Statistics and Applications {ISSN 2452-7395 (online)}
Volume 16, No. 1, 2018 (New Series), pp 375-397

Robust Mean Squared Error Estimation for ELL-based Poverty Estimates Under Heteroskedasticity - An Application to Poverty Estimation in Bangladesh

Sumonkanti Das[1], Hukum Chandra[2] and Ray Chambers[3]
[1]*Shahjalal University of Science and Technology, Bangladesh*
[2]*ICAR-Indian Agricultural Statistics Research Institute, India*
[3]*University of Wollongong, Australia*

Received: February 13, 2018; Revised: June 24, 2018; Accepted: June 25, 2018

Abstract

The ELL poverty mapping method (Elbers, Lanjouw and Lanjouw, 2003) has been criticized because of the risk of underestimating the mean squared error (MSE) of the corresponding poverty estimates when the area homogeneity assumption is violated. Das and Chambers (2017) describe a robust ELL-based MSE estimation approach that assumes unit-level homoscedasticity. However, applications of the ELL approach are typically based on an assumption of unit-level heteroskedasticity. Ignoring this behavior can lead to underestimation of the MSE of the poverty estimates generated by the ELL approach. This paper extends the idea of Das and Chambers to ELL-based MSE estimation assuming unit-level heteroskedasticity. The proposed method is then applied to poverty estimation in Bangladesh in order to evaluate its usefulness in a realistic data scenario.

Key words: Homoskedasticity, Small Area Estimation, Poverty Mapping, Stratification

1. Introduction

1.1 Poverty estimation in Bangladesh

Since 1983-84 Bangladesh poverty rates have been estimated at national and divisional levels using data collected in the Household Income and Expenditure Survey. To show the actual variation in poverty incidence (HCR) between local administrative units, the Bangladesh Bureau of Statistics (BBS) in conjunction with United Nation World Food Program (UNWFP) conducted a poverty mapping study using the Bangladesh 2001 Population and Housing Census (hereafter referred as Census 2001) and the Bangladesh 2000 Household Income and Expenditure Survey (hereafter referred as HIES 2000) datasets (BBS and UNWFP, 2004). This poverty map was updated using the HIES 2005 data (WB, BBS and WFP, 2009). Though the national level poverty incidence was about 40 percent in 2005 (BBS, 2011), sub-district level poverty incidence varied from about 0 to 55 percent (WB, BBS and UNWFP, 2009). The ELL method developed by the World Bank was used to obtain these sub-district poverty estimates. Both poverty maps show that areas close to the capital city Dhaka have lower poverty rates but the actual size of the poor population in these areas is large. In comparison, the sub-districts in the Chittagong Hill Tracks (south-eastern part of Bangladesh) have high poverty incidence but their population sizes are relatively small. On

Corresponding Author: Sumonkanti Das
E-mail: sumonkanti-sta@sust.edu

the other hand, the sub-districts in the northern part (well known for their seasonal food insecurity) have large population sizes as well as higher poverty rates (WB, BBS, and WFP, 2009).

1.2 Data sources and model specifications

In the first poverty mapping study of Bangladesh (BBS and UNWFP, 2004), hereafter referred as the BBS-2004 study, the BBS systematically sampled 5% of the enumeration areas (EAs) from each sub-district of Census 2001 instead of using the full census data. This 5% of Census 2001 covers 5 divisions, 64 districts (*Zila*), 507 sub-districts (*Upzila*), 12908 EAs, 1258240 households (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 Census 2001 and the 5% of Census 2001 are detailed in BBS and UNWFP (2004). The HIES 2000 is used as the accompanying economic survey dataset required by the ELL methodology to fit models that are then used to simulate poverty data based on the 5% of Census 2001. This survey dataset covers 295 out of 507 sub-districts. The sample for HIES 2000 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 should be noted that about two-thirds of sampled sub-districts (222 out of 295) had just a single sample cluster and so sub-district sample sizes are very small. The sampling design and the structure of HIES 2000 data are detailed in BBS-2004 study. In this paper we will use the HIES 2000 and the 5% of Census 2001 datasets to examine the empirical performance of the estimators proposed in what follows.

The main task when applying ELL methodology for poverty mapping is to identify appropriate explanatory variables at different levels in the hierarchies present in the population data in order to reduce overall residual variation given the model specification. In the BBS-2004 study, 27 HH specific and 3 sub-district specific explanatory variables are used to fit a two-level random effects model with HH and cluster as levels 1 and 2 respectively. The between-area variation is ignored, most likely because about 75% of the sampled sub-districts have only a single sampled cluster. In this study, we examine the contributions to overall variability due to the presence of cluster and sub-district hierarchies in the survey data by fitting both 2-level (2L) and 3-level (3L) models to these data with sub-district corresponding to level-three in the data hierarchy. We note that about 20% of the variation in HH expenditure is due to between cluster and between sub-district variability (see set-1 in Table 1). In particular, the contribution of sub-district level variability is negligible (about 5%) but statistically significant (p-value < 0.0001). The presence of negligible between-area variation and also lack of sufficient survey data to fit an appropriate 3-level model essentially enforce implementation of the standard 2-level model-based ELL method, which ignores sub-district level random effects.

