $\hat{G}_i^{p}$ is the Gini coefficient of the PG variable over the whole population belonging to the i^{th} area. The Sen index and its modified versions have the capability to answer the three questions simultaneously: (i) Are there more poor people? (ii) Are the poor people more poor? and (iii) Is there higher inequality among the poor people? (Haughton and Khandker, 2009). Since these complex indicators are functions of the FGT measures, the ELL approach can be easily implemented in order to estimate them. Moreover, we expect that the proposed CD-based method and the modified versions of both the ELL and the CD-based methods also be capable of producing the estimates of such poverty and inequality measures.

The performances of the proposed methods could also be examined for other poverty inequality measures such as Theil-index (Theil, 1967) (a special case of the General Entropy class), the decile dispersion ratio (the ratio of the average income/consumption of the richest 10% of the population to that of the poorest 10%), share of income/consumption of the poorest 10/20% population (Haughton and Khandker, 2009), and the quintile share ratio denoted by S80/S20 (the ratio of total income/consumption shared by the top quintile to that of the bottom quintile).

Idea 9: Application of SAE methods to a Multidimensional Poverty Index (MPI)

Generally, poverty is defined by a scalar measure, say per capita income. However, people experience poverty due to other monetary and non-monetary attributes, such as lack of education, health, housing, empowerment, humiliation, employment, personal security, and so on. One indicator cannot represent these multiple aspects of poverty. A household may be not poor in monetary terms but could still be deprived with respect to lack of education, health, and security. Income poverty alone fails to capture these other aspects of deprivation.

Alkire and Foster (2009, 2011) created a multidimensional poverty index (MPI) by including a number of non-monetary deprivation criteria with the monetary FGT poverty measures. Alkire and Foster (hereafter referred as AF) proposed an identification and agrregation method by considering multiple response variabes (either continuous or integer) simultaneously with corresponding cut-off points. Similar to the traditional FGT measures, AF defined adjusted head count ratio, adjusted poverty gap, and adjusted FGT measures which are together referred as the adjusted FGT class of multidimentional poverty measures. A number of countries including Bhutan, Colombia, Mexico, the Philippines, and the state government of Minas Gerais, Brazil have produced their national MPI using the AF methodology.

Calculation of an MPI index for disaggregated administrative units represents a new field of application for SAE methods. Since income/consumption is the main monetary variable used in preparation of such an MPI index, the main task is still one of predicting the small area income distribution. In the identification phase of the AF method, prediction of the non-sample values of a set of dependent variables including income is still required. This implies the use of multivariate multilevel models for joint prediction of dependent variables in the SAE method. Schmid and Tzavidis (2015) have implemented an SAE method by fitting a generalized linear mixed model considering a multinomial dependent varable with five categories which is based on a two-dimensional representation of poverty: economic and social deprivation.

Idea 10: Time lag between Census and Survey

The time lag between the census data and the survey data used in ELL is an important issue in an ELL-based poverty mapping study. If the time lag is large, caution should be exercised when carrying out SAE (Haslett, 2013). Lanjouw and van der Wiede (2006) have suggested the use of approximately time invariant structural variables in the modeling process; Isidro, Haslett and Jones (2010, 2016) on the other hand suggest fitting the SAE model using contemporaneous survey and census data, and then updating the model fit using an ESPREE (extension of structure preserving estimation) method by using a set of margins from an up to date survey that make allowance for associated sampling error. However, the question now arises as to how one allows for sampling error if there is between-area variability in the recent survey but not in the base survey. In such situation, our modified ELL methodology may be able to capture this between-area variability.

An investigation of the proposal of Isidro, Haslett and Jones (2010, 2016) is also worthwhile. This proposes utilizing 2001 Bangladesh Population and Housing Census and 2000 Household Income and Expenditure Survey (HIES) datasets as the contemporaneous datasets, updates the resulting poverty estimates utilizing 2010 HIES, and then compares these updated estimates with the standard estimates based on 2011 Census and 2010 HIES. A further complication here is that the boundaries of the small areas of interest will almost certainly have changed over this time period.

BIBLIOGRAPHY

Alderman, H., Babita, M., Demombynes, G., Makhatha, N., & Özler, B. (2002). How low can you go? Combining census and survey data for mapping poverty in South Africa. *Journal of African Economies, 11*(2), 169-200.

Alkire, S., & Foster, J. (2009). Counting and Multidimensional Poverty Measures. *OPHI Working Paper 32*. Oxford: University of Oxford.

Alkire, S., & Foster, J. (2011). Understandings and misunderstandings of multidimensional poverty measurement. *The Journal of Economic Inequality, 9*(2), 289-314.

Altman, N. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician, 46*(3), 175-185.

Avery, M. (2013). Literature Review for Local Polynomial Regression. Retrieved from http://www4.ncsu.edu/~mravery/AveryReview2.pdf

Battese, G. E., Harter, R. M., & Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, *83*(401), 28-36.

BBS & UNWFP. (2004). *Local Estimation of Poverty and Malnutrition in Bangladesh.* Dhaka: Bangladesh Bureau of Statistics and United Nations World Food Programme.

BBS. (2001). *2001 Popultion Census: Priliminary Report.* Ministry Planning, Government of Bangladesh, Statistics Division. Dhaka: Bangladesh Bureau of Statistics. Retrieved from file:///C:/SKD/BPHC%202001/Population%20Census%202001%20-%20Preliminary%20Report.pdf

BBS. (2003). *Report of the Household Income & Expenditure Survey 2000.* Dhaka: Bangladesh Bureau of Statistics.

BBS. (2011). *2011 Population and Housing Census: Priliminary Results.* Ministry Planning, Government of Bangladesh, Statistics Division. Dhaka: Bangladesh Bureau of Statistics.

Benavent, R., & Morales, D. (2016). Multivariate Fay–Herriot models for small area estimation. *Computational Statistics & Data Analysis, 94*, 372-390.

Betti, G., Cheli, B., Lemmi, A., & Verma, V. (2006). Multidimensional and longitudinal poverty: an integrated fuzzy approach. In A. Lemmi A, & G. Betti, *Fuzzy set approach to multidimensional poverty measurement* (pp. 111–137). New York: Springer.

Betti, G., Neri, L., & Salvati, N. (2007, September 3-5). Comparing M-quantile and ELL methods for poverty mapping. *IASS Satellite Conference on Small Area Estimation.* Pisa: University of Pisa.

Bigman, D., & Fofack, H. (2000). Geographical targeting for poverty alleviation: Methodology and applications. *The World Bank Economic Review, 14*(1), pp. 129-146.

Bouhlel, I. (2013). Heteroscedasticity under the linear mixed model: Diagnostics and Treatment. Retrieved from http://math.unice.fr/~massonr/Stages/Rapport5.pdf

Breckling, J., & Chambers, R. (1988). M-quantiles. *Biometrika*, *75*(4), 761–71.

Brewer, K. R. (1963). Ratio estimation and finite populations: Some results deducible from the assumption of an underlying stochastic process. *Australian Journal of Statistics, 5*(3), 93-105.

Carpenter, J., Goldstein, H., & Rasbash, J. (1999). A non-parametric bootstrap for multilevel models. *Multilevel modelling newsletter, 11*(1), 2-5.

Chambers, R. L., & Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika, 73*(3), 597-604.

Chambers, R., & Clark, R. (2012). *An introduction to model-based survey sampling with applications.* Oxford: Oxford University Press.

Chambers, R., & Pratesi, M. (2013). Small area methodology in poverty mapping: an introductory overview. In G. Betti, & A. Lemmi, *Poverty and Social Exclusion: New Methods of Analysis* (p. p. 213). Oxon: Routledge.

Chambers, R., & Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika, 93*(2), 255-268.

Chambers, R., Chandra, H., & Tzavidis, N. (2011). On Bias-Robust Mean Squared Error Estimation for Pseudo-Linear Small Area Estimators. *Survey Methodology, 37*, 153-170.

Chambers, R., Chandra, H., Salvati, N., & Tzavidis, N. (2014). Outlier robust small area estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76*(1), 47-69.

Cheli, B., & Lemmi, A. (1995). A "Totally" fuzzy and relative approach to the multidimensional analysis of poverty. *Economic Notes, 25*, 115-134.

Cheng, M., Fan, J., & Marron, J. (1997). On automatic boundary corrections. *The Annals of Statistics, 25*(4), 1691-1708.

Chu, C., & Marron, J. (1991). Choosing a kernel regression estimator. *Statistical Science, 6*(4), 404-419.

Citro, C., Cohen, M., Kalton, G., & West, K. (1997). *Small-area estimates of school-age children in poverty: Interim report I-evaluation of 1993 county estimates for title I allocations.* Washington, DC: National Academy Press.

Cochran, W. (1977). *Sampling Techniques.* New York: Wiley.

Coelho, P., & Pereira, L. (2011). A spatial unit level model for small area estimation. *REVSTAT–Statistical Journal, 9*(2), 155-180.

Cressie, N. (1993). *Statistics for Spatial Data.* New York: Wiley Series in Probability and Statistics.

Cuong, N. V. , Truong, T. N., & Van Der Weide, R. (2010). Poverty and inequality maps in rural Vietnam: An application of small area estimation. *Asian Economic Journal, 24*(4), 355-390.

Datta, G. S. (2009). Model-based approach to small area estimation. In C. R. Rao, *Handbook of Statistics - Sample Surveys: Inference and Analysis* (Vol. 29B, pp. 251-288). Amsterdam: Elsevier.

Datta, G., & Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica, 10*(2), 613-628.

Datta, G., & Ravallion, M. (1993). Regional disparities, targeting, and poverty in India. In M. Lipton, & J. van der Gaag, *Including the Poor* (pp. 91-114). Washington, DC: The World Bank.

Datta, G., Hall, P., & Mandal, A. (2011). Model selection by testing for the presence of small-area effects, and application to area-level data. *Journal of the American Statistical Association, 106*(493), 362-374.

Demombynes, G., & Özler, B. (2005). Crime and local inequality in South Africa. *Journal of Development Economics, 76*(2), 265-292.

Demombynes, G., Elbers, C., Lanjouw, J., & Lanjouw, P. (2007). How Good a Map? Putting Small Area Estimation to the Test. *World Bank Policy Research Working Paper, 4155*. Washington, DC: The World Bank.

Diallo, M. (2014). Small Area Estimation under Skew-Normal Nested Error Models. *Doctoral dissertation*. Ottawa: Carleton University .

Diallo, M., & Rao, J. (2014). Small Area Estimation of Complex Parameters Under Unit-level Models with Skew-Normal Errors. *Joint Statistical Meetings, Section on Survey Research Methods* (pp. 970-984). Alexandria, VA: American Statistical Association.

Dunstan, R., & Chambers, R. (1989). Estimating distribution functions from survey data with limited benchmark information. *Australian Journal of Statistics, 31*(1), 1-11.

Efron, B., & Morris, C. (1973). Stein's estimation rule and its competitors - an empirical Bayes approach. *Journal of the American Statistical Association, 68*(341), 117-130.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap* (Vol. 57). New York: Chapman & Hall/CRC.

Elbers, C., & Van der Weide, R. (2014). Estimation of normal mixtures in a nested error model with an application to small area estimation of poverty and inequality. *World Bank Policy Research Working Paper, 6962*. Washington, DC: The World Bank.

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2002). Micro-level Estimation of Welfare. *World Bank Policy Research Working Papaer, 2911*. Washington DC: The World Bank.

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro-Level Estimation of Poverty and Inequality. *Econometrica, 71*(1), 355–364.

Elbers, C., Lanjouw, J. O., Lanjouw, P., & Leite, P. G. (2004). Poverty and inequality in Brazil: New estimates from combined PPV-PNAD. In *Inequality and economic development in Brazil.* Washington, DC: The World Bank.

Elbers, C., Lanjouw, P., & Leite, P. G. (2008). Brazil within Brazil : Testing the Poverty Map Methodology in Minas Gerais. *World Bank Policy Research Working Paper, 4513*. Washington, DC: The World Bank.

EURAREA, C. (2004). Project Reference Volume. Unpublished research report of EU research project IST-2000-26290. Retrieved from http://www.ons.gov.uk/ons/guide-method/method-quality/general-methodology/spatial-analysis-and-modelling/eurarea/downloads/project-reference-volume--volume-one--main-report.pdf

Fay, R. E., & Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association, 74*(366a), 269-277.

Feder, M., Nathan, G., & Pfeffermann, D. (2000). Multilevel modelling of complex survey longitudinal data with time varying random effects. *Survey Methodology, 26*(1), 53-66.

Ferretti, C., & Molina, I. (2012). Fast EB method for estimating complex poverty indicators in large populations. *Journal of the Indian Society of Agricultural Statistics, 66*, 105-120.

Foster, J., Greer, J., & Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica: Journal of the Econometric Society*, 761-766.

Fujii, T. (2004). Commune-level estimation of poverty measures and its application in Cambodia. *Research paper, ISSN 1810-2611*. Helsinki: World Institute for Development Economics Research, United Nations University.

Fujii, T. (2010). Micro-level estimation of child undernutrition indicators in Cambodia. *The World Bank Economic Review, 24*(3), 520-553.

Gershunskaya, J., Jiang, J., & Lahiri, P. (2009). Resampling methods in surveys. In C. R. Rao, *Handbook of Statistics- Sample Surveys: Inference and Analysis* (Vol. 29B, pp. 121-151). Amsterdam: Elsevier.

Ghosh, M., & Rao, J. N. (1994). Small area estimation: an appraisal. *Statistical science*, 55-76.

Ghosh, M., Nangia, N., & Kim, D. (1996). Estimation of median income of four-person families: a Bayesian time series approach. *Journal of the American Statistical Association*, *91*(436), 1423-1431.

Gibson, J., & McKenzie, D. (2007). Using global positioning systems in household surveys for better economics and better policy. *The World Bank Research Observer, 22*(2), 217-241.

Gini, C. (1912). Variability and Mutability, Contribution to The Study of Statistical Distribution and Relaitons. *Studi Economico-Giuricici della R*. Cagliari: Universita de Cagliari.

Giusti, C., Marchetti, S., Pratesi, M., & Tzavidis, N. (2011). Small Area Methodologies for Poverty Estimation: an Application to Italian Data. *Proceedings of the 58th World Statistical Congress. Session CPS020*, pp. 6234-6237. Dublin: International Statistical Institute.

Giusti, C., Pratesi, M., & Salvati, N. (2009). Estimation of poverty indicators: a comparison of small area methods at LAU1-2 level in Tuscany. *New Techniques and Technologies for Statistics (NTTS Conference), S3 Paper 1* (pp. 18-20). Brussels: Eurostat.

Gonzalez, M. E. (1973). Use and Evaluation of Synthetic Estimates. *Annual meeting of American Statistical Association. Social Statistics Section*, pp. 33-36. New York: American Statistical Association.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., & Santamaría, L. (2008a). Bootstrap mean squared error of a small-area EBLUP. *Journal of Statistical Computation and Simulation, 78*(5), 443-462.

González-Manteiga, W., Lombardía, M., Molina, I., Morales, D., & Santamaría, L. (2007). Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model. *Computational statistics & data analysis, 51*(5), 2720-2733.

González-Manteiga, W., Lombardía, M., Molina, I., Morales, D., & Santamaría, L. (2008b). Analytic and bootstrap approximations of prediction errors under a multivariate Fay–Herriot model. *Computational Statistics & Data Analysis, 52*(12), 5242-5252.

Gordon, R. (2012). *Applied statistics for the social and health sciences.* New York: Routledge.

Guadarrama, M., Molina, I., & Rao, J. (2015, April). A Comparison of Small Area Estimation Methods for Poverty Mapping. *UC3M Working Papers, Statistics and Economics, 15-05*. Madrid: Universidad Carlos III de Madrid.

Gujarati, D. N. (2003). *Basic Econometrics. (Chapter 10)* (4th ed.). New York: McGraw-Hill.

Hall, P., & Maiti, T. (2006a). On parametric bootstrap methods for small area prediction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68*(2), 221-238.

Hall, P., & Maiti, T. (2006b). Nonparametric estimation of mean-squared prediction error in nested-error regression models. *The Annals of Statistics, 34*(4), 1733-1750.

Hansen, N., Quintaes, V., Silva, D., Pessoa, D., & Silva, P. (2011). A Fay–Herriot Model for Estimating the Proportion of Households in Poverty in Brazilian Municipalities. *Proceedings of the 58th World Statistical Congress, Session CPS016* (pp. 4218-4223). Dublin: International Statistical Institute.

Harville, D. A. (1991). Comments. *Statistical Science, 6*, 35-39.

Haslett, S. (2012). Practical guidelines for design and analysis of sample surveys for small area estimation, Special Issue on Small Area Estimation. *Journal of the Indian Society of Agricultural Statistics, 66*(1), 203-212.

Haslett, S. (2013). Small area estimation of poverty using the ELL/PovMap method, and its alternatives. In G. Betti, & A. Lemmi, *Poverty and Social Exclusion: New Methods of Analysis* (p. 224). Oxon: Routledge.

Haslett, S., Isidro, M., & Jones, G. (2010). Comparison of survey regression techniques in the context of small area estimation of poverty. *Survey Methodology, 36*(2), 157-170.

Haslett, S., & Jones, G. (2005). *Estimation of Local Poverty in the Philippines.* Manila: Philippines National Statistics Co-ordination Board and The World Bank.

Haslett, S., & Jones, G. (2006). *Small area estimation of poverty, caloric intake and malnutrition in Nepal.* Kathmandu: Nepal Central Bureau of Statistics, United Nations World Food Programme, and The World Bank.

Haslett, S., & Jones, G. (2010). Small-area estimation of poverty: the aid industry standard and its alternatives. *Australian & New Zealand Journal of Statistics, 52*(4), 341-362.

Hastie, T., & Loader , C. (1993). Local regression: Automatic kernel carpentry. *Statistical Science*, 120-129.

Haughton, J., & Khandker, S. (2009). *Handbook on poverty and inequality.* Washington, DC: The World Bank.

Hayes, A., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior research methods, 39*(4), 709-722.

Healy, A. J., Hitsuchon, S., & Vajaragupta, Y. (2003). *Spatially disaggregated estimates of poverty and inequality in Thailand.* Cambridge, MA & Bangkok: Massachusetts Institute of Technology and Thailand Development Research Institute.

Henderson, C. R. (1950). Estimation of genetic parameters. *Annals of Mathematical Statistics, 21*(2), 309-310.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics, 31*(2), 423-447.

Henninger, N. (1998). *Mapping and Geographic Analysis of Human Welfare and Poverty - Review and Assessment.* Washington, D.C.: World Resources Institute.

Henninger, N., & Snel, M. (2002). *Where are the poor? Experiences with the development and use of poverty maps.* Washington, DC & Arendal: World Resources Institute & UNEP/GRID.

Hentschel, J., & Lanjouw, P. (1996). Poverty Profile. In *Ecuador Poverty Report* (pp. 53–91). Washington, DC: The World Bank.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association, 47*(260), 663-685.

Horvitz, D., & Thompson, D. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association, 47*, 663–685.

Huang, R., & Hidiroglou, M. (2003). Design consistent estimators for a mixed linear model on survey data. *Joint Statistical Meetings, Section on Survey Research Methods* (pp. 1897-1904). Alexandria, VA: American Statistical Association.

Isidro, M., Haslett, S., & Jones, G. (2010). Extended Structure-Preserving Estimation Method for Updating Small Area Estimates of Poverty. *Joint Statistical Meetings, Section on Government Statistics.* Alexandria, VA: American Statistical Association.

Isidro, M., Haslett, S., & Jones, G. (2016). Extended Structure Preserving Estimation (ESPREE) for updating small area estimates of poverty. *The Annals of Applied Statistics, 10*(1), 451-476.

James, W., & Stein, C. (1961). Estimation with quadratic loss. *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability. Volume 1: Contributions to the Theory of Statistics*, pp. 361-379. Berkeley, California: University of California Press.

Jiang, J. (2003). Empirical best prediction for small-area inference based on generalized linear mixed models. *Journal of Statistical Planning and Inference, 111*(1), 117-127.

Jiang, J., & Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test, 15*(1), 1-96.

Jiang, J., & Nguyen, T. (2012). Small area estimation via heteroscedastic nested-error regression. *Canadian Journal of Statistics, 40*(3), 588-603.

Jiang, J., Lahiri, P., & Wan, S. (2002). A unified jackknife theory for empirical best prediction with M-estimation. *The Annals of Statistics, 30*(6), 1782-1810.

Jones, G., & Haslett, S. J. (2004). Local estimation of poverty and malnutrition in Bangladesh: some practical and statistical issues. *2004 International Conference on Official Poverty Statistics: Methodology and Comparability*. Manila: National Statistical Coordination Board, Philippines.

Kackar, R., & Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association, 79,* 853–862.

Kuk, A. Y., & Welsh, A. H. (2001). Robust estimation for finite populations based on a working model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63*(2), 277-292.

Lanjouw, P., & van der Wiede, R. (2006). Determining changes in welfare distributions at the micro-level: Updating poverty maps. *Powerpoint presentation at the NSCB Workshop for the NSCB/World Bank Intercensal Updating Project*. Manila, Philipines.

Lehtonen, R., & Veijanen, A. (1998). Logistic generalized regression estimators. *Survey Methodology, 24*, 51–55.

Lehtonen, R., & Veijanen, A. (2009). Design-based methods of estimation for domains and small areas. In C. R. Rao, *Handbook of statistics: Sample surveys - Inference and Analysis* (Vol. 29B, pp. 219–249). Amsterdam: Elsevier.

Lehtonen, R., & Veijanen, A. (2011). Estimation of poverty indicators for domains. *NTTS 2011 - New Techniques and Technologies for Statistics. S14 Paper 2.* Brussels: Eurostat.

Lehtonen, R., & Vejanen, A. (2012). Small area poverty estimation by model calibration. *Journal of the Indian Society of Agricultural Statistics*, *66*(1), 125-133.

Lehtonen, R., Sarndal , C., & Veijanen, A. (2009). Model calibration and generalized regression estimation for domains and small areas. *SAE2009 Conference on Small Area Estimation.* Elche: University Miguel Hernández de Elche.

Levy, P. S. (1979). Small area estimation-synthetic and other procedures, 1968-1978. In J. Steinberg, *Synthetic Estimates for Small Areas: Statistical Workshop Papers* (pp. 4-19). Rockville, Maryland: NIDA.

Licona, G. H., & Jimenez, R. A. (2015). Small Area of Multidimensional Poverty at the Municipality Level in Mexico. *Conference of European Statisticians. Working Paper 24.* Geneva: Economic Commission for Europe, UN.

Lombardía, M. J., González-Manteiga, W., & Prada-Sánchez, J. M. (2005). Estimation of a finite population distribution function based on a linear model with unknown heteroscedastic errors. *The Canadian Journal of Statistics*, 181-200.

Maiti, T., & Slud, E. (2002). *Comparison of Small Area Models in SAIPE.* Retrieved from http://www2.math.umd.edu/~slud//myr.html/SAErv5B.pdf.

Marchetti, S., Tzavidis, N., & Pratesi, M. (2012). Non-parametric bootstrap mean squared error estimation for M-quantile estimators of small area averages, quantiles and poverty indicators. *Computational Statistics & Data Analysis, 56*(10), 2889-2902.

Marhuenda, Y., Molina, I., & Morales, D. (2013). Small area estimation with spatio-temporal Fay–Herriot models. *Computational Statistics & Data Analysis, 58,* 308-325.

Molina, I., & Rao, J. N. (2010). Small area estimation of poverty indicators. *Canadian Journal of Statistics, 38*(3), 369-385.

Molina, I., Nandram, B., & Rao, J. (2014). Small area estimation of general parameters with application to poverty indicators: A hierarchical Bayes approach. *The Annals of Applied Statistics, 8*(2), 852-885.

Molina, I., Salvati, N., & Pratesi , M. (2009). Bootstrap for estimating the MSE of the Spatial EBLUP. *Computational Statistics, 24*(3), 441-458.

Münnich, R., & Burgard, J. (2012). On the influence of sampling design on small area estimates. *Journal of the Indian Society of Agricultural Statistics, 66*(1), 145-156.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133-142.

National Center for Health Statistics. (1968). *Synthetic State Estimates of Disability.* Washington DC: P.H.S. Publications, Government Printing Office.

Oehlert, G. W. (1992). A note on the delta method. *The American Statistician, 46*(1), 27-29.

Olivia, S., Gibson, J., Smith, A., Rozelle, S., & Deng, X. (2009). An Empirical Evaluation of Poverty Mapping Methodology: Explicitly Spatial versus Implicitly Spatial Approach. *The Australian Agricultural & Resource Economics Society' Conference.* Cairns.

Opsomer, J., Claeskens, G., Ranalli, M., Kauermann, G., & Breidt, F. (2008). Non-parametric small area estimation using penalized spline regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70*(1), 265-286.

Petrucci, A., & Salvati, N. (2006). Small area estimation for spatial correlation in watershed erosion assessment. *Journal of agricultural, biological, and environmental statistics, 11*(2), 169-182.

Pfeffermann, D. (2002). Small Area Estimation-New Developments and Directions. *International Statistical Review, 70*(1), 125-143.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science, 28*(1), 40-68.

Pfeffermann, D., & Glickman, H. (2004). Mean square error approximation in small area estimation by use of parametric and nonparametric bootstrap. *Joint Statistical Meetings, Section on Survey Research Methods* (pp. 4167–4178). Alexandria, VA: American Statistical Association.

Pfeffermann, D., Skinner, C., Holmes, D., Goldstein, H., & Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society: series B (statistical methodology), 60*(1), 23-40.

Pinheiro, J., & Bates, D. (2006). *Mixed-effects models in S and S-PLUS.* New York: Springer.

Prasad, N. G., & Rao, J. N. (1990). The estimation of mean squared error of small area estimators. *Journal of the American Statistical Association, 85,* 163–171.

Pratesi, M., & Salvati, N. (2008). Small area estimation: the EBLUP estimator based on spatially correlated random area effects. *Statistical methods and applications, 17*(1), 113-141.

Quintano, C., Castellano, R., & Punzo, G. (2007). Estimating poverty in the Italian provinces using small area estimation models. *Metodoloski zvezki, 4*(1), 37-70.

Rahman, A. (2008, October). A Review of Small Area Estimation Problems and Methodological Developments. *NATSEM Discussion paper 66*. Canberra: The University of Canberra.

Rahvar, A., & Ardakani, M. (2011). Boundary effect correction in k-nearest-neighbor estimation. *Physical Review E, 83*, 051121.

Rao, J. N. (1999). Some recent advances in model-based small area estimation. *Survey Methodology, 25*(2), 175-186.

Rao, J. N. (2003). *Small Area Estimation.* New Jersey: John Wiley & Sons, Inc.

Rao, J., & Molina, I. (2015). *Small area estimation* (2nd ed.). New Jersey: John Wiley & Sons.

Rao, J., & Yu, M. (1994). Small-area estimation by combining time-series and cross-sectional data. *Canadian Journal of Statistics, 22*(4), 511-528.

Royall, R. M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika, 57*(2), 377-387.

Salvati, N., Chandra, H., & Chambers, R. (2012). Model-based Direct Estimation of Small-area Distributions. *Australian & New Zealand Journal of Statistics, 54*(1), 103-123.

Salvati, N., Pratesi, M., Tzavidis, N., & Chambers, R. (2008). Spatial M-quantile Models for Small Area Estimation,. *Centre for Statistical and Survey Methodology Working Paper 15-08*. Wollongong, NSW: University of Wollongong.

Salvati, N., Tzavidis, N., Pratesi, M., & Chambers, R. (2012). Small area estimation via M-quantile geographically weighted regression. *Test, 21*(1), 1-28.

Särndal, C. E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling* (1st ed.). New York: Springler.

Särndal, C. E., Swensson, B., & Wretman, J. (2003). *Model assisted survey sampling* (2nd ed.). New York: Springer.

Schmid, T., & Tzavidis, N. (2015, June 15-17). Applications of Small Area Methodologies for Estimating Deprivation Indicators. *Expert workshop on Local statistics for decision-making on well-being and vulnerability*. Livorno: University of Pisa.

Searle, S., Casella, G., & McCulloch, C. (1992). *Variance Components.* New York: Wiley.

Sen, A. (1976). Poverty: An Ordinal Approach to Measurement. *Econometrica, 44*(2), 219-231.

Shorrocks, A. (1995). Revisiting the Sen poverty index. *Econometrica, 63*, 1225-1230.

Singh, A., Mantel, H., & Thomas, B. (1994). Time series EBLUPs for small areas using survey data. *Survey Methodology, 20*(1), 33-43.

Sinha, S. K., & Rao, J. N. (2009). Robust small area estimation. *Canadian Journal of Statistics, 37*(3), 381-399.

Slud, E., & Maiti, T. (2006). Mean-squared error estimation in transformed Fay–Herriot models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68*(2), 239-257.

Snijders, T., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2nd ed.). London: Sage Publishers.

Souza, D., Brendolin, N., Quintaes, V., Pessoa, D., & Onel, S. (2015, August 3-5). Comparing small area estimation methods for poverty indicators in the municipalities of Brazilian State of Minas Gerais. *ISI Satellite Meeting on Small Area Estimation*. Santiago: Universidad Católica de Chile.

Tarozzi, A., & Deaton, A. (2009). Using census and survey data to estimate poverty and inequality for small areas. *The review of economics and statistics, 91*(4), 773-792.

Theil, H. (1967). *Economics and information theory* (Vol. 7). Amsterdam: North-Holland.

Tibshirani, R., & Wasserman, L. (2013). Nonparametric Regression. Retrieved from http://www.stat.cmu.edu/~larry/=sml/NonparRegression.pdf

Tranmer, M. (1999). Using census data to investigate the multilevel structure of local populations. *Doctoral dissertation*. Southampton: University of Southampton.

Tranmer, M., & Steel, D. G. (2001a). Ignoring a level in a multilevel model: evidence from UK census data. *Environment and Planning A, 33*(5), 941-948.

Tranmer, M., & Steel, D. G. (2001b). Using Local Census Data to Investigate Scale Effects. In N. J. Tate, & P. M. Atkinson , *Modelling scale in geographical information science* (pp. 105-122). New York: John Wiley & Sons, Ltd.

Tzavidis, N., & Chambers, R. (2007). Robust prediction of small area means and distributions. *Centre for Statistical and Survey Methodology Working Paper 2007-08*. Wollongong, NSW: University of Wollongong.

Tzavidis, N., Marchetti, S., & Chambers, R. (2010). Robust estimation of small-area means and quantiles. *Australian & New Zealand Journal of Statistics, 52*(2), 167-186.

Tzavidis, N., Marchetti, S., & Donbavand, S. (2013). Outlier robust semi-parametric small area methods for poverty estimation. In G. Betti, & A. Lemmi, *Poverty and Social Exclusion: New Methods of Analysis.* Oxon: Routledge.

Tzavidis, N., Salvati, N., Pratesi, M., & Chambers, R. (2008). M-quantile models with application to poverty mapping. *Statistical Methods and Applications, 17*(3), 393-411.

Valliant, R., Dorfman, A., & Royall, M. (2000). *Finite Population Sampling and Inference: A Prediction Approach.* New York: John Wiley & Sons.

Van der Weide, R. (2014). GLS Estimation and Empirical Bayes Prediction for Linear Mixed Models with Heteroskedasticity and Sampling Weights. *World Bank Policy Research Working Paper, 7028*. Washington, DC: The World Bank.

Verbeek, M. (2008). *A guide to modern econometrics.* West Sussex: John Wiley & Sons Ltd.

WB, BBS, & WFP. (2009). *Updating Poverty Maps of Bangladesh.* Washingtion, DC & Dhaka: The World Bank, Bangladesh Bureau of Statistics and World Food Program.

Wolter, K. M. (1985). *Introduction to Variance Estimation.* New York: Springer-Verlag.

World Bank. (2013). *Nepal small area estimation of poverty 2011. Volume 1.* Washington DC: The World Bank.

World Bank. (2015). Poverty and Equality Databank. Retrieved from http://povertydata.worldbank.org/poverty/country/BGD

Wu, C., & Sitter, R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association, 96*(453), 185-193.

APPENDIX A

Proofs of Theoretical Results

A.1 Variance Component Estimation: 2-level Homoskedastic (HM) Population Model

Suppose y_{ijk} indicates the value of the response variable Y for k^{th} household (HH) belonging to j^{th} cluster of i^{th} area. Assuming households (HHs) at level-one and clusters at level-two, a 2-level nested error regression model can be written as

$$\begin{aligned} & y_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} + u_{ij} + \varepsilon_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} + e_{ijk} \\ & i = 1,2,...,D;\ j = 1,2,...,C_i;\ k = 1,2,....,N_{ij} \end{aligned} \tag{1}$$

where, $u_{ij} \sim N\left(0, \sigma_{u(2)}^2\right)$, and $\varepsilon_{ijk} \sim N\left(0, \sigma_{\varepsilon(2)}^2\right)$ are identically and independently distributed cluster- and HH-level random errors. The sub-script (l) stands for a perfectly specified l-level model. Since the HH are nested within cluster, the covariance between two observations becomes $\sigma_{u(2)}^2 + \sigma_{\varepsilon(2)}^2$ if $ij = ij^/$ & $k = k^/$, $\sigma_{u(2)}^2$ $ij \neq ij^/$, and zero otherwise. Suppose a sample of size n is drawn randomly through a two-stage cluster sampling design covering all D small areas, where $C_s = \sum_{i \in s} C_{s_i}$ clusters are randomly selected at the first stage and n_{ij} $\left(i = 1,...,D \ \& \ j = 1,...,C_{s_i}\right)$ HHs are randomly selected at the second stage from the corresponding selected clusters. Fitting the model (1) to the sample data via least square

(LS) method, the estimated residuals can be utilized to calculate moment-based estimates of cluster- and HH-level random effects as $\hat{u}_{ij} = n_{ij}^{-1} \sum_{k \in s} \hat{e}_{ijk}$ and $\hat{\varepsilon}_{ijk} = \hat{e}_{ijk} - \hat{u}_{ij}$ where $\hat{e}_{ijk} = y_{ijk} - \hat{y}_{ijk}$. Under the 2-level model (1), we have

$$\text{Var}_2\left(\hat{\bar{e}}_{ij.}\right) = E_2\left(\hat{\bar{e}}_{ij.}^2\right) = n_{ij}^{-2}\left[\sum_{k \in s} E_2\left(\hat{e}_{ijk}^2\right) + \sum_{k \neq k' \in s} E_2\left(\hat{e}_{ijk}\hat{e}_{i'j'k'}\right)\right] = \sigma_{u(2)}^2 + n_{ij}^{-1}\sigma_{\varepsilon(2)}^2,$$

$$\text{Cov}_2\left(\hat{u}_{ij}, \hat{u}_{i'j'}\right) = \text{Cov}_2\left(\hat{\bar{e}}_{ij.}, \hat{\bar{e}}_{i'j'.}\right) = \left(\sigma_{u(2)}^2 + n_{ij}^{-1}\sigma_{\varepsilon(2)}^2\right) I\left\{ij = i'j'\right\}, \text{ and}$$

$$\begin{aligned}\text{Var}_2\left(\hat{\bar{e}}_{...}\right) &= \text{Var}_2\left(n^{-1}\sum_{ij \in s} n_{ij}\hat{\bar{e}}_{ij.}\right) = n^{-2}\left[\sum_{ij \in s} n_{ij}^2 \text{Var}_2\left(\hat{\bar{e}}_{ij.}\right) + \sum_{ij \neq ij' \in s} n_{ij} n_{ij'} \text{Cov}_2\left(\hat{\bar{e}}_{ij.}, \hat{\bar{e}}_{ij'.}\right)\right] \\ &= n^{-2}\left[\sum_{ij \in s} n_{ij}^2\left\{\sigma_{u(2)}^2 + n_{ij}^{-1}\sigma_{\varepsilon(2)}^2\right\}\right] = n^{-2}\sum_{ij \in s} n_{ij}^2 \sigma_{u(2)}^2 + n^{-1}\sigma_{\varepsilon(2)}^2\end{aligned}$$

where $E_l(.)$ stands for expectation under a perfectly specified l-level model . The HH- and cluster-level sample residual variances are expressed as

$$s^{(1)} = (n-1)^{-1}\sum_{ijk \in s}\left(\hat{e}_{ijk} - \hat{\bar{e}}_{...}\right)^2 = (n-1)^{-1}\left[\sum_{ijk \in s}\hat{e}_{ijk}^2 - n^{-1}\left(\sum_{ijk \in s}\hat{e}_{ijk}\right)^2\right] \text{ and}$$

$$s^{(2)} = \left(C_s - 1\right)^{-1}\sum_{ij \in s} n_{ij}\left(\hat{\bar{e}}_{ij.} - \hat{\bar{e}}_{...}\right)^2 = \left(C_s - 1\right)^{-1}\left[\sum_{ij \in s} n_{ij}\hat{\bar{e}}_{ij.}^2 - n^{-1}\left(\sum_{ijk \in s}\hat{e}_{ijk}\right)^2\right]$$

where $s^{(p)}; p = 1,..,l$ indicates the sample residual variances calculated at p^{th} level of an l-level model. Under the population model (1), the expectation of different terms associated with $s^{(1)}$ and $s^{(2)}$ can be expressed as

$$E_2\left(\sum_{ijk \in s}\hat{e}_{ijk}^2\right) = \sum_{ijk \in s} E_2\left(\hat{e}_{ijk}^2\right) = n\left(\sigma_{\varepsilon(2)}^2 + \sigma_{u(2)}^2\right),$$

$$E_2\left(\sum_{ijk\in s}\hat{e}_{ijk}\right)^2 = \sum_{ijk\in s}E_2\left(\hat{e}_{ijk}^2\right)+\sum_{ij=i'j'\in s}\sum_{k\neq k'}E_2\left(\hat{e}_{ijk}\hat{e}_{i'j'k'}\right)+\sum_{ij\neq i'j'\in s}\sum_{k\neq k'}E_2\left(\hat{e}_{ijk}\hat{e}_{i'j'k'}\right)=n\sigma_{\varepsilon(2)}^2+\sum_{ij\in s}n_j^2\sigma_{u(2)}^2$$

and $E_2\left(\sum_{ij\in s}n_{ij}\hat{\bar{e}}_{ij.}^2\right)=\sum_{ij\in s}n_{ij}E_2\left(\hat{\bar{e}}_{ij.}^2\right)=\sum_{ij\in s}n_{ij}\left(\sigma_{u(2)}^2+n_{ij}^{-1}\sigma_{\varepsilon(2)}^2\right)=C_s\sigma_{\varepsilon(2)}^2+n\sigma_{u(2)}^2$.

Then the expectations of $s^{(1)}$ and $s^{(2)}$ become

$$E_2\left(s^{(1)}\right)=\left(n-1\right)^{-1}\left[n\left(\sigma_{u(2)}^2+\sigma_{\varepsilon(2)}^2\right)-n^{-1}\left(\sigma_{u(2)}^2\sum_{ij\in s}n_{ij}^2+n\sigma_{\varepsilon(2)}^2\right)\right]=\sigma_{\varepsilon(2)}^2+\frac{\left(n-\bar{n}_0^{(2)}\right)}{\left(n-1\right)}\sigma_{u(2)}^2,$$

$$E_2\left(s^{(2)}\right)=\left(C_s-1\right)^{-1}\left[\left(n\sigma_{u(2)}^2+C_s\sigma_{\varepsilon(2)}^2\right)-n^{-1}\left(\sigma_u^2\sum_{ij\in s}n_{ij}^2+n\sigma_{\varepsilon(2)}^2\right)\right]=\sigma_{\varepsilon(2)}^2+\frac{n-\bar{n}_0^{(2)}}{C_s-1}\sigma_{u(2)}^2$$

where $\bar{n}_0^{(2)}=n^{-1}\sum_{ij\in s}n_{ij}^2$. In matrix form, the expectations are expressed as

$$E_2\begin{bmatrix}s^{(1)}\\ s^{(2)}\end{bmatrix}=\mathbf{A}_{(2)}\mathbf{\Lambda}_{(2)} \text{ with } \mathbf{A}_{(2)}=\begin{bmatrix}1 & \dfrac{n-\bar{n}_0^{(2)}}{n-1}\\ 1 & \dfrac{n-\bar{n}_0^{(2)}}{C_s-1}\end{bmatrix} \;\&\; \mathbf{\Lambda}_{(2)}=\begin{bmatrix}\sigma_{\varepsilon(2)}^2\\ \sigma_{u(2)}^2\end{bmatrix}.$$

Now the unbiased estimator of the variance components can be easily obtained if the coefficient matrix $\mathbf{A}_{(2)}$ is non-singular (Tranmer, 1999). The unbiased estimators of the variance components can be expressed as

$$\hat{\mathbf{\Lambda}}_{(2)}=\begin{bmatrix}\hat{\sigma}_{\varepsilon(2)}^2\\ \hat{\sigma}_{u(2)}^2\end{bmatrix}=\mathbf{A}_{(2)}^{-1}\begin{bmatrix}s^{(1)}\\ s^{(2)}\end{bmatrix}=\begin{bmatrix}\dfrac{n-1}{n-C_s} & -\dfrac{C_s-1}{n-C_s}\\ -\dfrac{n-1}{n-C_s}\dfrac{C_s-1}{n-\bar{n}_0^{(2)}} & \dfrac{n-1}{n-C_s}\dfrac{C_s-1}{n-\bar{n}_0^{(2)}}\end{bmatrix}\begin{bmatrix}s^{(1)}\\ s^{(2)}\end{bmatrix} \tag{2}$$

and their unbiasedness can be easily checked as below

$$E_2\left(\hat{\sigma}_{\varepsilon(2)}^2\right)=\frac{n-1}{n-C_s}E_2\left(s^{(1)}\right)-\frac{C_s-1}{n-C_s}E_2\left(s^{(2)}\right)=\sigma_{\varepsilon(2)}^2 \text{ and}$$

$$E_2\left(\hat{\sigma}_{u(2)}^2\right)=\frac{n-1}{n-C_s}\frac{C_s-1}{n-\bar{n}_0^{(2)}}\left[-E_2\left(s^{(1)}\right)+E_2\left(s^{(2)}\right)\right]=\sigma_{u(2)}^2.$$

A.2 Variance Component Estimation: 3-level Homoskedastic (HM) Population Model

Consider an additional area-level random effect in the 2-level model (1) to construct a 3-level nested error linear regression model for the response variable of interest as

$$\begin{aligned} y_{ijk} &= \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(3)} + \eta_i + u_{ij} + \varepsilon_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(3)} + e_{ijk} \\ i &= 1,2,...,D;\ j = 1,2,...,C_i;\ k = 1,2,....,N_{ij} \end{aligned} \tag{3}$$

where HH- $\left(\varepsilon_{ijk}\right)$, cluster- $\left(u_{ij}\right)$ and area-level $\left(\eta_i\right)$ random errors are identically and independently distributed as respectively $N\left(0,\sigma^2_{\varepsilon(3)}\right)$, $N\left(0,\sigma^2_{u(3)}\right)$, and $N\left(0,\sigma^2_{\eta(3)}\right)$. Now the covariance between two observations becomes $\sigma^2_{\eta(3)} + \sigma^2_{u(3)}$ if $ijk \neq ijk'$, $\sigma^2_{\eta(3)}$ if $ijk \neq ij'k'$, and zero otherwise. As previous area-, cluster-, and HH-level random effects are estimated as $\hat{\eta}_i = n_i^{-1}\sum_{jk\in s}\hat{e}_{ijk}$, $\hat{u}_{ij} = n_{ij}^{-1}\sum_{k\in s}\hat{e}_{ijk}$ and $\hat{\varepsilon}_{ijk} = \hat{e}_{ijk} - \hat{\eta}_i - \hat{u}_{ij}$ where $\hat{e}_{ijk} = y_{ijk} - \hat{y}_{ijk}$. Under the 3-level model (3), we have

$\mathrm{Var}_3\left(\hat{\bar{e}}_{ij.}\right) = n_{ij}^{-1}\sigma^2_{\varepsilon(3)} + \sigma^2_{u(3)} + \sigma^2_{\eta(3)}$, $\mathrm{Var}_3\left(\hat{\bar{e}}_{i..}\right) = n_i^{-1}\sigma^2_{\varepsilon(3)} + \sigma^2_{u(3)} n_i^{-2}\sum_{j\in s} n_{ij}^2 + \sigma^2_{\eta(3)}$ and

$\mathrm{Var}_3\left(\hat{\bar{e}}_{...}\right) = n^{-1}\sigma^2_{\varepsilon(3)} + n^{-2}\sum_{ij\in s} n_{ij}^2 \sigma^2_{u(3)} + n^{-2}\sum_{i\in s} n_i^2 \sigma^2_{\eta(3)}$.

The sample residual variances at area level is defined as $s^{(3)} = \left(D-1\right)^{-1}\sum_{i\in s} n_i\left(\hat{\bar{e}}_{i..} - \hat{\bar{e}}_{...}\right)^2$ $= \left(D-1\right)^{-1}\left[\sum_{i\in s} n_i \bar{e}^2_{i..} - n^{-1}\left(\sum_{ijk\in s}\hat{e}_{ijk}\right)^2\right]$. Now the expectations of the core terms under model (3)

$E_3\left(\sum_{ijk\in s}\hat{e}^2_{ijk}\right) = n\left(\sigma^2_{\varepsilon(3)} + \sigma^2_{u(3)} + \sigma^2_{\eta(3)}\right)$, $E_3\left(\sum_{ij\in s} n_{ij}\hat{\bar{e}}^2_{ij.}\right) = C_s\sigma^2_{\varepsilon(3)} + n\sigma^2_{u(3)} + n\sigma^2_{\eta(3)}$,

$E_3\left(\sum_{i\in s} n_i\hat{\bar{e}}^2_{i..}\right) = D\sigma^2_{\varepsilon(3)} + \sum_{i\in s}\bar{n}_{0i}^{(2)}\sigma^2_{u(3)} + n\sigma^2_{\eta(3)}$, $E_3\left(\sum_{ijk\in s}\hat{e}_{ijk}\right)^2 = n\sigma^2_{\varepsilon(3)} + \sigma^2_{u(3)}\sum_{ij\in s} n_{ij}^2 + \sigma^2_{\eta(3)}\sum_{i\in s} n_i^2$

lead to

$$E_3\left(\sum_{ijk\in s}\left(\hat{e}_{ijk}-\hat{\hat{e}}_{...}\right)^2\right)=\left(n-1\right)\sigma^2_{\varepsilon(3)}+\left(n-\overline{n}_0^{(2)}\right)\sigma^2_{u(3)}+\left(n-\overline{n}_0^{(3)}\right)\sigma^2_{\eta(3)},$$

$$E_3\left(\sum_{ij\in s}n_{ij}\left(\hat{\hat{e}}_{ij.}-\hat{\hat{e}}_{...}\right)^2\right)=\left(C_s-1\right)\sigma^2_{\varepsilon(3)}+\left(n-\overline{n}_0^{(2)}\right)\sigma^2_{u(3)}+\left(n-\overline{n}_0^{(3)}\right)\sigma^2_{\eta(3)}, \text{ and}$$

$$E_3\left(\sum_{i\in s}n_{i}\left(\hat{\hat{e}}_{i..}-\hat{\hat{e}}_{...}\right)^2\right)=\left(D-1\right)\sigma^2_{\varepsilon(3)}+\left(\sum_{i\in s}\overline{n}_{0i}^{(2)}-\overline{n}_0^{(2)}\right)\sigma^2_{u(3)}+\left(n-\overline{n}_0^{(3)}\right)\sigma^2_{\eta(3)}$$

with $\overline{n}_{0i}^{(2)}=n_i^{-1}\sum_{j\in s}n_{ij}^2$ and $\overline{n}_0^{(3)}=n^{-1}\sum_{i\in s}n_i^2$. After simplification, the expectations of the sample residual variances become

$$E_3\left(s^{(1)}\right)=\sigma^2_{\varepsilon(3)}+\left(n-1\right)^{-1}\left(n-\overline{n}_0^{(2)}\right)\sigma^2_{u(3)}+\left(n-1\right)^{-1}\left(n-\overline{n}_0^{(3)}\right)\sigma^2_{\eta(3)},$$

$$E_3\left(s^{(2)}\right)=\sigma^2_{\varepsilon(3)}+\left(C_s-1\right)^{-1}\left(n-\overline{n}_0^{(2)}\right)\sigma^2_{u(3)}+\left(C_s-1\right)^{-1}\left(n-\overline{n}_0^{(3)}\right)\sigma^2_{\eta(3)}, \text{ and}$$

$$E_3\left(s^{(3)}\right)=\sigma^2_{\varepsilon(3)}+\left(D-1\right)^{-1}\left(\sum_{i\in s}\overline{n}_{0i}^{(2)}-\overline{n}_0^{(2)}\right)\sigma^2_{u(3)}+\left(D-1\right)^{-1}\left(n-\overline{n}_0^{(3)}\right)\sigma^2_{\eta(3)}$$

which can be expressed in matrix form as below.

$$E\begin{bmatrix}s^{(1)}\\ s^{(2)}\\ s^{(3)}\end{bmatrix}=\mathbf{A}_{(3)}\mathbf{\Lambda}_{(3)} \text{ with } \mathbf{A}_{(3)}=\begin{bmatrix}1 & \left(n-1\right)^{-1}\left(n-\overline{n}_0^{(2)}\right) & \left(n-1\right)^{-1}\left(n-\overline{n}_0^{(3)}\right)\\ 1 & \left(C_s-1\right)^{-1}\left(n-\overline{n}_0^{(2)}\right) & \left(C_s-1\right)^{-1}\left(n-\overline{n}_0^{(3)}\right)\\ 1 & \left(D-1\right)^{-1}\left\{\sum_{i\in s}\overline{n}_{0i}^{(2)}-\overline{n}_0^{(2)}\right\} & \left(D-1\right)^{-1}\left(n-\overline{n}_0^{(3)}\right)\end{bmatrix} \ \& \ \mathbf{\Lambda}_{(3)}=\begin{bmatrix}\sigma^2_{\varepsilon(3)}\\ \sigma^2_{u(3)}\\ \sigma^2_{\eta(3)}\end{bmatrix}$$

The coefficient matrix $\mathbf{A}_{(3)}$ is comparable to Tranmer and Steel (2001b). If the coefficient matrix $\mathbf{A}_{(3)}$ is non-singular, the unbiased estimators of the variance components are easily obtained from

$$\hat{\mathbf{\Lambda}}_{(3)}=\begin{bmatrix}\hat{\sigma}^2_{\varepsilon(3)} & \hat{\sigma}^2_{u(3)} & \hat{\sigma}^2_{\eta(3)}\end{bmatrix}^T=\mathbf{A}_{(3)}^{-1}\begin{bmatrix}s^{(1)} & s^{(2)} & s^{(3)}\end{bmatrix}^T \tag{4}$$

with

$$\mathbf{A}_{(3)}^{-1}=\begin{bmatrix} \left(n-C_s\right)^{-1}\left(n-1\right) & -\left(n-C_s\right)^{-1}\left(C_s-1\right) & 0 \\ -\dfrac{n-1}{n-C_s}\dfrac{C_s-D}{n-\sum_{i\in s}\overline{n}_{0i}^{(2)}} & \dfrac{n-D}{n-\sum_{i\in s}\overline{n}_{0i}^{(2)}}\dfrac{C_s-1}{n-C_s} & -\dfrac{D-1}{n-\sum_{i\in s}\overline{n}_{0i}^{(2)}} \\ \dfrac{\left(n-1\right)\left\{\begin{array}{l}\left(\sum_{i\in s}\overline{n}_{0i}^{(2)}-\overline{n}_{0}^{(2)}\right)\left(C_s-1\right)\\ -\left(D-1\right)\left(n-\overline{n}_{0}^{(2)}\right)\end{array}\right\}}{\left(n-\overline{n}_{0}^{(2)}\right)\left(n-\sum_{i\in s}\overline{n}_{0i}^{(2)}\right)\left(n-C_s\right)} & -\dfrac{\left(C_s-1\right)\left\{\begin{array}{l}\left(\sum_{i\in s}\overline{n}_{0i}^{(2)}-\overline{n}_{0}^{(2)}\right)\left(n-1\right)\\ -\left(D-1\right)\left(n-\overline{n}_{0}^{(2)}\right)\end{array}\right\}}{\left(n-\overline{n}_{0}^{(3)}\right)\left(n-\sum_{i\in s}\overline{n}_{0i}^{(2)}\right)\left(n-C_s\right)} & \dfrac{D-1}{n-\sum_{i\in s}\overline{n}_{0i}^{(2)}}\dfrac{n-\overline{n}_{0}^{(2)}}{n-\overline{n}_{0}^{(3)}} \end{bmatrix}.$$

Unbiasedness of these estimators can be easily checked by taking the expectation under the true 3-level population model.

Suppose the first two variance components of the 3-level model are estimated using the estimators $\hat{\sigma}_{\varepsilon(2)}^2$ and $\hat{\sigma}_{u(2)}^2$ which are based on a perfectly specified 2-level model. The expectations of these incorrect estimators under the 3-level model become

$$\begin{aligned} E_3\left(\hat{\sigma}_{\varepsilon(2)}^2\right) &= \begin{bmatrix} \dfrac{n-1}{n-C_s}\left\{\sigma_{\varepsilon(3)}^2+\dfrac{n-\overline{n}_0^{(2)}}{n-1}\sigma_{u(3)}^2+\dfrac{n-\overline{n}_0^{(3)}}{n-1}\sigma_{\eta(3)}^2\right\} \\ -\dfrac{C_s-1}{n-C_s}\left\{\sigma_{\varepsilon(3)}^2+\dfrac{n-\overline{n}_0^{(2)}}{C_s-1}\sigma_{u(3)}^2+\dfrac{n-\overline{n}_0^{(3)}}{C_s-1}\sigma_{\eta(3)}^2\right\} \end{bmatrix} \\ &= \sigma_{\varepsilon(3)}^2\left\{\frac{n-1}{n-C_s}-\frac{C_s-1}{n-C_s}\right\}+\sigma_{u(3)}^2\left\{\frac{n-\overline{n}_0^{(2)}}{n-C_s}-\frac{n-\overline{n}_0^{(2)}}{n-C_s}\right\}-\sigma_{\eta(3)}^2\left\{\frac{n-\overline{n}_0^{(3)}}{n-C_s}-\frac{n-\overline{n}_0^{(3)}}{n-C_s}\right\} \\ &= \sigma_{\varepsilon(3)}^2\text{, and} \end{aligned}$$

$$\begin{aligned} E_2\left(\hat{\sigma}_{u(2)}^2\right) &= \frac{n-1}{n-C_s}\frac{C_s-1}{n-\overline{n}_0^{(2)}}\begin{bmatrix} -\left\{\sigma_{\varepsilon(3)}^2+\dfrac{n-\overline{n}_0^{(2)}}{n-1}\sigma_{u(3)}^2+\dfrac{n-\overline{n}_0^{(3)}}{n-1}\sigma_{\eta(3)}^2\right\} \\ +\left\{\sigma_{\varepsilon(3)}^2+\dfrac{n-\overline{n}_0^{(2)}}{C_s-1}\sigma_{u(3)}^2+\dfrac{n-\overline{n}_0^{(3)}}{C_s-1}\sigma_{\eta(3)}^2\right\} \end{bmatrix} \\ &= \frac{n-1}{n-C_s}\frac{C_s-1}{n-\overline{n}_0^{(2)}}\left[\left\{\frac{n-\overline{n}_0^{(2)}}{C_s-1}-\frac{n-\overline{n}_0^{(2)}}{n-1}\right\}\sigma_{u(3)}^2+\left\{\frac{n-\overline{n}_0^{(3)}}{C_s-1}-\frac{n-\overline{n}_0^{(3)}}{n-1}\right\}\sigma_{\eta(3)}^2\right] \end{aligned}$$

$$= \frac{1}{\left(n-C_s\right)\left(n-\bar{n}_0^{(2)}\right)}\left[\left(n-\bar{n}_0^{(2)}\right)\left(n-C_s\right)\sigma^2_{u(3)} + \left(n-\bar{n}_0^{(3)}\right)\left(n-C_s\right)\sigma^2_{\eta(3)}\right]$$
$$= \left[\sigma^2_{u(3)} + R\sigma^2_{\eta(3)}\right]$$

where $R = \frac{n-\bar{n}_0^{(3)}}{n-\bar{n}_0^{(2)}} < 1$. Thus the estimator of level-one variance component $\left(\hat{\sigma}^2_{\varepsilon(2)}\right)$ under a 2-level model provides unbiased estimate of the level-one variance component $\left(\sigma^2_{\varepsilon(3)}\right)$ of a 3-level model. However, the estimator of level-two variance component $\left(\hat{\sigma}^2_{u(2)}\right)$ under the 2-level model doesn't provide unbiased estimate of the sum of level-two and level-three variance components $\left(\sigma^2_{u(3)} + \sigma^2_{\eta(3)}\right)$ of the 3-level model.

A.3 Variance Component Estimation by Ignoring a Level of 3-level Homoskedastic (HM) Population Model

Suppose the level-three of the 3-level model (3) is ignored. Then the estimators of level-two and level-three variance components can be expressed as

$$\begin{bmatrix} \hat{\sigma}^2_{\varepsilon(3/3)} \\ \hat{\sigma}^2_{u(3/3)} \end{bmatrix} = \mathbf{A}_1^{-1}\begin{bmatrix} s^{(1)} \\ s^{(2)} \end{bmatrix} \text{ with } \mathbf{A}_1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 1 & \left(n-1\right)^{-1}\left(n-\bar{n}_0^{(2)}\right) \\ 1 & \left(C-1\right)^{-1}\left(n-\bar{n}_0^{(2)}\right) \end{bmatrix}$$

where $\mathbf{A}_1$ is a sub-set of $\mathbf{A}_{(3)}$ and the sub-script (3/3) is used for an estimator under a **3-**level model of which the **3**rd level is ignored. Under the 3-level model, expectations of the estimated variance components can be expressed as

$$E_3\begin{bmatrix} \hat{\sigma}^2_{\varepsilon(3/3)} \\ \hat{\sigma}^2_{u(3/3)} \end{bmatrix} = \mathbf{A}_1^{-1}E_3\begin{bmatrix} s^{(1)} \\ s^{(2)} \end{bmatrix} = \mathbf{A}_1^{-1}\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}\begin{bmatrix} \sigma^2_{\varepsilon} \\ \sigma^2_{u} \\ \sigma^2_{\eta} \end{bmatrix} = \mathbf{A}_1^{-1}\mathbf{A}_{1,2}\mathbf{\Lambda}_{(3)}$$

where $\mathbf{A}_{1,2}$ consists of first and second rows of $\mathbf{A}_{(3)}$. After the simplification, the expectations become

$$E_3\begin{bmatrix}\hat{\sigma}^2_{\varepsilon(3/3)}\\ \hat{\sigma}^2_{u(3/3)}\end{bmatrix}=\begin{bmatrix}\sigma^2_{\varepsilon(3)}\\ \sigma^2_{u(3)}+R\sigma^2_{\eta(3)}\end{bmatrix},\text{ with } R=\frac{n-\bar{n}_0^{(3)}}{n-\bar{n}_0^{(2)}}<1\ .$$

Suppose the level-two of the 3-level model (3) is ignored, then the estimators of the remaining variance components and their expectations under the true 3-level model become

$$\begin{bmatrix}\hat{\sigma}^2_{\varepsilon(3/2)}\\ \hat{\sigma}^2_{\eta(3/2)}\end{bmatrix}=\mathbf{A}_2^{-1}\begin{bmatrix}s^{(1)}\\ s^{(3)}\end{bmatrix}\text{ with }\mathbf{A}_2=\begin{bmatrix}a_{11} & a_{13}\\ a_{31} & a_{33}\end{bmatrix}=\begin{bmatrix}1 & (n-1)\left(n-\bar{n}_0^{(3)}\right)\\ 1 & (D-1)\left(n-\bar{n}_0^{(3)}\right)\end{bmatrix}$$

$$\text{and } E_3\begin{bmatrix}\hat{\sigma}^2_{\varepsilon(3/2)}\\ \hat{\sigma}^2_{\eta(3/2)}\end{bmatrix}=\mathbf{A}_2^{-1}E_3\begin{bmatrix}s^{(1)}\\ s^{(3)}\end{bmatrix}=\mathbf{A}_2^{-1}\begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{31} & a_{32} & a_{33}\end{bmatrix}\begin{bmatrix}\sigma^2_{\varepsilon}\\ \sigma^2_{u}\\ \sigma^2_{\eta}\end{bmatrix}=\mathbf{A}_2^{-1}\mathbf{A}_{1,3}\mathbf{\Lambda}_{(3)}$$

respectively where $\mathbf{A}_2$ and $\mathbf{A}_{1,3}$ are sub-set of matrix $\mathbf{A}_{(3)}$. The term $\mathbf{A}_2^{-1}\mathbf{A}_{1,3}$ becomes

$$\mathbf{A}_2^{-1}\mathbf{A}_{1,3}=\begin{bmatrix}\dfrac{n-1}{n-D} & \dfrac{-(D-1)}{n-D}\\ \dfrac{-(n-1)(D-1)}{\left(n-\bar{n}_0^{(3)}\right)(n-D)} & \dfrac{(n-1)(D-1)}{\left(n-\bar{n}_0^{(3)}\right)(n-D)}\end{bmatrix}\begin{bmatrix}1 & \dfrac{n-\bar{n}_0^{(2)}}{n-1} & \dfrac{n-\bar{n}_0^{(3)}}{n-1}\\ 1 & \dfrac{\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}}{D-1} & \dfrac{n-\bar{n}_0^{(3)}}{D-1}\end{bmatrix}$$

$$=\begin{bmatrix}1 & (n-D)^{-1}\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right) & 0\\ 0 & \dfrac{(n-1)\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)-(D-1)\left(n-\bar{n}_0^{(2)}\right)}{(n-D)\left(n-\bar{n}_0^{(3)}\right)} & 1\end{bmatrix}$$

$$\text{and hence } E_3\begin{bmatrix}\hat{\sigma}^2_{\varepsilon(3/2)}\\ \hat{\sigma}^2_{\eta(3/2)}\end{bmatrix}=\begin{bmatrix}\sigma^2_{\varepsilon(3)}+(n-D)^{-1}\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)\sigma^2_{u(3)}\\ \dfrac{(n-1)\left\{\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right\}-(D-1)\left(n-\bar{n}_0^{(2)}\right)}{(n-D)\left(n-\bar{n}_0^{(3)}\right)}\sigma^2_{u(3)}+\sigma^2_{\eta(3)}\end{bmatrix}.$$

An approximation following Tranmer and Steel (2001b) can be done as

$$\frac{(n-1)n\left\{n^{-1}\sum_{i\in s}\overline{n}_{0i}^{(2)}-n^{-1}\overline{n}_0^{(2)}\right\}-(D-1)n\left(1-n^{-1}\overline{n}_0^{(2)}\right)}{(n-D)n\left(1-n^{-1}\overline{n}_0^{(3)}\right)}\approx\frac{\sum_{i\in s}\overline{n}_{0i}^{(2)}-D}{n-D}=1-\frac{n-\sum_{i\in s}\overline{n}_{0i}^{(2)}}{n-D},$$

which leads to the following relationship

$$\frac{n-\sum_{i\in s}\overline{n}_{0i}^{(2)}}{n-D}\approx\frac{(n-D)\left(n-\overline{n}_0^{(3)}\right)-(n-1)\left\{\sum_{i\in s}\overline{n}_{0i}^{(2)}-\overline{n}_0^{(2)}\right\}+(D-1)\left(n-\overline{n}_0^{(2)}\right)}{(n-D)\left(n-\overline{n}_0^{(3)}\right)}$$

$$\Rightarrow\frac{n-\sum_{i\in s}\overline{n}_{0i}^{(2)}}{n-D}\approx\frac{(n-1)\left(n-\sum_{i\in s}\overline{n}_{0i}^{(2)}\right)+(n-D)\left(\overline{n}_0^{(2)}-\overline{n}_0^{(3)}\right)}{(n-D)\left(n-\overline{n}_0^{(3)}\right)}$$

$$\Rightarrow\frac{n-\sum_{i\in s}\overline{n}_{0i}^{(2)}}{n-D}\left[1-\frac{n-1}{n-\overline{n}_0^{(3)}}\right]\approx\frac{\overline{n}_0^{(2)}-\overline{n}_0^{(3)}}{n-\overline{n}_0^{(3)}}$$

$$\Rightarrow\frac{n-\sum_{i\in s}\overline{n}_{0i}^{(2)}}{n-D}\approx\frac{\left(\overline{n}_0^{(2)}-\overline{n}_0^{(3)}\right)}{\left(n-\overline{n}_0^{(3)}\right)}\frac{\left(n-\overline{n}_0^{(3)}\right)}{\left(1-\overline{n}_0^{(3)}\right)}=\frac{\left(\overline{n}_0^{(2)}-\overline{n}_0^{(3)}\right)}{\left(1-\overline{n}_0^{(3)}\right)}=1-\frac{1-\overline{n}_0^{(2)}}{1-\overline{n}_0^{(3)}}\approx1-\frac{\overline{n}_0^{(2)}}{\overline{n}_0^{(3)}}.$$

Thus, considering the relationship as

$$\frac{n-\sum_{i\in s}\overline{n}_{0i}^{(2)}}{n-D}\approx1-\frac{(n-1)\left\{\sum_{i\in s}\overline{n}_{0i}^{(2)}-\overline{n}_0^{(2)}\right\}-(D-1)\left(n-\overline{n}_0^{(2)}\right)}{(n-D)\left(n-\overline{n}_0^{(3)}\right)}\approx1-\frac{\overline{n}_0^{(2)}}{\overline{n}_0^{(3)}}$$

the expectations of the variance component estimators can be approximated as

$$E_3\begin{bmatrix}\hat{\sigma}^2_{\varepsilon(3/2)}\\ \hat{\sigma}^2_{\eta(3/2)}\end{bmatrix}\approx\begin{bmatrix}\sigma_\varepsilon^2+\left(1-\dfrac{\overline{n}_0^{(2)}}{\overline{n}_0^{(3)}}\right)\sigma^2_{u(3)}\\ \dfrac{\overline{n}_0^{(2)}}{\overline{n}_0^{(3)}}\sigma^2_{u(3)}+\sigma^2_{\eta(3)}\end{bmatrix}.$$

Equality holds only if the relationship is true in a particular case. These approximations are comparable to those of Tranmer and Steel (2001a).

Suppose the level-one of the 3-level model (3) is ignored, then the estimators of the remaining variance components and their expectations under the true 3-level model become

$$\begin{bmatrix} \hat{\sigma}^2_{u(3/1)} \\ \hat{\sigma}^2_{\eta(3/1)} \end{bmatrix} = \mathbf{A}_3^{-1} \begin{bmatrix} s^{(2)} \\ s^{(3)} \end{bmatrix} \text{ with } \mathbf{A}_3 = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} \dfrac{n-\bar{n}_0^{(2)}}{C_s - 1} & \dfrac{n-\bar{n}_0^{(3)}}{C_s - 1} \\ \dfrac{\sum_{i\in s} \bar{n}_{0i}^{(2)} - \bar{n}_0^{(2)}}{D-1} & \dfrac{n-\bar{n}_0^{(3)}}{D-1} \end{bmatrix} \text{ and}$$

$$E_3 \begin{bmatrix} \hat{\sigma}^2_{u(3/1)} \\ \hat{\sigma}^2_{\eta(3/1)} \end{bmatrix} = \mathbf{A}_3^{-1} E_3 \begin{bmatrix} s^{(2)} \\ s^{(3)} \end{bmatrix} = \mathbf{A}_3^{-1} \begin{bmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \sigma^2_\varepsilon \\ \sigma^2_u \\ \sigma^2_\eta \end{bmatrix} = \mathbf{A}_3^{-1} \mathbf{A}_{2,3} \mathbf{\Lambda}_{(3)}$$

respectively where $\mathbf{A}_3$ and $\mathbf{A}_{2,3}$ are sub-sets of matrix $\mathbf{A}_{(3)}$. The term $\mathbf{A}_3^{-1}\mathbf{A}_{2,3}$ becomes

$$\mathbf{A}_3^{-1}\mathbf{A}_{2,3} = \frac{1}{n - \sum_{i\in s} \bar{n}_{0i}^{(2)}} \begin{bmatrix} (C_s - 1) & -(D-1) \\ -\dfrac{(C_s-1)\left(\sum_{i\in s} \bar{n}_{0i}^{(2)} - \bar{n}_0^{(2)}\right)}{\left(n-\bar{n}_0^{(3)}\right)} & \dfrac{(D-1)\left(n-\bar{n}_0^{(2)}\right)}{\left(n-\bar{n}_0^{(3)}\right)} \end{bmatrix} \begin{bmatrix} 1 & \dfrac{n-\bar{n}_0^{(2)}}{C_s - 1} & \dfrac{n-\bar{n}_0^{(3)}}{C_s - 1} \\ 1 & \dfrac{\sum_{i\in s} \bar{n}_{0i}^{(2)} - \bar{n}_0^{(2)}}{(D-1)} & \dfrac{\left(n-\bar{n}_0^{(3)}\right)}{(D-1)} \end{bmatrix}$$

$$= \begin{bmatrix} \left(n - \sum_{i\in s} \bar{n}_{0i}^{(2)}\right)^{-1} (C_s - D) & 1 & 0 \\ \dfrac{(D-1)\left(n-\bar{n}_0^{(2)}\right) - (C_s-1)\left(\sum_{i\in s} \bar{n}_{0i}^{(2)} - \bar{n}_0^{(2)}\right)}{\left(n-\bar{n}_0^{(3)}\right)\left(n - \sum_{i\in s} \bar{n}_{0i}^{(2)}\right)} & 0 & 1 \end{bmatrix}$$

The complex term of the above matrix can be approximated as

$$\frac{(D-1)\left(n-\bar{n}_0^{(2)}\right) - (C_s-1)\left(\sum_{i\in s} \bar{n}_{0i}^{(2)} - \bar{n}_0^{(2)}\right)}{\left(n-\bar{n}_0^{(3)}\right)\left(n - \sum_{i\in s} \bar{n}_{0i}^{(2)}\right)} = \frac{(D-1)n\left(1-n^{-1}\bar{n}_0^{(2)}\right) - (C_s-1)n\left(n^{-1}\sum_{i\in s} \bar{n}_{0i}^{(2)} - n^{-1}\bar{n}_0^{(2)}\right)}{n\left(1-n^{-1}\bar{n}_0^{(3)}\right)\left(n - \sum_{i\in s} \bar{n}_{0i}^{(2)}\right)}$$

$$\approx \frac{(D-1)n - n(C_s-1)n^{-1}\sum_{i\in s} \bar{n}_{0i}^{(2)}}{n\left(n - \sum_{i\in s} \bar{n}_{0i}^{(2)}\right)} \approx \frac{D - C_s n^{-1} \sum_{i\in s} \bar{n}_{0i}^{(2)}}{n - \sum_{i\in s} \bar{n}_{0i}^{(2)}}.$$

Then the approximate expected values of the variance component estimators become

$$E_3\begin{bmatrix}\hat{\sigma}^2_{u(3/1)}\\ \hat{\sigma}^2_{\eta(3/1)}\end{bmatrix}\approx\begin{bmatrix}\left(n-\sum_{i\in s}\bar{n}^{(2)}_{0i}\right)^{-1}\left(C_s-D\right)\sigma^2_{\varepsilon(3)}+\sigma^2_{u(3)}\\ \left(n-\sum_{i\in s}\bar{n}^{(2)}_{0i}\right)^{-1}\left(D-C_s n^{-1}\sum_{i\in s}\bar{n}^{(2)}_{0i}\right)\sigma^2_{\varepsilon(3)}+\sigma^2_{\eta(3)}\end{bmatrix}$$

where the complex term of $E_3\left[\hat{\sigma}^2_{\eta(3/1)}\right]$ tends to zero if the overall sample size n becomes very large. Thus the estimator of area-level variance component can be approximately unbiased under this situation. These approximate results are comparable to those of Tranmer and Steel (2001a).

A.4 Scaling Raw Residuals under 2-level Homoskedastic (HM) Population Model

The empirical (sample) variances of HH- and cluster-level residuals under the 2-level population model (1) are respectively

$$\Sigma^{(1)}=\left(n-1\right)^{-1}\sum_{ijk\in s}\left(\hat{e}_{ijk}-\hat{\bar{e}}_{...}\right)^2=\left(n-1\right)^{-1}\left[\sum_{ijk\in s}\hat{e}^2_{ijk}-\frac{1}{n}\left(\sum_{ijk\in s}\hat{e}_{ijk}\right)^2\right]=s^{(1)}\neq\hat{\sigma}^2_{\varepsilon(2)}, \text{ and}$$

$$\begin{aligned}\Sigma^{(2)}&=\left(C_s-1\right)^{-1}\sum_{ij\in s}\left(\hat{\bar{e}}_{ij.}-\hat{\bar{e}}_{...}\right)^2=\left(C_s-1\right)^{-1}\left[\sum_{ij\in s}n_{ij}\left(\hat{\bar{e}}_{ij.}-\hat{\bar{e}}_{...}\right)^2-\sum_{ij\in s}\left(n_{ij}-1\right)\left(\hat{\bar{e}}_{ij.}-\hat{\bar{e}}_{...}\right)^2\right]\\&=s^{(2)}-\left(C_s-1\right)^{-1}\sum_{ij\in s}\left(n_{ij}-1\right)\left(\hat{\bar{e}}_{ij.}-\hat{\bar{e}}_{...}\right)^2\neq\hat{\sigma}^2_{u(2)}\end{aligned}$$

where C_s is the total number of sample clusters. The adjustments are needed in such a way that the empirical variances $\left(\Sigma^{(1)},\Sigma^{(2)}\right)$ in the bootstrap procedure are approximately equal to the estimated variance components $\left(\hat{\sigma}^2_{\varepsilon(2)},\hat{\sigma}^2_{u(2)}\right)$. It can be shown that, scaling the HH-level residuals by the term $\left(n-1\right)^{1/2}\sqrt{\sum_{ijk\in s}\hat{e}^2_{ijk}}\,\hat{\sigma}_{\varepsilon(2)}$ can approximate the estimated variation as

$$\Sigma^{*(1)}=\hat{\sigma}^2_{\varepsilon(2)}-\hat{\sigma}^2_{\varepsilon(2)}\left(\sum_{ijk\in s}\hat{e}^2_{ijk}\right)^{-1}n^{-1}\left(\sum_{ijk\in s}\hat{e}_{ijk}\right)^2\approx\hat{\sigma}^2_{\varepsilon(2)} \text{ when } \sum_{ijk\in s}\hat{e}_{ijk}\approx 0.$$

In the similar way, scaling the cluster-level residuals by the term $\left(C_s - 1\right)^{1/2} \sqrt{\sum_{ij\in s} \hat{\hat{e}}_{ij.}^2} \hat{\sigma}_{u(2)}$ leads to an approximation

$$\Sigma^{*(2)} = \hat{\sigma}_{u(2)}^2 - \frac{\hat{\sigma}_{u(2)}^2}{C_s \sum_{ij\in s} \hat{\hat{e}}_{ij.}^2} \left(\sum_{ij\in s} \hat{\hat{e}}_{ij.}\right)^2 \approx \hat{\sigma}_{u(2)}^2 \text{ when } \sum_{ij\in s} \hat{\hat{e}}_{ij.} \approx 0$$

that implies that $\Sigma^{*(2)} \approx \hat{\sigma}_{u(2)}^2$ if $\sum_{ij\in s} \hat{\hat{e}}_{ij.}$ tends to zero. The scaling procedure of the raw residuals under a random slope model is described by Carpenter *et al.* (1999).

A.5 Motivation of Modified ELL Methodology (MELL) under Misspecified Model with Homoskedastic (HM) Level-one errors

Under the 3-level population model (3), let assume m_{ijk} indicate the HH weight (say, HH size) of k^{th} HH belonging to the j^{th} cluster in the i^{th} area. Then the area-specific weighted mean of the response variable become $\bar{Y}_i = M_i^{-1} \sum_{j=1}^{C_i} \sum_{k=1}^{N_{ij}} m_{ijk} y_{ijk}$ where $M_i = \sum_j^{C_i} \sum_k^{N_{ij}} m_{ijk} = \sum_j^{C_i} M_{ij}$ and C_i are total number of individuals and clusters respectively in i^{th} area; $M_{ij} = \sum_k^{N_{ij}} m_{ijk}$ and N_{ij} are total number of individuals (population) and HHs respectively in j^{th} cluster of i^{th} area. Under the considered 3-level model (3), the variance of cluster-specific mean $\bar{Y}_{ij} = M_{ij}^{-1} \sum_{k=1}^{N_{ij}} m_{ijk} y_{ijk}$ can be expressed as

$$\begin{aligned}
\text{Var}_{(3)}\left(\bar{Y}_{ij}\right) &= M_{ij}^{-2}\left[\sum_{k=1}^{N_{ij}} m_{ijk}^2 \text{Var}_{(3)}\left(y_{ijk}\right) + \sum_{k\neq k'}^{N_{ij}} m_{ijk} m_{ijk'} \text{Cov}_3\left(y_{ijk}, y_{ijk'}\right)\right] \\
&= M_{ij}^{-2}\left[\left(\sigma_{\eta(3)}^2 + \sigma_{u(3)}^2 + \sigma_{\varepsilon(3)}^2\right)\sum_{k=1}^{N_{ij}} m_{ijk}^2 + \left(\sigma_{\eta(3)}^2 + \sigma_{u(3)}^2\right)\sum_{k\neq k'}^{N_{ij}} m_{ijk} m_{ijk'}\right] \\
&= M_{ij}^{-2}\left[\left(\sigma_{\eta(3)}^2 + \sigma_{u(3)}^2\right)\left\{\sum_{k=1}^{N_{ij}} m_{ijk}^2 + \sum_{k\neq k'}^{N_{ij}} m_{ijk} m_{ijk'}\right\} + \sigma_{\varepsilon(3)}^2 \sum_{k=1}^{N_{ij}} m_{ijk}^2\right] \\
&= \sigma_{\eta(3)}^2 + \sigma_{u(3)}^2 + \sigma_{\varepsilon(3)}^2 M_{ij}^{-2}\sum_{k=1}^{N_{ij}} m_{ijk}^2
\end{aligned}$$

where the suffix (3) stands for the 3-level model. The covariance between $\bar{Y}_{ij}$ and $\bar{Y}_{ij'}$

$$\text{Cov}_3\left(\bar{Y}_{ij}, \bar{Y}_{ij'}\right) = M_{ij}^{-1} M_{ij'}^{-1}\sum_{k}^{N_{ij}}\sum_{k'}^{N_{ij'}} m_{ijk} m_{ij'k'} \text{Cov}_3\left(y_{ijk}, y_{ij'k'}\right) = \sigma_{\eta(3)}^2 M_{ij}^{-1} M_{ij'}^{-1}\sum_{k}^{N_{ij}}\sum_{k'}^{N_{ij'}} m_{ijk} m_{ij'k'} = \sigma_{\eta(3)}^2$$

leads to the variance of i^{th} area mean $\bar{Y}_i$ as

$$\begin{aligned}
\text{Var}_{(3)}\left(\bar{Y}_i\right) &= M_i^{-2}\left[\sum_{j=1}^{C_i} M_{ij}^2\left\{\sigma_{\eta(3)}^2 + \sigma_{u(3)}^2 + \sigma_{\varepsilon(3)}^2 M_{ij}^{-2}\sum_{k=1}^{N_{ij}} m_{ijk}^2\right\} + \sigma_{\eta(3)}^2 \sum_{j\neq j'}^{C_i} M_{ij} M_{ij'}\right] \\
&= M_i^{-2}\left[\sigma_{\eta(3)}^2\left\{\sum_{j=1}^{C_i} M_{ij}^2 + \sum_{j\neq j'}^{C_i} M_{ij} M_{ij'}\right\} + \sigma_{u(3)}^2 \sum_{j=1}^{C_i} M_{ij}^2 + \sigma_{\varepsilon(3)}^2 \sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}} m_{ijk}^2\right] \\
&= \left[\sigma_{\eta(3)}^2 + \sigma_{u(3)}^2 M_i^{-2}\sum_{j=1}^{C_i} M_{ij}^2 + \sigma_{\varepsilon(3)}^2 M_i^{-2}\sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}} m_{ijk}^2\right] = \left[\sigma_{\eta(3)}^2 + \sigma_{u(3)}^2 \bar{m}_{Ui}^{(2)} + \sigma_{\varepsilon(3)}^2 \bar{m}_{Ui}^{(3)}\right]
\end{aligned}$$

where $\bar{m}_{Ui}^{(2)} = M_i^{-2}\sum_{j=1}^{C_i} M_{ij}^2 < 1$ and $\bar{m}_{Ui}^{(3)} = M_i^{-2}\sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}} m_{ijk}^2$. Then a plug-in estimator of $\text{Var}_{(3)}\left(\bar{Y}_i\right)$ can be considered as $\hat{\text{V}}_{(3)}\left(\bar{Y}_i\right) = \hat{\sigma}_{\eta(3)}^2 + \hat{\sigma}_{u(3)}^2 \bar{m}_{Ui}^{(2)} + \hat{\sigma}_{\varepsilon(3)}^2 \bar{m}_{Ui}^{(3)}$, where $\hat{\sigma}_{\eta(3)}^2$, $\hat{\sigma}_{u(3)}^2$, and $\hat{\sigma}_{\varepsilon(3)}^2$ are unbiased estimators of the variance components under the considered 3-level model. If HHs have same weights (say, $m_{ijk} = m$), the area-specific constant terms of $\text{Var}_{(3)}\left(\bar{Y}_i\right)$ and $\hat{\text{V}}_{(3)}\left(\bar{Y}_i\right)$ become $\bar{m}_{Ui}^{(2)} = N_i^{-2}\sum_{j=1}^{C_i} N_{ij}^2 = \bar{n}_{Ui}^{(2)}$ and $\bar{m}_{Ui}^{(3)} = N_i^{-1}$.

Under the 2-level model (1), $\mathrm{Var}_{(2)}\left(\bar{Y}_i\right)$ and its plug-in estimator can be written as

$$\mathrm{Var}_{(2)}\left(\bar{Y}_i\right)=\sigma^2_{u(2)}\bar{m}^{(2)}_{Ui}+\sigma^2_{\varepsilon(2)}\bar{m}^{(3)}_{Ui} \text{ and } \hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)=\hat{\sigma}^2_{u(2)}\bar{m}^{(2)}_{Ui}+\hat{\sigma}^2_{\varepsilon(2)}\bar{m}^{(3)}_{Ui}\,.$$

The expectations of the variance estimators $\hat{\mathrm{V}}_{(3)}\left(\bar{Y}_i\right)$ and $\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)$ under the true 3-level model are

$$E_3\left[\hat{\mathrm{V}}_{(3)}\left(\bar{Y}_i\right)\right]=\sigma^2_{\eta(3)}+\sigma^2_{u(3)}\bar{m}^{(2)}_{Ui}+\sigma^2_{\varepsilon(3)}\bar{m}^{(3)}_{Ui} \text{ and } E_3\left[\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)\right]=\left\{R\sigma^2_{\eta(3)}+\sigma^2_{u(3)}\right\}\bar{m}^{(2)}_{Ui}+\sigma^2_{\varepsilon(3)}\bar{m}^{(3)}_{Ui}$$

where $\bar{m}^{(2)}_{Ui}<1$ and $R<1$ lead to $E_3\left[\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)\right]<E_3\left[\hat{\mathrm{V}}_{(3)}\left(\bar{Y}_i\right)\right]$. Thus the expected value of the variance estimator $\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)$ under the true 3-level model always underestimates the true variation of area mean. We can adjust (or 'robustify') $\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)$ to make it unbiased or approximately unbiased under the 3-level model. This leads to an adjusted estimator as

$$\hat{\mathrm{V}}^{M}_{(2)}\left(\bar{Y}_i\right)=\left\{1\big/\bar{m}^{(2)}_{Ui}\;\hat{\sigma}^2_{\eta(3)}+\hat{\sigma}^2_{u(3)}\right\}\bar{m}^{(2)}_{Ui}+\hat{\sigma}^2_{\varepsilon(2)}\bar{m}^{(3)}_{Ui}$$

which is unbiased under the true 3-level model.

Based on this modified variance estimator of $\bar{Y}_i$ under the 2-level model, the second variance component estimator $\hat{\sigma}^2_{u(2)}$ in the ELL methodology is adjusted. The first adjustment factor is proposed based on the sample information as below:

$$k_1=\hat{\sigma}^{-2}_{u(2)}\left[\hat{\sigma}^2_{\eta(3)}D_s^{-1}\sum_{i=1}^{D_s}\left(1\big/\bar{m}^{(2)}_{si}\right)+\hat{\sigma}^2_{u(3)}\right]$$

where $\bar{m}^{(2)}_{si}=m_i^{-2}\sum_{j=1}^{C_{s_i}}m^2_{ij}$, D_s is number sampled areas, m_i is total members observed in i^{th} sampled area, m_{ij} is total members observed in j^{th} sampled cluster of i^{th} sampled area. The new adjusted cluster-level variance component will be then $\hat{\hat{\sigma}}^2_u=k_1\hat{\sigma}^2_{u(2)}$. The second adjustment factor and the corresponding corrected $\hat{\sigma}^2_{u(2)}$ are based on the area-specific population size (M_i) recorded in the census as respectively

$$k_2 = \hat{\sigma}^{-2}_{u(2)}\left[\hat{\sigma}^2_{\eta(3)} D^{-1}\sum_{i=1}^{D}\left(1\big/\bar{m}^{(2)}_{Ui}\right) + \hat{\sigma}^2_{u(3)}\right] \text{ and } \hat{\hat{\sigma}}^2_u = k_2\hat{\sigma}^2_{u(2)}.$$

The third adjustment factor is based on the stratification of all the small areas based on their population size. The adjustment factors and the corresponding corrected $\hat{\sigma}^2_{u(2)}$s can be expressed as

$$k_{3(h)} = \hat{\sigma}^{-2}_{u(2)}\left[\hat{\sigma}^2_{\eta(3)} D^{-1}_{(h)}\sum_{i=1}^{D_{(h)}}\left(1\big/\bar{m}^{(2)}_{Ui}\right) + \hat{\sigma}^2_{u(3)}\right] \text{ and } \hat{\hat{\sigma}}^2_{u(h)} = k_{3(h)}\hat{\sigma}^2_{u(2)};\ \ h = 1,\ldots,H.$$

When the HH weights are equal, the adjustment factors becomes

$$k_1 = \hat{\sigma}^{-2}_{u(2)}\left[\hat{\sigma}^2_{\eta(3)} D^{-1}_{s}\sum_{i=1}^{D_s}\left(1\big/\bar{n}^{(2)}_{si}\right) + \hat{\sigma}^2_{u(3)}\right],\ k_2 = \hat{\sigma}^{-2}_{u(2)}\left[\hat{\sigma}^2_{\eta(3)} D^{-1}\sum_{i=1}^{D}\left(1\big/\bar{n}^{(2)}_{Ui}\right) + \hat{\sigma}^2_{u(3)}\right], \text{ and}$$

$$k_{3(h)} = \hat{\sigma}^{-2}_{u(2)}\left[\hat{\sigma}^2_{\eta(3)} D^{-1}_{(h)}\sum_{i=1}^{D_{(h)}}\left(1\big/\bar{n}^{(2)}_{Ui}\right) + \hat{\sigma}^2_{u(3)}\right]$$

where $\bar{n}^{(2)}_{si} = n_i^{-2}\sum_{j=1}^{C_{s_i}} n^2_{ij}$, $n_i = m_i/m$ is total HHs observed in i^{th} sampled area, $n_{ij} = m_{ij}/m$ is total HHs observed in j^{th} sampled cluster of i^{th} sampled area.

A.6 Variance Component Estimation: 2-level Model with Heteroskedastic (HT) Level-one Errors

Now consider the HH-level random errors are heteroskedastic (HT) and distributed as $\varepsilon_{ijk} \sim N\left(0, \sigma^2_{\varepsilon(2),ijk}\right)$ where $\sigma^2_{\varepsilon(2),ijk} = \varphi\left(x_{ijk}\right)$ is assumed to be an unknown function of explanatory variables. Then the 2-level nested error regression model can be considered as

$$y_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} + u_{ij} + \varepsilon_{ijk} \text{ with } u_{ij} \sim N\left(0, \sigma_{u(2)}^{2(ht)}\right) \text{ \& } \varepsilon_{ijk} \sim N\left(0, \sigma_{\varepsilon(2),ijk}^2\right). \tag{5}$$

As previous cluster- and HH-level random effects can be estimated at first by the moment estimators $\hat{u}_{ij} = n_{ij}^{-1} \sum_{k \in s} \hat{e}_{ijk}$ and $\hat{\varepsilon}_{ijk} = \hat{e}_{ijk} - \hat{u}_{ij}$ respectively using the LS residuals $\hat{e}_{ijk} = y_{ijk} - \hat{y}_{ijk}$. Under the considered HT 2-level model,

$$\text{Var}_2\left(\hat{\bar{e}}_{ij.}\right) = E_2\left(\hat{\bar{e}}_{ij.}^2\right) = n_{ij}^{-2}\left[\sum_{k \in s} E_2\left(\hat{e}_{ijk}^2\right) + \sum_{k \neq k' \in s} E_2\left(\hat{e}_{ijk}\hat{e}_{i'j'k'}\right)\right] = \sigma_u^{2(ht)} + n_{ij}^{-2} \sum_{k \in s} \sigma_{\varepsilon(2),ijk}^2 ,$$

$$\text{Cov}_2\left(\hat{u}_{ij}, \hat{u}_{i'j'}\right) = \text{Cov}_2\left(\hat{\bar{e}}_{ij.}, \hat{\bar{e}}_{i'j'.}\right) = \left(\sigma_u^{2(ht)} + n_{ij}^{-2} \sum_{k \in s} \sigma_{\varepsilon(2),ijk}^2\right) I\left\{ij = i^/ j^/\right\},$$

$$\begin{aligned}\text{Var}_2\left(\hat{\bar{e}}_{...}\right) &= \text{Var}_2\left(n^{-1} \sum_{ij} n_{ij} \hat{\bar{e}}_{ij.}\right) = n^{-2}\left[\sum_{ij} n_{ij}^2 \text{Var}_2\left(\hat{\bar{e}}_{ij.}\right) + \sum_{ij \neq ij^/} n_{ij} n_{ij^/} \text{Cov}_2\left(\hat{\bar{e}}_{ij.}, \hat{\bar{e}}_{ij^/.}\right)\right] \\ &= n^{-2}\left[\sum_{ij} n_{ij}^2 \left\{\sigma_u^{2(ht)} + n_{ij}^{-2} \sum_{k \in s} \sigma_{\varepsilon(2),ijk}^2\right\}\right] = n^{-2} \sum_{ij} n_{ij}^2 \sigma_u^{2(ht)} + n^{-2} \sum_{ijk \in s} \sigma_{\varepsilon(2),ijk}^2 ,\end{aligned}$$

$$E_2\left(\sum_{ijk \in s} \hat{e}_{ijk}^2\right) = \sum_{ijk \in s} E_2\left(\hat{e}_{ijk}^2\right) = n\sigma_u^{2(ht)} + \sum_{ijk \in s} \sigma_{\varepsilon(2),ijk}^2 ,$$

$$E_2\left(\sum_{ijk \in s} \hat{e}_{ijk}\right)^2 = \sum_{ijk \in s} E_2\left(\hat{e}_{ijk}^2\right) + \sum_{ij = i'j' \in s} \sum_{k \neq k'} E_2\left(\hat{e}_{ijk}\hat{e}_{i'j'k'}\right) + \sum_{ij \neq i'j' \in s} \sum_{k \neq k'} E_2\left(\hat{e}_{ijk}\hat{e}_{i'j'k'}\right) = \sigma_u^{2(ht)} \sum_{ij \in s} n_{ij}^2 + \sum_{ijk \in s} \sigma_{\varepsilon(2),ijk}^2$$

and $$E_2\left(\sum_{ij \in s} n_{ij} \hat{\bar{e}}_{ij.}^2\right) = \sum_{ij \in s} n_{ij} E_2\left(\hat{\bar{e}}_{ij.}^2\right) = \sum_{ij \in s} n_{ij}\left(\sigma_u^{2(ht)} + n_{ij}^{-2} \sum_{k} \sigma_{\varepsilon(2),ijk}^2\right) = n\sigma_u^{2(ht)} + \sum_{ij \in s} n_{ij}^{-1} \sum_{k} \sigma_{\varepsilon(2),ijk}^2 .$$

Using these expected values, the expectation of the HH- and cluster-level sample residual variances $s^{(1)}$ and $s^{(2)}$ can be expressed as

$$E_2\left(s^{(1)}\right) = \frac{1}{n-1}\left[n\sigma_u^{2(ht)} + \sum_{ij \in s} \sigma_{\varepsilon(2),ijk}^2 - \frac{1}{n}\left(\sigma_u^{2(ht)} \sum_{ij \in s} n_{ij}^2 + \sum_{ijk \in s} \sigma_{\varepsilon(2),ijk}^2\right)\right] = \frac{n - \bar{n}_0^{(2)}}{n-1} \sigma_u^{2(ht)} + \frac{1}{n} \sum_{ijk \in s} \sigma_{\varepsilon(2),ijk}^2$$

and $E_2\left(s^{(2)}\right)=\frac{1}{C_s-1}\left[\left(n\sigma_u^{2(ht)}+\sum_{ij\in s}\frac{1}{n_{ij}}\sum_k\sigma^2_{\varepsilon(2),ijk}\right)-\frac{1}{n}\left(\sigma_u^2\sum_{ij\in s}n_{ij}^{2(ht)}+\sum_{ijk\in s}\sigma^2_{\varepsilon(2),ijk}\right)\right]$

$$=\frac{n-\bar{n}_0^{(2)}}{C_s-1}\sigma_u^{2(ht)}+\frac{1}{C_s-1}\sum_{ij\in s}\left(\frac{1}{n_{ij}}-\frac{1}{n}\right)\sum_k\sigma^2_{\varepsilon(2),ijk}.$$

Under the assumption of known HH-level variances $\sigma^2_{\varepsilon(2),ijk}$, the expectation of sample residual variances can be expressed in terms of true variance components as

$$E_2\begin{bmatrix}s^{(1)}\\ s^{(2)}\end{bmatrix}=\mathbf{A}_4\mathbf{\Lambda}_4 \text{ with } \mathbf{A}_4=\begin{bmatrix}n^{-1}\sum_{ijk\in s}\sigma^2_{\varepsilon(2),ijk} & \frac{n-\bar{n}_0^{(2)}}{n-1}\\ \frac{1}{C_s-1}\sum_{ij\in s}\left(\frac{1}{n_{ij}}-\frac{1}{n}\right)\sum_k\sigma^2_{\varepsilon(2),ijk} & \frac{n-\bar{n}_0^{(2)}}{C_s-1}\end{bmatrix} \text{ and } \mathbf{\Lambda}_4=\begin{bmatrix}1\\ \sigma_u^{2(ht)}\end{bmatrix}$$

which will be the same as HM case shown in **Appendix A.1** if $\sigma^2_{\varepsilon(2),ijk}=\sigma^2_{\varepsilon(2)}$. Now the estimator of $\mathbf{\Lambda}_4$ can be easily obtained as $\hat{\mathbf{\Lambda}}_4=\mathbf{A}_4^{-1}\mathbf{S}$ if the coefficient matrix $\mathbf{A}_4$ is non-singular. Putting $\xi_1=\sum_{ijk\in s}\sigma^2_{\varepsilon(2),ijk}$ and $\xi_2=\sum_{ij\in s}\left(\frac{1}{n_{ij}}-\frac{1}{n}\right)\sum_k\sigma^2_{\varepsilon(2),ijk}$, the determinant and conjugate of $\mathbf{A}_4$ can be expressed as

$$\left|A_4\right|=\frac{n-\bar{n}_0^{(2)}}{C_s-1}\frac{\xi_1}{n}-\frac{n-\bar{n}_0^{(2)}}{n-1}\frac{\xi_2}{C_s-1}=\frac{n-\bar{n}_0^{(2)}}{C_s-1}\left[\frac{\xi_1}{n}-\frac{\xi_2}{n-1}\right] \ \& \ \mathrm{Con}\left(A_4\right)=\begin{bmatrix}\frac{n-\bar{n}_0^{(2)}}{C_s-1} & -\frac{n-\bar{n}_0^{(2)}}{n-1}\\ -\frac{\xi_2}{C_s-1} & \frac{\xi_1}{n}\end{bmatrix}$$

which lead to $A_4^{-1}=\left[\frac{\xi_1}{n}-\frac{\xi_2}{n-1}\right]^{-1}\begin{bmatrix}1 & -\frac{C_s-1}{n-1}\\ -\frac{\xi_2}{n-\bar{n}_0^{(2)}} & \frac{C_s-1}{n-\bar{n}_0^{(2)}}\frac{\xi_1}{n}\end{bmatrix}$.

Then the estimator of $\sigma_{u(2)}^{2(ht)}$ is

$$\hat{\sigma}_{u(2)}^{2(ht)} = \left[\frac{\xi_1}{n} - \frac{\xi_2}{n-1}\right]^{-1} \left\{ -\frac{\xi_2}{n-\bar{n}_0^{(2)}} s^{(1)} + \frac{C_s - 1}{n-\bar{n}_0^{(2)}} \frac{\xi_1}{n} s^{(2)} \right\}.$$

To obtain a simplest form of $\hat{\sigma}_{u(2)}^{2(ht)}$, the HH-level sample residual variance $s^{(1)}$ can be expressed as a function of $\sum_{ijk\in s} \sigma_{\varepsilon(2),ijk}^2$ and $s^{(2)}$ from $A_4^{-1}S$ as below.

$$1 = \left[\frac{\xi_1}{n} - \frac{\xi_2}{n-1}\right]^{-1} \left\{ s^{(1)} - \frac{C_s - 1}{n-1} s^{(2)} \right\} \Rightarrow s^{(1)} = \left[\frac{\xi_1}{n} - \frac{\xi_2}{n-1}\right] + \frac{C_s - 1}{n-1} s^{(2)}$$

Then the estimator $\hat{\sigma}_{u(2)}^{2(ht)}$ can be expressed as

$$\begin{aligned}
\hat{\sigma}_{u(2)}^{2(ht)} &= \left[\frac{\xi_1}{n} - \frac{\xi_2}{n-1}\right]^{-1} \left\{ -\frac{\xi_2}{n-\bar{n}_0^{(2)}} \left\{ \left[\frac{\xi_1}{n} - \frac{\xi_2}{n-1}\right] + \frac{C_s - 1}{n-1} s^{(2)} \right\} + \frac{C_s - 1}{n-\bar{n}_0^{(2)}} \frac{\xi_1}{n} s^{(2)} \right\} \\
&= \left[\frac{\xi_1}{n} - \frac{\xi_2}{n-1}\right]^{-1} \left\{ -\frac{\xi_2}{n-\bar{n}_0^{(2)}} \left[\frac{\xi_1}{n} - \frac{\xi_2}{n-1}\right] - \frac{\xi_2}{n-\bar{n}_0^{(2)}} \frac{C_s - 1}{n-1} s^{(2)} + \frac{C_s - 1}{n-\bar{n}_0^{(2)}} \frac{\xi_1}{n} s^{(2)} \right\} \\
&= \left[\frac{\xi_1}{n} - \frac{\xi_2}{n-1}\right]^{-1} \left\{ -\frac{\xi_2}{n-\bar{n}_0^{(2)}} \left[\frac{\xi_1}{n} - \frac{\xi_2}{n-1}\right] + \frac{C_s - 1}{n-\bar{n}_0^{(2)}} s^{(2)} \left[\frac{\xi_1}{n} - \frac{\xi_2}{n-1}\right] \right\} \\
&= \frac{1}{n-\bar{n}_0^{(2)}} \left[(C_s - 1) s^{(2)} - \xi_2 \right] = \frac{1}{n-\bar{n}_0^{(2)}} \left[(C_s - 1) s^{(2)} - \sum_{ij\in s} \left(\frac{1}{n_{ij}} - \frac{1}{n} \right) \sum_k \sigma_{\varepsilon(2),ijk}^2 \right]
\end{aligned}$$

The estimator $\hat{\sigma}_{u(2)}^{2(ht)}$ can also be obtained directly from equation of $E_2\left(s^{(2)}\right)$ as below:

$$s^{(2)} = \frac{n-\bar{n}_0^{(2)}}{C_s - 1} \hat{\sigma}_{u(2)}^{2(ht)} + \frac{1}{C_s - 1} \sum_{ij\in s} \left(\frac{1}{n_{ij}} - \frac{1}{n} \right) \sum_k \sigma_{\varepsilon(2),ijk}^2$$

$\Rightarrow \dfrac{n-\bar{n}_0^{(2)}}{C_s - 1} \hat{\sigma}_{u(2)}^{2(ht)} = s^{(2)} - \dfrac{1}{C_s - 1} \sum_{ij\in s} \left(\dfrac{1}{n_{ij}} - \dfrac{1}{n} \right) \sum_k \sigma_{\varepsilon(2),ijk}^2$ and hence

$$\hat{\sigma}_{u(2)}^{2(ht)} = \frac{1}{n-\bar{n}_0^{(2)}} \left\{ (C_s - 1) s^{(2)} - \sum_{ij\in s} \left(n_{ij}^{-1} - n \right) \sum_k \sigma_{\varepsilon(2),ijk}^2 \right\}.$$

The terms involved in $\hat{\sigma}_{u(2)}^{2(ht)}$ can be expressed putting $w_{ij} = \dfrac{n_{ij}}{n}$ as $n - \bar{n}_0^{(2)} = n\sum_{ij} \dfrac{n_{ij}}{n}\left(1 - \dfrac{n_{ij}}{n}\right)$

$= n\sum_{ij} w_{ij}\left(1 - w_{ij}\right)$, $\sum_{ij\in s}\left(\dfrac{1}{n_{ij}} - \dfrac{1}{n}\right) = n\sum_{ij\in s} w_{ij}\left(1 - w_{ij}\right)\dfrac{1}{n_{ij}^2}$, and $\left(C_s - 1\right)s^{(2)} = n\sum_{ij\in s} w_{ij}\left(\hat{\bar{e}}_{ij.} - \hat{\bar{e}}_{...}\right)^2$.

Then the estimator of cluster level variance component becomes

$$\hat{\sigma}_{u(2)}^{2(ht)} = \left\{\sum_{ij} w_{ij}\left(1 - w_{ij}\right)\right\}^{-1}\left\{\sum_{ij\in s} w_{ij}\left(\hat{\bar{e}}_{ij.} - \hat{\bar{e}}_{...}\right)^2 - \sum_{ij\in s} w_{ij}\left(1 - w_{ij}\right)\frac{1}{n_{ij}^2}\tau_{ij}^2\right\}$$

where $\tau_{ij}^2 = n_{ij}^{-2}\sum_{k\in s}\sigma_{\varepsilon(2),ijk}^2$. Thus the estimator depends on a function of cluster-level components $\left(\tau_{ij}^2\right)$ which itself a function of HH-level error variances $\sigma_{\varepsilon(2),ijk}^2$. An unbiased estimator of τ_{ij}^2 is $\hat{\tau}_{ij}^2 = n_{ij}^{-1}\left(n_{ij} - 1\right)^{-1}\sum_{k\in s}\left(\hat{\varepsilon}_{ijk} - \hat{\varepsilon}_{ij.}\right)^2 = n_{ij}^{-1}\left(n_{ij} - 1\right)^{-1}\sum_{k\in s}\left(\hat{e}_{ijk} - \hat{\bar{e}}_{ij.}\right)^2$ where $\hat{\varepsilon}_{ij.} = n_{ij}^{-1}\sum_{k\in s}\hat{\varepsilon}_{ijk}$. It is easy to show that $E_2\left(\hat{\tau}_{ij}^2\right) = n_{ij}^{-2}\sum_{k\in s}\sigma_{\varepsilon(2),ijk}^2 = \tau_{ij}^2$ since

$$E_2\left[\sum_{k\in s}\left(\hat{e}_{ijk} - \hat{\bar{e}}_{ij.}\right)^2\right] = \sum_{k\in s}\left(\sigma_{u(2)}^{2(ht)} + \sigma_{\varepsilon(2),ijk}^2\right) - n_{ij}\left(\sigma_{u(2)}^{2(ht)} + n_{ij}^{-2}\sum_{k\in s}\sigma_{\varepsilon(2),ijk}^2\right) = \left(n_{ij} - 1\right)n_{ij}^{-1}\sum_{k\in s}\sigma_{\varepsilon(2),ijk}^2 .$$

Thus the ultimate plug-in estimator of $\sigma_{u(2)}^2$ considering heteroskedasticity at the level-one is

$$\hat{\sigma}_{u(2)}^2 = \max\left\{\left(\sum_{ij\in s} w_{ij}\left(1 - w_{ij}\right)\right)^{-1}\left\{\left[\sum_{ij\in s} w_{ij}\left(\hat{\bar{e}}_{ij.} - \hat{\bar{e}}_{...}\right)^2\right] - \sum_{ij\in s} w_{ij}\left(1 - w_{ij}\right)\hat{\tau}_{ij}^2\right\}, 0\right\} .$$

An important characteristics of this HT estimator $\hat{\sigma}_{u(2)}^{2(ht)}$ is that it will be same as the HM estimator $\hat{\sigma}_{u(2)}^2$ if $n_{ij} = m$. In such case, the key term of $\hat{\sigma}_{u(2)}^{2(ht)}$ becomes

$$\sum_{ij\in s} w_{ij}\left(1 - w_{ij}\right)\hat{\tau}_{ij}^2 = \sum_{ij\in s}\frac{m}{n}\left(1 - \frac{m}{n}\right)m^{-1}\left(m - 1\right)^{-1}\sum_{k\in s}\left\{\left(\hat{\varepsilon}_{ijk} - \hat{\varepsilon}_{...}\right) - \left(\hat{\varepsilon}_{ij.} - \hat{\varepsilon}_{...}\right)\right\}^2$$
$$= \frac{n - m}{n^2\left(m - 1\right)}\sum_{ij\in s}\left\{\sum_{k=1}^{n_{ij}}\left(\hat{\varepsilon}_{ijk} - \hat{\varepsilon}_{...}\right)^2 - m\left(\hat{\varepsilon}_{ij.} - \hat{\varepsilon}_{...}\right)^2\right\} = \frac{n - m}{n^2\left(m - 1\right)}\left[\left(n - 1\right)s^{(1)} - \left(C_s - 1\right)s^{(2)}\right].$$

which leads to

$$
\begin{aligned}
\hat{\sigma}_u^{2(ht)} &= \frac{n}{\left(n-\bar{n}_0^{(2)}\right)}\left[\frac{(C_s-1)}{n}s^{(2)} - \left\{\frac{(n-m)}{n^2(m-1)}\left[(n-1)s^{(1)}-(C_s-1)s^{(2)}\right]\right\}\right] \\
&= \frac{n}{\left(n-\bar{n}_0^{(2)}\right)}\left[\frac{(C_s-1)}{n}s^{(2)}\left\{1+\frac{(n-m)}{n(m-1)}\right\} - \frac{(n-m)(n-1)}{n^2(m-1)}s^{(1)}\right] \\
&= \frac{(n-1)}{\left(n-\bar{n}_0^{(2)}\right)}\left[\frac{m(C_s-1)}{n(m-1)}s^{(2)} - \frac{(n-m)}{n(m-1)}s^{(1)}\right] \\
&= \frac{(n-1)}{\left(n-\bar{n}_0^{(2)}\right)}\frac{(C_s-1)}{(n-C_s)}\left[\frac{m(n-C_s)}{n(m-1)}s^{(2)} - \frac{(n-C_s)(n-m)}{n(m-1)(C_s-1)}s^{(1)}\right] \\
&= \frac{(n-1)}{\left(n-\bar{n}_0^{(2)}\right)}\frac{(C_s-1)}{(n-C_s)}\left[s^{(2)}-s^{(1)}\right] = \hat{\sigma}_{u(2)}^2 \quad [\because n = mC_s] \; .
\end{aligned}
$$

Thus the estimators $\hat{\sigma}_{u(2)}^{2(ht)}$ under heteroskedasticity will provide the same estimate as $\hat{\sigma}_{u(2)}^2$ if the number of sampled HHs per cluster is same for all the sampled clusters.

A.7 Variance Component Estimation: 3-level Model with Heteroskedastic (HT) Level-one Errors

Consider a 3-level population model including an area-level random effect in the superpopulation model (5) as

$$y_{ijk} = \mathbf{x}_{ijk}^T\boldsymbol{\beta}_{(3)} + \eta_i + u_{ij} + \varepsilon_{ijk} \text{ with } \eta_i \sim N\left(0, \sigma_{\eta(3)}^{2(ht)}\right), u_{ij} \sim N\left(0, \sigma_{u(3)}^{2(ht)}\right), \varepsilon_{ijk} \sim N\left(0, \sigma_{\varepsilon(3),ijk}^2\right) \quad (6)$$

where η_i, u_{ij}, and ε_{ijk} are identically and independently distributed from each other. As previous, the estimated OLS residuals are utilized to calculate moment-based estimates of area-level, cluster-level and HH-level random effects as $\hat{\eta}_i = n_i^{-1}\sum_{jk\in s}\hat{e}_{ijk}$, $\hat{u}_{ij} = n_{ij}^{-1}\sum_{k\in s}\hat{e}_{ijk}$, and $\hat{\varepsilon}_{ijk} = \hat{e}_{ijk} - \hat{\eta}_i - \hat{u}_{ij}$ where $\hat{e}_{ijk} = y_{ijk} - \hat{y}_{ijk}$.

Under the 3-level model (6) with heteroskedasticity at HH-level, we have

$$\mathrm{Var}_3\left(\hat{\bar{e}}_{ij.}\right)=E_3\left(\hat{\bar{e}}_{ij.}^2\right)=n_{ij}^{-2}\left[n_{ij}\left(\sigma_{\eta(3)}^{2(ht)}+\sigma_{u(3)}^{2(ht)}\right)+\sum_{k\in s}\sigma_{\varepsilon(3),ijk}^2+n_{ij}\left(n_{ij}-1\right)\left(\sigma_{\eta(3)}^{2(ht)}+\sigma_{u(3)}^{2(ht)}\right)\right]$$
$$=\sigma_{\eta(3)}^{2(ht)}+\sigma_{u(3)}^{2(ht)}+n_{ij}^{-2}\sum_{k\in s}\sigma_{\varepsilon(3),ijk}^2 ,$$

$$\mathrm{Var}_3\left(\hat{\bar{e}}_{i..}\right)=E_3\left(\hat{\bar{e}}_{i..}^2\right)=\mathrm{Var}_3\left(n_i^{-1}\sum_{j\in s}n_{ij}\hat{\bar{e}}_{ij.}\right)=n_i^{-2}\left[\sum_{j\in s}n_{ij}^2\,\mathrm{var}\left(\hat{\bar{e}}_{ij.}\right)+\sum_{j\neq j'\in s}n_{ij}n_{ij'}\,\mathrm{cov}\left(\hat{\bar{e}}_{ij.},\hat{\bar{e}}_{ij'.}\right)\right]$$
$$=n_i^{-2}\left[\left(\sum_{j\in s}n_{ij}^2+\sum_{j\neq j'\in s}n_{ij}n_{ij'}\right)\sigma_{\eta(3)}^{2(ht)}+\sum_{j\in s}n_{ij}^2\sigma_{u(3)}^{2(ht)}+\sum_{jk\in s}\sigma_{\varepsilon(3),ijk}^2\right]=\sigma_{\eta(3)}^{2(ht)}+n_i^{-2}\sum_{j\in s}n_{ij}^2\sigma_{u(3)}^{2(ht)}+n_i^{-2}\sum_{jk\in s}\sigma_{\varepsilon(3),ijk}^2$$

$$\mathrm{Var}_3\left(\hat{\bar{e}}_{...}\right)=\mathrm{Var}_3\left(n^{-1}\sum_{i\in s}n_i\hat{\bar{e}}_{i..}\right)=n^{-2}\left[\sum_{i\in s}n_i^2\mathrm{Var}_3\left(\hat{\bar{e}}_{i..}\right)+\sum_{i\neq i'\in s}n_in_{i'}\mathrm{Cov}_3\left(\hat{\bar{e}}_{i..},\hat{\bar{e}}_{i'..}\right)\right]$$
$$=n^{-2}\left[\sum_{i\in s}n_i^2\left\{\sigma_{\eta(3)}^{2(ht)}+n_i^{-2}\sum_{j\in s}n_{ij}^2\sigma_{u(3)}^{2(ht)}+n_i^{-2}\sum_{jk\in s}\sigma_{\varepsilon(3),ijk}^2\right\}\right]=n^{-2}\left(\sum_{i\in s}n_i^2\sigma_{\eta(3)}^{2(ht)}+\sum_{ij\in s}n_{ij}^2\sigma_{u(3)}^{2(ht)}+\sum_{ijk\in s}\sigma_{\varepsilon(3),ijk}^2\right),$$

$$E_3\left(\sum_{ijk\in s}\hat{e}_{ijk}^2\right)=\sum_{ijk\in s}E_3\left(\hat{e}_{ijk}^2\right)=n\left(\sigma_{\eta(3)}^{2(ht)}+\sigma_{u(3)}^{2(ht)}\right)+\sum_{ijk\in s}\sigma_{\varepsilon(3),ijk}^2 ,$$

$$E_3\left(\sum_{ij\in s}n_{ij}\hat{\bar{e}}_{ij.}^2\right)=\sum_{ij\in s}E_3\left(n_{ij}\hat{\bar{e}}_{ij.}^2\right)=\sum_{ij\in s}n_{ij}\left\{\sigma_{\eta(3)}^2+\sigma_{u(3)}^2+\frac{1}{n_{ij}^2}\sum_{k\in s}\sigma_{\varepsilon(3),ijk}^2\right\}$$
$$=n\left(\sigma_{\eta(3)}^{2(ht)}+\sigma_{u(3)}^{2(ht)}\right)+\sum_{ij\in s}\frac{1}{n_{ij}}\sum_{k\in s}\sigma_{\varepsilon(3),ijk}^2 ,$$

$$E_3\left(\sum_{i\in s}n_i\hat{\bar{e}}_{i..}^2\right)=\sum_{i\in s}n_i\left\{\sigma_{\eta(3)}^2+n_i^{-2}\sum_{j\in s}n_{ij}^2\sigma_{u(3)}^2+n_i^{-2}\sum_{jk\in s}\sigma_{\varepsilon(3),ijk}^2\right\}$$
$$=n\sigma_{\eta(3)}^{2(ht)}+\sum_{i\in s}\bar{n}_{0i}^{(2)}\sigma_{u(3)}^{2(ht)}+\sum_{i\in s}n_i^{-1}\sum_{jk\in s}\sigma_{\varepsilon(3),ijk}^2 \text{, and}$$

$$E_3\left(\sum_{ijk\in s}\hat{e}_{ijk}\right)^2=\sum_{ijk\in s}E_3\left(\hat{e}_{ijk}^2\right)+\sum_{ij\in s}\sum_{k\neq k'}E_3\left(\hat{e}_{ijk}\hat{e}_{ijk'}\right)+\sum_{i\in s}\sum_{j\neq j'\in s}\sum_{k\neq k'}E_3\left(\hat{e}_{ijk}\hat{e}_{ij'k'}\right)+\sum_{i\neq i'\in s}\sum_{j\neq j'\in s}\sum_{k\neq k'}E_3\left(\hat{e}_{ijk}\hat{e}_{i'j'k'}\right)$$
$$=n\left(\sigma_{\eta(3)}^{2(ht)}+\sigma_{u(3)}^{2(ht)}\right)+\sum_{ijk\in s}\sigma_{\varepsilon(3),ijk}^2+\sum_{ij\in s}n_{ij}\left(n_{ij}-1\right)\sigma_{u(3)}^{2(ht)}+\sigma_{\eta(3)}^{2(ht)}\sum_{i\in s}n_i\left(n_i-1\right)$$
$$=\sum_{ijk\in s}\sigma_{\varepsilon(3),ijk}^2+\sigma_{u(3)}^{2(ht)}\sum_{ij\in s}n_{ij}^2+\sigma_{\eta(3)}^{2(ht)}\sum_{i\in s}n_i^2 .$$

Now the expectation of the sample residual variances become

$$E_3\left[s^{(1)}\right]=(n-1)^{-1}\left[n\left(\sigma^{2(ht)}_{\eta(3)}+\sigma^{2(ht)}_{u(3)}\right)+\sum_{ijk\in s}\sigma^2_{\varepsilon(3),ijk}-n^{-1}\left\{\sum_{ijk\in s}\sigma^2_{\varepsilon(3),ijk}+\sigma^{2(ht)}_{u(3)}\sum_{ij\in s}n_{ij}^2+\sigma^{2(ht)}_{\eta(3)}\right\}\right]$$

$$=(n-1)^{-1}\left[\sum_{ijk\in s}\sigma^2_{\varepsilon(3),ijk}\left(1-n^{-1}\right)+\sigma^{2(ht)}_{u(3)}\left\{n-n^{-1}\sum_{ij\in s}n_{ij}^2\right\}+\sigma^{2(ht)}_{\eta(3)}\left\{n-n^{-1}\sum_{i\in s}n_i^2\right\}\right]$$

$$=n^{-1}\sum_{ijk\in s}\sigma^2_{\varepsilon(3),ijk}+\frac{n-\overline{n}_0^{(2)}}{n-1}\sigma^{2(ht)}_{u(3)}+\frac{n-\overline{n}_0^{(3)}}{n-1}\sigma^{2(ht)}_{\eta(3)}=a_{11}+a_{12}\sigma^{2(ht)}_{u(3)}+a_{13}\sigma^{2(ht)}_{\eta(3)},$$

$$E_3\left[s^{(2)}\right]=(C_s-1)^{-1}\begin{bmatrix}n\left(\sigma^{2(ht)}_{\eta(3)}+\sigma^{2(ht)}_{u(3)}\right)+\sum_{ij\in s}n_{ij}^{-1}\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}-\\ n^{-1}\left\{\sum_{ijk\in s}\sigma^2_{\varepsilon(3),ijk}+\sigma^{2(ht)}_{u(3)}\sum_{ij\in s}n_{ij}^2+\sigma^{2(ht)}_{\eta(3)}\sum_{i\in s}n_i^2\right\}\end{bmatrix}$$

$$=(C_s-1)^{-1}\left[\sum_{ij\in s}\left(n_{ij}^{-1}-n^{-1}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}\right]+(C_s-1)^{-1}\left(n-\overline{n}_0^{(2)}\right)\sigma^{2(ht)}_{u(3)}+(C_s-1)^{-1}\left(n-\overline{n}_0^{(3)}\right)\sigma^{2(ht)}_{\eta(3)}$$

$=a_{21}+a_{22}\sigma^{2(ht)}_{u(3)}+a_{23}\sigma^{2(ht)}_{\eta(3)}$, and

$$E_3\left[s^{(3)}\right]=(D-1)^{-1}\begin{bmatrix}n\sigma^{2(ht)}_{\eta(3)}+\sum_{i\in s}\overline{n}_{0i}^{(2)}\sigma^{2(ht)}_{u(3)}+\sum_{i\in s}n_i^{-1}\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}-\\ n^{-1}\left(\sum_{ijk\in s}\sigma^2_{\varepsilon(3),ijk}+\sigma^{2(ht)}_{u(3)}\sum_{ij\in s}n_{ij}^2+\sigma^{2(ht)}_{\eta(3)}\sum_{i\in s}n_i^2\right)\end{bmatrix}$$

$$=(D-1)^{-1}\sum_{i\in s}\left(n_i^{-1}-n^{-1}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}+\frac{\sum_{i\in s}\overline{n}_{0i}^{(2)}-\overline{n}_0^{(2)}}{D-1}\sigma^{2(ht)}_{u(3)}+\frac{n-\overline{n}_0^{(3)}}{D-1}\sigma^{2(ht)}_{\eta(3)}$$

$$=a_{31}+a_{32}\sigma^{2(ht)}_{u(3)}+a_{33}\sigma^{2(ht)}_{\eta(3)}$$

where $\overline{n}_0^{(2)}=n^{-1}\sum_{ij\in s}n_{ij}^2$ and $\overline{n}_0^{(3)}=n^{-1}\sum_{i\in s}n_i^2$. Under the assumption of known $\sigma^2_{\varepsilon(3),ijk}$, the expectations can be expresses in matrix form as

$$E_3\begin{bmatrix}s^{(1)}\\ s^{(2)}\\ s^{(3)}\end{bmatrix}=\begin{bmatrix}\frac{\Psi_{11}}{n} & \frac{n-\overline{n}_0^{(2)}}{n-1} & \frac{n-\overline{n}_0^{(3)}}{n-1}\\ \frac{\Psi_{21}}{C_s-1} & \frac{n-\overline{n}_0^{(2)}}{C_s-1} & \frac{n-\overline{n}_0^{(3)}}{C_s-1}\\ \frac{\Psi_{31}}{D-1} & \frac{\sum_{i\in s}\overline{n}_{0i}^{(2)}-\overline{n}_0^{(2)}}{D-1} & \frac{n-\overline{n}_0^{(3)}}{D-1}\end{bmatrix}\begin{bmatrix}1\\ \sigma^{2(ht)}_{u(3)}\\ \sigma^{2(ht)}_{\eta(3)}\end{bmatrix}=A_5\Lambda_5$$

putting $\psi_{11}=\sum_{ijk\in s}\sigma^2_{\varepsilon(3),ijk}$, $\psi_{21}=\sum_{ij\in s}\left(n_{ij}^{-1}-n^{-1}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}$, and $\psi_{31}=\sum_{i\in s}\left(n_i^{-1}-n^{-1}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}$.

The coefficient matrix will be the same as $A_{(3)}$ shown in **Appendix A.2** if $\sigma^2_{\varepsilon(3).ijk}=\sigma^2_{\varepsilon(3)}$. In such case, the unbiased estimators of the variance components $\left(\sigma^2_{\eta(3)},\sigma^2_{u(3)},\sigma^2_{\varepsilon(3)}\right)$ can be easily obtained if the new coefficient matrix A_5 is non-singular (Tranmer & Steel, 2001a). Assuming the coefficients φ_{11}, φ_{21}, and φ_{31} as known, the estimator of Λ_5 can be easily obtained from $\hat{\Lambda}_5=A_5^{-1}S_1$. The determinant and conjugate of A_5 are

$$|A_5|=a_{11}\left(a_{22}a_{33}-a_{32}a_{23}\right)-a_{21}\left(a_{33}a_{12}-a_{32}a_{13}\right)+a_{31}\left(a_{23}a_{12}-a_{22}a_{13}\right)$$

$$=\frac{\psi_{11}}{n}\left\{\frac{n-\bar{n}_0^{(2)}}{C_s-1}\frac{n-\bar{n}_0^{(3)}}{D-1}-\frac{\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}}{D-1}\frac{n-\bar{n}_0^{(3)}}{C_s-1}\right\}$$
$$-\frac{\psi_{21}}{C_s-1}\left(\frac{n-\bar{n}_0^{(3)}}{D-1}\frac{n-\bar{n}_0^{(2)}}{n-1}-\frac{\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}}{D-1}\frac{n-\bar{n}_0^{(3)}}{n-1}\right)$$
$$+\frac{\psi_{31}}{D-1}\left\{\frac{n-\bar{n}_0^{(3)}}{C_s-1}\frac{n-\bar{n}_0^{(2)}}{n-1}-\frac{n-\bar{n}_0^{(2)}}{C_s-1}\frac{n-\bar{n}_0^{(3)}}{n-1}\right\}$$

$$=\frac{n^{-1}\psi_{11}}{D-1}\frac{n-\bar{n}_0^{(3)}}{C_s-1}\left\{\left(n-\bar{n}_0^{(2)}\right)-\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\right\}$$
$$-\frac{\psi_{21}}{(C_s-1)(n-1)}\frac{n-\bar{n}_0^{(3)}}{D-1}\left\{\left(n-\bar{n}_0^{(2)}\right)-\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\right\}+0$$

$$=\frac{n^{-1}\psi_{11}}{D-1}\frac{n-\bar{n}_0^{(3)}}{C_s-1}\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)-\frac{\psi_{21}}{(C_s-1)(n-1)}\frac{n-\bar{n}_0^{(3)}}{D-1}\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)$$

$$=\frac{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}{D-1}\frac{n-\bar{n}_0^{(3)}}{C_s-1}\left\{\frac{\psi_{11}}{n}-\frac{\psi_{21}}{n-1}\right\}=\frac{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}{D-1}\frac{n-\bar{n}_0^{(3)}}{C_s-1}\frac{1}{n-1}\left\{\frac{n-1}{n}\psi_{11}-\psi_{21}\right\}$$

and $\text{Conjugate}(A_5)=\begin{bmatrix} \frac{n-\bar{n}_0^{(3)}}{C_s-1}\frac{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}{D-1} & -\frac{n-\bar{n}_0^{(3)}}{n-1}\frac{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}{D-1} & 0 \\ -\frac{n-\bar{n}_0^{(3)}}{D-1}\frac{\varphi_{21}}{C_s-1} & \frac{n-\bar{n}_0^{(3)}}{D-1}\frac{\varphi_{22}}{(n-1)} & -\frac{n-\bar{n}_0^{(3)}}{C_s-1}\frac{\varphi_{00}}{(n-1)} \\ \frac{\varphi_{31}}{(D-1)(C_s-1)} & -\frac{\varphi_{32}}{(n-1)(D-1)} & \frac{n-\bar{n}_0^{(2)}}{C_s-1}\frac{\varphi_{00}}{(n-1)} \end{bmatrix}$

where,

$$\varphi_{00}=\frac{n-1}{n}\psi_{11}-\psi_{21}\ ,\ \varphi_{21}=\psi_{21}-\psi_{31},\ \varphi_{22}=\frac{n-1}{n}\psi_{11}-\psi_{31},$$

$$\varphi_{31}=\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\psi_{21}-\left(n-\bar{n}_0^{(2)}\right)\psi_{31}, \text{ and } \varphi_{32}=\frac{n-1}{n}\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\psi_{11}-\left(n-\bar{n}_0^{(2)}\right)\psi_{31}.$$

There are some relationships among the coefficients as:

$$\varphi_{21}-\varphi_{22}=\psi_{21}-\frac{n-1}{n}\psi_{11}=-\varphi_{00} \text{ and}$$

$$(\varphi_{31}-\varphi_{32})=\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\left(\psi_{21}-\frac{n-1}{n}\psi_{11}\right)=-\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\varphi_{00}.$$

Then the inverse of A_5 is

$$A_5^{-1}=\begin{bmatrix} \frac{(n-1)}{\varphi_{00}} & -\frac{(C_s-1)}{\varphi_{00}} & 0 \\ -\frac{n-1}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}\frac{\varphi_{21}}{\varphi_{00}} & \frac{\varphi_{22}}{\varphi_{00}}\frac{C_s-1}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}} & -\frac{D-1}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}} \\ \frac{(n-1)\varphi_{31}}{\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)\left(n-\bar{n}_0^{(3)}\right)\varphi_{00}} & -\frac{(C_s-1)\varphi_{32}}{\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)\left(n-\bar{n}_0^{(3)}\right)\varphi_{00}} & \frac{(D-1)\left(n-\bar{n}_0^{(2)}\right)}{\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)\left(n-\bar{n}_0^{(3)}\right)} \end{bmatrix}$$

from which the estimators of $\sigma_{u(3)}^2$ and $\sigma_{\eta(3)}^2$ can be written as

$$\hat{\sigma}_{u(3)}^{2(ht)} = -\frac{(n-1)\varphi_{21}}{\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)\varphi_{00}} s^{(1)} + \frac{\varphi_{22}}{\varphi_{00}} \frac{(C_s-1)}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}} s^{(2)} - \frac{D-1}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}} s^{(3)} \text{ and}$$

$$\hat{\sigma}_{\eta(3)}^{2(ht)} = \frac{s^{(1)}(n-1)\varphi_{31}}{\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)\left(n-\bar{n}_0^{(3)}\right)\varphi_{00}} - \frac{s^{(2)}(C_s-1)\varphi_{32}}{\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)\left(n-\bar{n}_0^{(3)}\right)\varphi_{00}} + \frac{s^{(3)}(D-1)\left(n-\bar{n}_0^{(2)}\right)}{\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)\left(n-\bar{n}_0^{(3)}\right)}.$$

To obtain a simplest form of $\hat{\sigma}_{u(3)}^{2(ht)}$ and $\hat{\sigma}_{\eta(3)}^{2(ht)}$, the HH-level sample residual variance $s^{(1)}$ is expressed as a function of $\sum_{ijk\in s}\sigma_{\varepsilon(3),ijk}^2$ and $s^{(2)}$ from $A_5^{-1}S$ as

$$1 = \varphi_{00}^{-1}\left\{(n-1)s^{(1)} - (C_s-1)s^{(2)}\right\} \Rightarrow s^{(1)} = \frac{\varphi_{00}}{n-1} + \frac{C_s-1}{n-1}s^{(2)}.$$

Putting the value of $s^{(1)}$ in $\hat{\sigma}_{u(3)}^{2(ht)}$ and $\hat{\sigma}_{\eta(3)}^{2(ht)}$ we have,

$$\begin{aligned}
\hat{\sigma}_{u(3)}^{2(ht)} &= -\frac{(n-1)\varphi_{21}}{\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)\varphi_{00}}\left\{\frac{\varphi_{00}}{n-1} + \frac{C_s-1}{n-1}s^{(2)}\right\} + \frac{\varphi_{22}}{\varphi_{00}}\frac{(C_s-1)}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}s^{(2)} - \frac{D-1}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}s^{(3)} \\
&= -\frac{\varphi_{21}}{\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)} - \frac{\varphi_{21}(C_s-1)}{\varphi_{00}\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)}s^{(2)} + \frac{\varphi_{22}}{\varphi_{00}}\frac{(C_s-1)}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}s^{(2)} - \frac{D-1}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}s^{(3)} \\
&= -\frac{\varphi_{21}}{\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)} - (\varphi_{21}-\varphi_{22})\frac{(C_s-1)}{\varphi_{00}\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)}s^{(2)} - \frac{D-1}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}s^{(3)} \\
&= -\frac{\varphi_{21}}{\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)} + \varphi_{00}\frac{(C_s-1)}{\varphi_{00}\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)}s^{(2)} - \frac{D-1}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}s^{(3)} \quad [\because \varphi_{21}+\varphi_{22} = -\varphi_{00}] \\
&= -\frac{\varphi_{21}}{\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)} + \frac{(C_s-1)}{\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)}s^{(2)} - \frac{D-1}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}s^{(3)} \\
&= \left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)^{-1}\left\{-\varphi_{21} + (C_s-1)s^{(2)} - (D-1)s^{(3)}\right\} \\
&= \frac{1}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}\left[-\left\{\sum_{ij\in s}\left(\frac{1}{n_{ij}}-\frac{1}{n}\right)\sum_{k\in s}\sigma_{\varepsilon(3),ijk}^2 - \sum_{i\in s}\left(\frac{1}{n_i}-\frac{1}{n}\right)\sum_{jk\in s}\sigma_{\varepsilon(3),ijk}^2\right\} + (C_s-1)s^{(2)} - (D-1)s^{(3)}\right]
\end{aligned}$$

$$
\begin{aligned}
\hat{\sigma}^{2(ht)}_{\eta(3)} &= \frac{(n-1)\varphi_{31}}{\left(n-\sum_{i\in s}\bar{n}^{(2)}_{0i}\right)\left(n-\bar{n}^{(3)}_{0}\right)\varphi_{00}} s^{(1)} - \frac{(C_s-1)\varphi_{32}}{\left(n-\sum_{i\in s}\bar{n}^{(2)}_{0i}\right)\left(n-\bar{n}^{(3)}_{0}\right)\varphi_{00}} s^{(2)} + \frac{(D-1)\left(n-\bar{n}^{(2)}_{0}\right)}{\left(n-\sum_{i\in s}\bar{n}^{(2)}_{0i}\right)\left(n-\bar{n}^{(3)}_{0}\right)} s^{(3)} \\
&= \frac{\left(n-\bar{n}^{(3)}_{0}\right)^{-1}}{n-\sum_{i\in s}\bar{n}^{(2)}_{0i}} \left\{ \frac{(n-1)\varphi_{31}}{\varphi_{00}} s^{(1)} - \frac{(C_s-1)\varphi_{32}}{\varphi_{00}} s^{(2)} + (D-1)\left(n-\bar{n}^{(2)}_{0}\right) s^{(3)} \right\} \\
&= \frac{\left(n-\bar{n}^{(3)}_{0}\right)^{-1}}{n-\sum_{i\in s}\bar{n}^{(2)}_{0i}} \left\{ \frac{(n-1)\varphi_{31}}{\varphi_{00}} \left(\frac{\varphi_{00}}{n-1} + \frac{C_s-1}{n-1} s^{(2)} \right) - \frac{(C_s-1)\varphi_{32}}{\varphi_{00}} s^{(2)} + (D-1)\left(n-\bar{n}^{(2)}_{0}\right) s^{(3)} \right\} \\
&= \frac{\left(n-\bar{n}^{(3)}_{0}\right)^{-1}}{n-\sum_{i\in s}\bar{n}^{(2)}_{0i}} \left\{ \varphi_{31} + \frac{\varphi_{31}}{\varphi_{00}} (C_s-1) s^{(2)} - \frac{(C_s-1)\varphi_{32}}{\varphi_{00}} s^{(2)} + (D-1)\left(n-\bar{n}^{(2)}_{0}\right) s^{(3)} \right\} \\
&= \frac{\left(n-\bar{n}^{(3)}_{0}\right)^{-1}}{n-\sum_{i\in s}\bar{n}^{(2)}_{0i}} \left\{ \varphi_{31} + (\varphi_{31}-\varphi_{32}) \frac{(C_s-1)}{\varphi_{00}} s^{(2)} + (D-1)\left(n-\bar{n}^{(2)}_{0}\right) s^{(3)} \right\}.
\end{aligned}
$$

Putting the values of φ_{31} and $(\varphi_{31}-\varphi_{32}) = -\left(\sum_{i\in s}\bar{n}^{(2)}_{0i} - \bar{n}^{(2)}_{0}\right)\varphi_{00}$, we get

$$
\hat{\sigma}^{2}_{\eta(3)} = \frac{1}{\left(n-\bar{n}^{(3)}_{0}\right)\left(n-\sum_{i\in s}\bar{n}^{(2)}_{0i}\right)} \times
\left\{ \begin{array}{l}
\left(\sum_{i\in s}\bar{n}^{(2)}_{0i} - \bar{n}^{(2)}_{0}\right) \sum_{ij\in s}\left(n_{ij}^{-1}-n^{-1}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk} - \left(n-\bar{n}^{(2)}_{0}\right)\sum_{i\in s}\left(n_i^{-1}-n^{-1}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk} \\
-\left(\sum_{i\in s}\bar{n}^{(2)}_{0i} - \bar{n}^{(2)}_{0}\right)(C_s-1)s^{(2)} + (D-1)\left(n-\bar{n}^{(2)}_{0}\right)s^{(3)}
\end{array} \right\}.
$$

Under homoskedasticity at HH-level, the complex terms become $\varphi_{00} = n - C_s$, $\varphi_{22} = n - D$, and $\varphi_{21} = C_s - D$ which lead to the unbiased estimators of variance components as (4) under the HM model (3).

The moment-based estimators of $\sigma^{2(ht)}_{u(3)}$ and $\sigma^{2(ht)}_{\eta(3)}$ depend on a cluster-specific term $\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}$ and an area-specific term $\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}$ which are themselves functions of level-one variances $\sigma^2_{\varepsilon(3),ijk}$. Assume a cluster-specific term $\tau^2_{ij} = n_{ij}^{-2}\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}$ and an area-specific term $\tau^2_{i} = n_{i}^{-2}\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}$ and their corresponding unbiased estimators as $\hat{\tau}^2_{ij} = n_{ij}^{-1}\left(n_{ij}-1\right)^{-1}\sum_{k\in s}\left(\hat{\varepsilon}_{ijk}-\hat{\varepsilon}_{ij.}\right)^2$ and $\hat{\tau}^2_{i} = n_{i}^{-2}\sum_{j\in s}n_{ij}^2\hat{\tau}^2_{ij}$ under the considered 3-level model with $\hat{\varepsilon}_{ij.} = n_{ij}^{-1}\sum_{k\in s}\hat{\varepsilon}_{ijk}$ and $\hat{\varepsilon}_{i..} = n_{i}^{-1}\sum_{jk\in s}\hat{\varepsilon}_{ijk}$. It can be easily shown that $E\left(\hat{\tau}^2_{ij}\right) = n_{ij}^{-2}\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk} = \tau^2_{ij}$ and $E\left(\hat{\tau}^2_{i}\right) = n_{i}^{-2}\sum_{j\in s}n_{ij}^2 E\left(\hat{\tau}^2_{ij}\right) = n_{i}^{-2}\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk} = \tau^2_{i}$, since $E\left[\sum_{k\in s}\left(\hat{\varepsilon}_{ijk}-\hat{\varepsilon}_{ij.}\right)^2\right] = E\left[\sum_{k\in s}\left(\hat{e}_{ijk}-\hat{\bar{e}}_{ij.}\right)^2\right] = \sum_{k\in s}E\left(\hat{e}^2_{ijk}\right) - n_{ij}E\left(\hat{\bar{e}}^2_{ij.}\right) = \frac{n_{ij}-1}{n_{ij}}\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}$. Thus the estimators can be expressed as

$$\hat{\sigma}^{2(ht)}_{u(3)} = \left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)^{-1}\left[\left(C-1\right)s^{(2)} - \left(D-1\right)s^{(3)} - \sum_{ij\in s}\left(\frac{1}{n_{ij}}-\frac{1}{n}\right)n_{ij}^2\hat{\tau}^2_{ij} + \sum_{i\in s}\left(\frac{1}{n_i}-\frac{1}{n}\right)n_i^2\hat{\tau}^2_i\right] \text{ and}$$

$$\hat{\sigma}^{2(ht)}_{\eta(3)} = \frac{1}{\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)\left(n-\bar{n}_{0}^{(3)}\right)}\left[\begin{array}{l}\left(n-\bar{n}_0^{(2)}\right)\left\{\left(D-1\right)s^{(3)} - \sum_{i\in s}\left(\frac{1}{n_i}-\frac{1}{n}\right)n_i^2\hat{\tau}^2_i\right\} \\ -\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\left\{\left(C-1\right)s^{(2)} - \sum_{ij\in s}\left(\frac{1}{n_{ij}}-\frac{1}{n}\right)n_{ij}^2\hat{\tau}^2_{ij}\right\}\end{array}\right].$$

The complex known terms involved in $\hat{\sigma}^{2(ht)}_{u(3)}$ and $\hat{\sigma}^{2(ht)}_{\eta(3)}$ can be expressed in terms of $w_i = n_i/n$ and $w_{ij} = n_{ij}/n$ as

$$n-\sum_{i\in s}\bar{n}_{0i}^{(2)} = n\sum_{ij}\frac{n_{ij}}{n} - \sum_{i\in s}\frac{1}{n_i}\sum_{j\in s}n_{ij}^2 = n\left(\sum_{ij}\frac{n_{ij}}{n} - \sum_{i\in s}\frac{n}{n_i}\sum_{j\in s}\frac{n_{ij}^2}{n^2}\right) = n\left(\sum_{ij}w_{ij} - \sum_{i\in s}w_i^{-1}\sum_{j\in s}w_{ij}^2\right),$$

$$\left(C-1\right)s^{(2)} = \sum_{ij\in s}n_{ij}\left(\hat{\bar{e}}_{ij.}-\hat{\bar{e}}_{...}\right)^2 = n\sum_{ij\in s}w_{ij}\left(\hat{\bar{e}}_{ij.}-\hat{\bar{e}}_{...}\right)^2,$$

$$\left(D-1\right)s^{(3)} = \sum_{i\in s}n_{i}\left(\hat{\bar{e}}_{i..}-\hat{\bar{e}}_{...}\right)^2 = n\sum_{i\in s}w_{i}\left(\hat{\bar{e}}_{i..}-\hat{\bar{e}}_{...}\right)^2,$$

$$\sum_{ij\in s}\left(\frac{1}{n_{ij}}-\frac{1}{n}\right)=n\sum_{ij\in s}\left(\frac{1}{nn_{ij}}-\frac{1}{n^2}\right)=n\sum_{ij\in s}\left(\frac{n_{ij}^2}{nn_{ij}}-\frac{n_{ij}^2}{n^2}\right)n_{ij}^{-2}=n\sum_{ij\in s}\left(\frac{n_{ij}}{n}-\frac{n_{ij}^2}{n^2}\right)n_{ij}^{-2}=n\sum_{ij\in s}w_{ij}\left(1-w_{ij}\right)n_{ij}^{-2},$$

$$\sum_{i\in s}\left(\frac{1}{n_i}-\frac{1}{n}\right)=n\sum_{i\in s}\left(\frac{1}{nn_i}-\frac{1}{n^2}\right)=n\sum_{i\in s}\left(\frac{n_i^2}{nn_i}-\frac{n_i^2}{n^2}\right)\frac{1}{n_i^2}=n\sum_{i\in s}\frac{n_i}{n}\left(1-\frac{n_i}{n}\right)\frac{1}{n_i^2}=n\sum_{i\in s}w_i\left(1-w_i\right)\frac{1}{n_i^2},$$

$$n-\bar{n}_0^{(3)}=n\sum_{i\in s}\frac{n_i}{n}-\frac{1}{n}\sum_{i\in s}n_i^2=n\sum_{i\in s}\left(\frac{n_i}{n}-\frac{n_i^2}{n^2}\right)=n\sum_{i\in s}\frac{n_i}{n}\left(1-\frac{n_i}{n}\right)=n\sum_{i\in s}w_i\left(1-w_i\right),$$

$$n-\bar{n}_0^{(2)}=n\sum_{ij\in s}\frac{n_{ij}}{n}-\frac{1}{n}\sum_{ij\in s}n_{ij}^2=n\sum_{ij\in s}\left(\frac{n_{ij}}{n}-\frac{n_{ij}^2}{n^2}\right)=n\sum_{ij\in s}\frac{n_{ij}}{n}\left(1-\frac{n_{ij}}{n}\right)=n\sum_{ij\in s}w_{ij}\left(1-w_{ij}\right),\text{ and}$$

$$\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}=\sum_{i\in s}\frac{1}{n_i}\sum_{j\in s}n_{ij}^2-\frac{1}{n}\sum_{ij\in s}n_{ij}^2=\sum_{i\in s}\left(\frac{1}{n_i}-\frac{1}{n}\right)\sum_{j\in s}n_{ij}^2=n\sum_{i\in s}\left(\frac{1}{w_i}-1\right)\sum_{j\in s}w_{ij}^2\,.$$

Then the estimators become

$$\hat{\sigma}_{u(3)}^{2(ht)}=\frac{1}{\sum_{ij}w_{ij}-\sum_{i\in s}w_i^{-1}\sum_{j\in s}w_{ij}^2}\left[\begin{array}{l}\sum_{ij\in s}w_{ij}\left(\hat{\bar{e}}_{ij.}-\hat{\bar{e}}_{...}\right)^2-\sum_{i\in s}w_i\left(\hat{\bar{e}}_{i..}-\hat{\bar{e}}_{...}\right)^2\\-\sum_{ij\in s}w_{ij}\left(1-w_{ij}\right)\hat{\tau}_{ij}^2+\sum_{i\in s}w_i\left(1-w_i\right)\hat{\tau}_i^2\end{array}\right]\text{ and}$$

$$\hat{\sigma}_{\eta(3)}^{2(ht)}=\frac{1}{\sum_{i\in s}w_i\left(1-w_i\right)}\frac{1}{\sum_{ij}w_{ij}-\sum_{i\in s}w_i^{-1}\sum_{j\in s}w_{ij}^2}\times$$
$$\left[\begin{array}{l}\sum_{ij\in s}w_{ij}\left(1-w_{ij}\right)\left\{\sum_{i\in s}w_i\left(\hat{\bar{e}}_{i..}-\hat{\bar{e}}_{...}\right)^2-\sum_{i}w_i\left(1-w_i\right)\hat{\tau}_i^2\right\}\\-\sum_{i\in s}\left(w_i^{-1}-1\right)\sum_{j\in s}w_{ij}^2\left\{\sum_{ij\in s}w_{ij}\left(\hat{\bar{e}}_{ij.}-\hat{\bar{e}}_{...}\right)^2-\sum_{ij}w_{ij}\left(1-w_{ij}\right)\hat{\tau}_{ij}^2\right\}\end{array}\right].$$

If the estimators produce negative values, the variance components will be valued as zero.

The estimators can be obtained directly from the estimating equations as follows. From the second and third simultaneous equations, we have $\hat{\sigma}_{u(3)}^{2(ht)}=\left[s^{(2)}-a_{21}-a_{23}\hat{\sigma}_{\eta(3)}^{2(ht)}\right]\big/a_{22}$ and $\hat{\sigma}_{\eta(3)}^{2(ht)}=\left[s^{(3)}-a_{31}-a_{32}\hat{\sigma}_{u(3)}^{2(ht)}\right]\big/a_{33}$. Putting $\hat{\sigma}_{u(3)}^{2(ht)}$ into $\hat{\sigma}_{\eta(3)}^{2(ht)}$, we can obtain

$$\hat{\sigma}^{2(ht)}_{\eta(3)} = \frac{1}{a_{33}}\left[s^{(3)} - a_{31} - a_{32}\left\{ \frac{1}{a_{22}}\left[s^{(2)} - a_{21} - a_{23}\hat{\sigma}^{2(ht)}_{\eta(3)} \right] \right\} \right]$$

$$\Rightarrow \hat{\sigma}^{2(ht)}_{\eta(3)} = \frac{s^{(3)}}{a_{33}} - \frac{a_1^{(3)}}{a_{33}} - \frac{a_{32}}{a_{22}a_{33}}\left[s^{(2)} - a_{21} - a_{23}\hat{\sigma}^{2(ht)}_{\eta(3)} \right]$$

$$\Rightarrow \hat{\sigma}^{2(ht)}_{\eta(3)} = \frac{1}{a_{33}} s^{(3)} - \frac{a_{32}}{a_{22}a_{33}} s^{(2)} - \left[\frac{a_{31}}{a_{33}} - \frac{a_{21}a_{32}}{a_{22}a_{33}} \right] + \frac{a_{32}a_{23}}{a_{22}a_{33}} \hat{\sigma}^{2(ht)}_{\eta(3)}$$

$$\Rightarrow \sigma^{2(ht)}_{\eta(3)}\left[1 - \frac{a_{32}a_{23}}{a_{22}a_{33}} \right] = \frac{1}{a_{33}} s^{(3)} - \frac{a_{32}}{a_{22}a_{33}} s^{(2)} - \left[\frac{a_{31}a_{22} - a_{21}a_{32}}{a_{22}a_{33}} \right]$$

$$\Rightarrow \left(a_{22}a_{33} - a_{32}a_{23} \right)\hat{\sigma}^{2(ht)}_{\eta(3)} = a_{22}s^{(3)} - a_{32}s^{(2)} - \left(a_{31}a_{22} - a_{21}a_{32} \right)$$

$$\Rightarrow \hat{\sigma}^{2(ht)}_{\eta(3)} = \left[\left(a_{22}a_{33} - a_{32}a_{23} \right) \right]^{-1} \left[a_{22}s^{(3)} - a_{32}s^{(2)} - \left(a_{31}a_{22} - a_{21}a_{32} \right) \right].$$

In similar way, putting new $\hat{\sigma}^{2(ht)}_{\eta(3)}$ in $\hat{\sigma}^{2(ht)}_{u(3)}$ we obtain

$$\hat{\sigma}^{2(ht)}_{u(3)} = \frac{1}{a_{22}}\left[s^{(2)} - a_{21} - a_{23}\hat{\sigma}^{2(ht)}_{\eta(3)} \right]$$

$$\Rightarrow \hat{\sigma}^{2(ht)}_{u(3)} = \frac{s^{(2)}}{a_{22}} - \frac{a_{21}}{a_{22}} - \frac{a_{23}}{a_{22}\left(a_{22}a_{33} - a_{32}a_{23} \right)}\left[a_{22}s^{(3)} - a_{32}s^{(2)} - \left(a_{31}a_{22} - a_{21}a_{32} \right) \right]$$

$$\Rightarrow \hat{\sigma}^{2(ht)}_{u(3)} = \left(a_{22}a_{33} - a_{32}a_{23} \right)^{-1} \times \left[\begin{array}{l} \left\{ \dfrac{a_{22}a_{33} - a_{32}a_{23}}{a_{22}} + \dfrac{a_{32}a_{23}}{a_{22}} \right\} s^{(2)} - a_{23}s^{(3)} \\ - \left\{ \dfrac{a_{21}\left(a_{22}a_{33} - a_{32}a_{23} \right)}{a_{22}} - \dfrac{a_{23}\left(a_{31}a_{22} - a_{21}a_{32} \right)}{a_{22}} \right\} \end{array} \right]$$

$$\Rightarrow \hat{\sigma}^{2(ht)}_{u(3)} = \left(a_{22}a_{33} - a_{32}a_{23} \right)^{-1} \times \left[a_{33}s^{(2)} - a_{23}s^{(3)} - \left(a_{21}a_{33} - a_{31}a_{23} \right) \right].$$

The terms involved in $\hat{\sigma}^{2(ht)}_{\eta(3)}$ and $\hat{\sigma}^{2(ht)}_{u(3)}$ can be simplified as

$$a_{22}a_{33} - a_{32}a_{23} = \frac{n - \overline{n}_0^{(2)}}{C_s - 1}\frac{n - \overline{n}_0^{(3)}}{D - 1} - \frac{\sum_{i \in s}\overline{n}_{0i}^{(2)} - \overline{n}_0^{(2)}}{D - 1}\frac{n - \overline{n}_0^{(3)}}{C_s - 1} = \frac{n - \overline{n}_0^{(3)}}{C_s - 1}\frac{n - \sum_{i \in s}\overline{n}_{0i}^{(2)}}{D - 1},$$

$$a_{31}a_{22}-a_{21}a_{32}=\frac{1}{D-1}\sum_{i\in s}\left(\frac{1}{n_i}-\frac{1}{n}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}\,\frac{n-\bar{n}_0^{(2)}}{C_s-1}-\frac{1}{C_s-1}\sum_{ij\in s}\left(\frac{1}{n_{ij}}-\frac{1}{n}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}\,\frac{\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}}{D-1}$$

$$=\frac{1}{(C_s-1)(D-1)}\left\{\left(n-\bar{n}_0^{(2)}\right)\sum_{i\in s}\left(\frac{1}{n_i}-\frac{1}{n}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}-\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\sum_{ij\in s}\left(\frac{1}{n_{ij}}-\frac{1}{n}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}\right\},$$

$$a_{21}a_{33}-a_{31}a_{23}=\frac{1}{C_s-1}\sum_{ij\in s}\left(\frac{1}{n_{ij}}-\frac{1}{n}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}\,\frac{n-\bar{n}_0^{(3)}}{D-1}-\frac{1}{D-1}\sum_{i\in s}\left(\frac{1}{n_i}-\frac{1}{n}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}\,\frac{n-\bar{n}_0^{(3)}}{C_s-1}$$

$$=\frac{n-\bar{n}_0^{(3)}}{(C_s-1)(D-1)}\left\{\sum_{ij\in s}\left(\frac{1}{n_{ij}}-\frac{1}{n}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}-\sum_{i\in s}\left(\frac{1}{n_i}-\frac{1}{n}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}\right\}.$$

Then the ultimate estimators of cluster- and area-level variance components become

$$\hat{\sigma}^{2(ht)}_{u(3)}=\frac{C_s-1}{n-\bar{n}_0^{(3)}}\frac{D-1}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}\left[\begin{array}{l}(D-1)^{-1}\left(n-\bar{n}_0^{(3)}\right)s^{(2)}-(C_s-1)^{-1}\left(n-\bar{n}_0^{(3)}\right)s^{(3)}-\\ \dfrac{n-\bar{n}_0^{(3)}}{(C_s-1)(D-1)}\left\{\sum_{ij\in s}\left(\frac{1}{n_{ij}}-\frac{1}{n}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}-\sum_{i\in s}\left(\frac{1}{n_i}-\frac{1}{n}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}\right\}\end{array}\right]$$

$$=\frac{1}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}\left[-\sum_{ij\in s}\left(\frac{1}{n_{ij}}-\frac{1}{n}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}+\sum_{i\in s}\left(\frac{1}{n_i}-\frac{1}{n}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}+(C_s-1)s^{(2)}-(D-1)s^{(3)}\right],$$

$$\hat{\sigma}^{2(ht)}_{\eta(3)}=\frac{C_s-1}{n-\bar{n}_0^{(3)}}\frac{D-1}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}\left[\begin{array}{l}(C_s-1)^{-1}\left(n-\bar{n}_0^{(2)}\right)s^{(3)}-(D-1)^{-1}\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)s^{(2)}\\ -\dfrac{1}{(C_s-1)(D-1)}\left\{\begin{array}{l}\left(n-\bar{n}_0^{(2)}\right)\sum_{i\in s}\left(\frac{1}{n_i}-\frac{1}{n}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}\\ -\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\sum_{ij\in s}\left(\frac{1}{n_{ij}}-\frac{1}{n}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}\end{array}\right\}\end{array}\right]$$

$$=\frac{\left(n-\bar{n}_0^{(3)}\right)^{-1}}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}\left[\begin{array}{l}\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\sum_{ij\in s}\left(\frac{1}{n_{ij}}-\frac{1}{n}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}-\left(n-\bar{n}_0^{(2)}\right)\sum_{i\in s}\left(\frac{1}{n_i}-\frac{1}{n}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}\\ -(C_s-1)\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)s^{(2)}+(D-1)\left(n-\bar{n}_0^{(2)}\right)s^{(3)}\end{array}\right]$$

which are similar to previous estimators.

A.8 Motivation of Modified ELL Methodology (MELL) under Misspecified Model with Heteroskedastic (HT) Level-one Errors

Under the considered HT 3-level model (6), the variances of cluster-specific mean $\bar{Y}_{ij} = M_{ij}^{-1}\sum_{k=1}^{N_{ij}} m_{ijk} y_{ijk}$ and area-specific mean $\bar{Y}_i = M_i^{-1}\sum_{j=1}^{C_i}\sum_{j=1}^{N_{ij}} m_{ijk} y_{ijk}$ can be expressed as

$$\begin{aligned}
\mathrm{Var}_{(3)}\left(\bar{Y}_{ij}\right) &= M_{ij}^{-2}\left[\sum_{k=1}^{N_{ij}} m_{ijk}^2 \mathrm{Var}_{(3)}\left(y_{ijk}\right) + \sum_{k\neq k'}^{N_{ij}} m_{ijk} m_{ijk'} \mathrm{Cov}_3\left(y_{ijk}, y_{ijk'}\right)\right] \\
&= M_{ij}^{-2}\left[\left(\sigma_{\eta(3)}^{2(ht)} + \sigma_{u(3)}^{2(ht)}\right)\sum_{k=1}^{N_{ij}} m_{ijk}^2 + \sum_{k=1}^{N_{ij}} m_{ijk}^2 \sigma_{\varepsilon(3),ijk}^2 + \left(\sigma_{\eta(3)}^{2(ht)} + \sigma_{u(3)}^{2(ht)}\right)\sum_{k\neq k'}^{N_{ij}} m_{ijk} m_{ijk'}\right] \\
&= M_{ij}^{-2}\left[\left(\sigma_{\eta(3)}^{2(ht)} + \sigma_{u(3)}^{2(ht)}\right)\left\{\sum_{k=1}^{N_{ij}} m_{ijk}^2 + \sum_{k\neq k'}^{N_{ij}} m_{ijk} m_{ijk'}\right\} + \sum_{k=1}^{N_{ij}} m_{ijk}^2 \sigma_{\varepsilon(3),ijk}^2\right] \\
&= \sigma_{\eta}^{2(ht)} + \sigma_{u}^{2(ht)} + M_{ij}^{-2}\sum_{k=1}^{N_{ij}} m_{ijk}^2 \sigma_{\varepsilon(3),ijk}^2 \text{, and}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}_{(3)}\left(\bar{Y}_{i}\right) &= M_{i}^{-2}\left[\sum_{j=1}^{C_i} M_{ij}^2\left\{\sigma_{\eta(3)}^{2(ht)} + \sigma_{u(3)}^{2(ht)} + M_{ij}^{-2}\sum_{k=1}^{N_{ij}} m_{ijk}^2 \sigma_{\varepsilon(3),ijk}^2\right\} + \sigma_{\eta(3)}^{2}\sum_{j\neq j'}^{C_i} M_{ij} M_{ij'}\right] \\
&= M_{i}^{-2}\left[\sigma_{\eta(3)}^{2(ht)}\left\{\sum_{j=1}^{C_i} M_{ij}^2 + \sum_{j\neq j'}^{C_i} M_{ij} M_{ij'}\right\} + \sigma_{u(3)}^{2(ht)}\sum_{j=1}^{C_i} M_{ij}^2 + \sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}} m_{ijk}^2 \sigma_{\varepsilon(3),ijk}^2\right] \\
&= \sigma_{\eta(3)}^{2(ht)} + \sigma_{u(3)}^{2(ht)} M_i^{-2}\sum_{j=1}^{C_i} M_{ij}^2 + M_i^{-2}\sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}} m_{ijk}^2 \sigma_{\varepsilon(3),ijk}^2 = \sigma_{\eta(3)}^{2(ht)} + \sigma_{u(3)}^{2(ht)} \bar{m}_{Ui}^{(2)} + \xi_{ijk}^{(3)}
\end{aligned}$$

where $\mathrm{Cov}\left(\bar{Y}_{ij}, \bar{Y}_{ij'}\right) = \sigma_{\eta(3)}^{2(ht)}$, $\bar{m}_{Ui}^{(2)} = M_i^{-2}\sum_{j=1}^{C_i} M_{ij}^2$, and $\xi_{ijk}^{(3)} = M_i^{-2}\sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}} m_{ijk}^2 \sigma_{\varepsilon(3),ijk}^2$. Under the HT 3-level population model (6), we have already shown that $\hat{\sigma}_{\eta(3)}^{2(ht)}$ and $\hat{\sigma}_{u(3)}^{2(ht)}$ given in **Appendix A.7** are unbiased and consistent estimators of $\sigma_{\eta(3)}^{2(ht)}$ and $\sigma_{u(3)}^{2(ht)}$ under the assumption of known HH-level variances $\sigma_{\varepsilon(3),ijk}^2$. Assume a consistent estimator of $\sigma_{\varepsilon(3),ijk}^2$ can be obtained via ELL parametric method or proposed non-parametric method (discussed in Chapter Five). Then a plug-in estimator of $\mathrm{Var}_{(3)}\left(\bar{Y}_i\right)$ can be considered as

$$\hat{V}_{(3)}\left(\bar{Y}_i\right) = \hat{\sigma}^{2(ht)}_{\eta(3)} + \hat{\sigma}^{2(ht)}_{u(3)}\bar{m}^{(2)}_{Ui} + \hat{\xi}^{(3)}_{ijk}$$

where $\hat{\xi}^{(3)}_{ijk} = M_i^{-2}\sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}} m^2_{ijk}\hat{\sigma}^2_{\varepsilon(3),ijk}$ and $\hat{\sigma}^2_{\varepsilon(3),ijk} = \phi\left(x_{ijk}\right)$ is a consistent estimator of $\sigma^2_{\varepsilon(3),ijk}$ which will be a function of explanatory variables or predicted values. In the similar way, under the 2-level HT model (5), the variance of area-specific mean $\mathrm{Var}_{(2)}\left(\bar{Y}_i\right)$ and its plug-in estimator can be written as

$$\mathrm{Var}_{(2)}\left(\bar{Y}_i\right) = \sigma^{2(ht)}_{u(2)}\bar{m}^{(2)}_{Ui} + \xi^{(2)}_{ijk} \text{ and } \hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right) = \hat{\sigma}^{2(ht)}_{u(2)}\bar{m}^{(2)}_{Ui} + \hat{\xi}^{(2)}_{ijk}$$

respectively where $\xi^{(2)}_{ijk} = M_i^{-2}\sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}} m^2_{ijk}\sigma^2_{\varepsilon(2),ijk}$ and $\hat{\xi}^{(2)}_{ijk} = M_i^{-2}\sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}} m^2_{ijk}\hat{\sigma}^2_{\varepsilon(2),ijk}$. The variance component estimators $\hat{\sigma}^{2(ht)}_{u(2)}$ and $\hat{\sigma}^2_{\varepsilon(2),ijk}$ are consistent estimator of $\sigma^{2(ht)}_{u(2)}$ and $\sigma^2_{\varepsilon(2),ijk}$ under the HT 2-level population model (5) with the assumption of known $\sigma^2_{\varepsilon(2),ijk}$. Expectations of the variance component estimator $\hat{\sigma}^{2(ht)}_{u(2)}$ under the true HT 3-level model (6) is

$$\begin{aligned}
E_3\left[\hat{\sigma}^{2(ht)}_{u(2)}\right] &= \frac{1}{n-\bar{n}^{(2)}_0}\left\{\left(C_s-1\right)E_3\left[s^{(2)}\right] - \sum_{ij\in s}\left(n_{ij}^{-1}-n\right)\sum_k \sigma^2_{\varepsilon(3),ijk}\right\} \\
&= \frac{1}{n-\bar{n}^{(2)}_0}\left\{\sum_{ij\in s}\left(n_{ij}^{-1}-n^{-1}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk} + \left(n-\bar{n}^{(2)}_0\right)\sigma^{2(ht)}_{u(3)} + \left(n-\bar{n}^{(3)}_0\right)\sigma^{2(ht)}_{\eta(3)} - \sum_{ij\in s}\left(n_{ij}^{-1}-n\right)\sum_k \sigma^2_{\varepsilon(3),ijk}\right\} \\
&= \frac{1}{n-\bar{n}^{(2)}_0}\left\{\left(n-\bar{n}^{(2)}_0\right)\sigma^{2(ht)}_{u(3)} + \left(n-\bar{n}^{(3)}_0\right)\sigma^{2(ht)}_{\eta(3)}\right\} = \sigma^{2(ht)}_{u(3)} + \frac{n-\bar{n}^{(3)}_0}{n-\bar{n}^{(2)}_0}\sigma^{2(ht)}_{\eta(3)}.
\end{aligned}$$

In the similar way, the expectations of the variance components estimators $\hat{\sigma}^{2(ht)}_{\eta(3)}$ and $\hat{\sigma}^{2(ht)}_{u(3)}$ under the true HT 3-level model are

$$E_3\left[\hat{\sigma}_{u(3)}^{2(h)}\right]=\frac{1}{n-\sum_{i\in s}\bar{n}_{0i}^{(2)}}\left[\begin{array}{l}-\left\{\sum_{ij\in s}\left(n_{ij}^{-1}-n^{-1}\right)\sum_{k\in s}\sigma_{\varepsilon(3),ijk}^2-\sum_{i\in s}\left(n_i^{-1}-n^{-1}\right)\sum_{jk\in s}\sigma_{\varepsilon(3),ijk}^2\right\}+\\ \left\{\sum_{ij\in s}\left(n_{ij}^{-1}-n^{-1}\right)\sum_{k\in s}\sigma_{\varepsilon(3),ijk}^2+\left(n-\bar{n}_0^{(2)}\right)\sigma_u^{2(ht)}+\left(n-\bar{n}_0^{(3)}\right)\sigma_\eta^{2(ht)}\right\}-\\ \left\{\begin{array}{l}\sum_{i\in s}\left(n_i^{-1}-n^{-1}\right)\sum_{jk\in s}\sigma_{\varepsilon(3),ijk}^2+\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\sigma_u^{2(ht)}\\ +\left(n-\bar{n}_0^{(3)}\right)\sigma_\eta^{2(ht)}\end{array}\right\}\end{array}\right]=\sigma_{u(3)}^{2(ht)}$$

$$E_3\left[\hat{\sigma}_{\eta(3)}^2\right]=\frac{\left[\begin{array}{l}\varphi_{31}+\frac{\varphi_{31}-\varphi_{32}}{\varphi_{00}}\left\{\sum_{ij\in s}\left(n_{ij}^{-1}-n^{-1}\right)\sum_{k\in s}\sigma_{\varepsilon(3),ijk}^2+\left(n-\bar{n}_0^{(2)}\right)\sigma_u^{2(ht)}+\left(n-\bar{n}_0^{(3)}\right)\sigma_\eta^{2(ht)}\right\}\\ +\left(n-\bar{n}_0^{(2)}\right)\left\{\sum_{i\in s}\left(n_i^{-1}-n^{-1}\right)\sum_{jk\in s}\sigma_{\varepsilon(3),ijk}^2+\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\sigma_u^{2(ht)}+\left(n-\bar{n}_0^{(3)}\right)\sigma_\eta^{2(ht)}\right\}\end{array}\right]}{\left(n-\bar{n}_0^{(3)}\right)\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)}$$

$$=\frac{\left[\begin{array}{l}\varphi_{31}-\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\left\{\left[\sum_{ij\in s}\left(n_{ij}^{-1}-n^{-1}\right)\sum_{k\in s}\sigma_{\varepsilon(3),ijk}^2\right]+\left(n-\bar{n}_0^{(2)}\right)\sigma_u^{2(ht)}+\left(n-\bar{n}_0^{(3)}\right)\sigma_\eta^{2(ht)}\right\}\\ +\left(n-\bar{n}_0^{(2)}\right)\left\{\sum_{i\in s}\left(n_i^{-1}-n^{-1}\right)\sum_{jk\in s}\sigma_{\varepsilon(3),ijk}^2+\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\sigma_u^{2(ht)}+\left(n-\bar{n}_0^{(3)}\right)\sigma_\eta^{2(ht)}\right\}\end{array}\right]}{\left(n-\bar{n}_0^{(3)}\right)\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)}$$

$$=\frac{\left[\begin{array}{l}\varphi_{31}-\left\{\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\sum_{ij\in s}\left(n_{ij}^{-1}-n^{-1}\right)\sum_{k\in s}\sigma_{\varepsilon,ijk}^2-\left(n-\bar{n}_0^{(2)}\right)\sum_{i\in s}\left(n_i^{-1}-n^{-1}\right)\sum_{jk\in s}\sigma_{\varepsilon,ijk}^2\right\}\\ +\left\{\left(n-\bar{n}_0^{(2)}\right)\left(n-\bar{n}_0^{(3)}\right)-\left(\sum_{i\in s}\bar{n}_{0i}^{(2)}-\bar{n}_0^{(2)}\right)\left(n-\bar{n}_0^{(3)}\right)\right\}\sigma_\eta^{2(ht)}\end{array}\right]}{\left(n-\bar{n}_0^{(3)}\right)\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)}$$

$$=\left(n-\bar{n}_0^{(3)}\right)^{-1}\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)^{-1}\times\left[\varphi_{31}-\varphi_{31}+\left(n-\bar{n}_0^{(3)}\right)\left(n-\sum_{i\in s}\bar{n}_{0i}^{(2)}\right)\sigma_\eta^{2(ht)}\right]=\sigma_\eta^{2(ht)}.$$

Then the expectations of $\hat{\mathrm{V}}_{(3)}\left(\bar{Y}_i\right)$ and $\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)$ under the HT 3-level model are

$$E_3\left[\hat{\mathrm{V}}_{(3)}\left(\bar{Y}_i\right)\right]\approx\sigma_{\eta(3)}^{2(ht)}+\sigma_{u(3)}^{2(ht)}\bar{m}_{Ui}^{(2)}+\xi_{ijk}^{(3)}\text{ and }E_3\left[\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)\right]\approx\left\{R\sigma_{\eta(3)}^{2(ht)}+\sigma_{u(3)}^{2(ht)}\right\}\bar{m}_{Ui}^{(2)}+\xi_{ijk}^{(3)}$$

where $\bar{m}_{Ui}^{(2)} < 1$ and $R < 1$ lead to $E_3\left[\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)\right] < E_3\left[\hat{\mathrm{V}}_{(3)}\left(\bar{Y}_i\right)\right]$ under the assumption of $E_3\left[\hat{\xi}_{ijk}^{(3)}\right] \approx E_3\left[\hat{\xi}_{ijk}^{(2)}\right] \approx \xi_{ijk}^{(3)}$. As the HM case, an area-specific adjustment can be made to make $\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)$ unbiased or approximately unbiased under the 3-level model. This leads to an adjusted estimator as

$$\hat{\mathrm{V}}_{(2)}^{M}\left(\bar{Y}_i\right) = \left\{\left(1/\bar{m}_{Ui}^{(2)}\right)\hat{\sigma}_{\eta(3)}^{2(ht)} + \hat{\sigma}_{u(3)}^{2(ht)}\right\}\bar{m}_{Ui}^{(2)} + \hat{\xi}_{ijk}^{(2)}$$

which is approximately unbiased under the true 3-level model. When HH weights are equal, the adjusted estimator becomes

$$\hat{\mathrm{V}}_{(2)}^{M}\left(\bar{Y}_i\right) = \left\{\left(1/\bar{n}_{Ui}^{(2)}\right)\hat{\sigma}_{\eta(3)}^{2(ht)} + \hat{\sigma}_{u(3)}^{2(ht)}\right\}\bar{n}_{Ui}^{(2)} + \hat{\sigma}_{\varepsilon(2)}^{2(ht)} N_i^{-1}.$$

where $\bar{n}_{Ui}^{(2)} = N_i^{-2}\sum_{j=1}^{C_i} N_{ij}^2$ is an-area-specific term based on the number of HHs. In the similar way of **Appendix A.5**, the adjustment factors for the cluster-variance component can be considered as below:

$$k_1 = \hat{\sigma}_{u(2)}^{-2(ht)}\left[\hat{\sigma}_{\eta(3)}^{2(ht)} D_s^{-1}\sum_{i=1}^{D_s}\left(1/\bar{m}_{si}^{(2)}\right) + \hat{\sigma}_{u(3)}^{2(ht)}\right],\; k_2 = \hat{\sigma}_{u(2)}^{-2(ht)}\left[\hat{\sigma}_{\eta(3)}^{2(ht)} D^{-1}\sum_{i=1}^{D}\left(1/\bar{m}_{Ui}^{(2)}\right) + \hat{\sigma}_{u(3)}^{2(ht)}\right] \text{, and}$$

$$k_{3(p)} = \hat{\sigma}_{u(2)}^{-2(ht)}\left[\hat{\sigma}_{\eta(3)}^{2(ht)} D_{(p)}^{-1}\sum_{i=1}^{D_{(p)}}\left(1/\bar{m}_{Ui}^{(2)}\right) + \hat{\sigma}_{u(3)}^{2(ht)}\right];\;\; p = 1,\ldots,P$$

where $\bar{m}_{si}^{(2)} = m_i^2 \Big/ \sum_{j=1}^{C_i} m_{ij}^2$, D_s is number sampled areas, m_i is total members observed in i^{th} sampled area, m_{ij} is total members observed in j^{th} sampled cluster of i^{th} sampled area. The adjusted factor will be based on the number of HHs if no HH-level weights are considered.

APPENDIX B

R Scripts

R scripts for the contributed four chapters are available in the following links. Bangladesh datasets, which are used in chapters Three and Six, are not attached with the link.

Chapter Three: https://github.com/sumon148/Chapter-Three.git

Chapter Four: https://github.com/sumon148/Chapter-Four.git

Chapter Five: https://github.com/sumon148/Chapter-Five.git

Chapter Six: https://github.com/sumon148/Chapter-Six.git

Contact information of Bangladesh Bureau of Statistics (BBS) for Bangladesh Census and Survey Datasets: http://www.bbs.gov.bd/PageWebMenuContent.aspx?MenuKey=62