Table 1: Estimate of variance component parameters obtained by method of moment estimation under 2-level (2L) and 3-level (3L) homoskedastic models

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 HIES 2000 data set out in Table 2 suggests that the sampled sub-districts with a single sampled cluster are mainly rural (197 out of 222 rural clusters), while sampled sub-districts with multiple sampled clusters are mainly urban (183 out of 220 urban clusters). In order to investigate the presence of significant between area variability in the urban parts of Bangladesh, a new sample dataset (hereinafter, set-2) was created containing only sampled sub-districts with multiple clusters. This allowed both 2-level and 3-level random effects models to be fitted and the significance of between-area variability to be tested. The results for set-2 in Table 1 confirm the presence of statistically significant sub-district level random effects in the urban parts of Bangladesh. Both survey and census datasets indicate that a significant number of sub-districts have both urban and rural parts (Table 2) and so it is reasonable to partition each dataset into rural and urban sub-sets, and then estimate between area variability using data from residential areas. In the rural sample data (hereafter set-3), only 35 out of 232 sub-districts have multiple clusters, while 71 out of 96 sub-districts in the urban sample of HIES 2000 data (hereafter set-4) have multiple clusters. The results for set-4 in Table 1 confirm the significance of between area variability in the urban sample. These results therefore questions regarding the suitability of applying naïve ELL methodology (i.e. based on a 2-level model) to the data set as a whole since this might not capture the actual between area variability present in the HIES 2000 data.

Table 2: Distribution of household (HH), cluster and sub-districts (area) by type of residence in HIES 2000 and Census 2001

Type of Residence	Census 2001			HIES 2000							
				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

1.3 Suggested alternatives to ELL methodology

Note that the ELL method of poverty estimation is typically based on an assumption of cluster-heterogeneity. If this assumption is violated, the method produces approximately unbiased poverty estimates, but with MSE estimates that are biased low, which in turn leads to poor coverage rates for the poverty estimates (Tarozzi and Deaton, 2009). The MSE estimates help to prioritize the small areas according to their corresponding poverty estimates. An area with an underestimated MSE will receive less priority compared to an area with the same value of the associated poverty estimate, but with MSE estimated correctly. Estimation of MSE depends on several issues including specification of the fitted model, unexplained variation in the response variable after accounting for the explanatory variables, estimated variation at higher levels (e.g. cluster/area levels), and the population size of a small area. In typical applications of the ELL method, HH specific random errors are usually assumed to be heteroskedastic (HT) while the cluster random effects are usually assumed to be homoskedastic (HM), in large part because there are often few clusters per sampled area in the survey data. Ignoring level one heteroskedasticity generally leads to biased estimates of distribution functions and hence biased estimates of the Foster, Greer and Thorbecke (1984, hereafter FGT) measures of poverty incidence. In these cases the HM cluster-specific variance component is estimated assuming heteroskedasticity at HH-level, using a HH level heteroskedasticity model (typically referred as an "alpha" model) for between HH variances which is a function of the potential explanatory variables. Under either HM or HT level-one errors, a 2-level nested-error regression model is then fitted ignoring area-specific random effects. It is well known that if between area variation remains significant after incorporating area-level contextual variables in the regression model, then the ELL method leads to estimated MSEs that are biased low, and hence under coverage of the true population values of the poverty measures. The modified ELL (MELL) method of Das and Chambers (2017) will perform better in such situations. But the MELL is based on an assumption of HM level one errors. Consequently the MELL needs to be modified in order to account for heteroskedastic HH-level errors. Alternatively, one could consider implementing either the optimistic or the conservative ELL methods (Elbers, *et al.*, 2008; World Bank, 2013) assuming HH-level heteroskedasticity at HH-level. Unfortunately, however, both these methods assume HM random effects for both 2-level and 3-level random effects models fitted to the survey data. Furthermore, both these approaches perform poorly when there is between area variability (Das and Chambers, 2017). In this article, we develop and test a modification to the MELL approach assuming heteroskedasticity at HH-level. In the next Section we review the ELL method and its modifications (including the modified MELL) for capturing potential between area variability under both HM and HT household-level errors. In Section 3 we then demonstrate the application of these methods using the HIES 2000 data of Bangladesh. Section 4 contains a discussion of our empirical results and summarizes our major findings. Finally, in Section 5 we provide some concluding remarks and indicate further research avenues.

2. The ELL Methodology and Its Extensions

To start, let E_{ijk} and m_{ijk} denote the per capita household expenditure (i.e. welfare measure) and the number of family members (i.e. family size) respectively of the k^{th} household (HH) belonging to the j^{th} cluster in the i^{th} small area. Let $y_{ijk} = \log(E_{ijk})$ denote the log transformed per capita household expenditure. The area-specific FGT poverty indicators are then calculated as $F_{\alpha i} = M_i^{-1} \sum_{j=1}^{C_i} \sum_{k=1}^{N_{ij}} m_{ijk} \left(1 - E_{ijk}/t\right)^{\alpha} I\left(E_{ijk} < t\right);\ \alpha = 0,1,2$ where $M_i = \sum_j^{C_I} \sum_k^{N_{ij}} m_{ijk} = \sum_j^{C_i} M_{ij}$ and C_i are respectively the total number of individuals and clusters in i^{th} area; $M_{ij} = \sum_k^{N_{ij}} m_{ijk}$ and N_{ij} are respectively the total number of individuals (population) and households (HHs) in the j^{th} cluster of the i^{th} area. When HH-specific weights are ignored or equal, the FGT indicator becomes

$F_{\alpha i} = N_i^{-1} \sum_{j=1}^{C_i} \sum_{k=1}^{N_{ij}} \left(1 - E_{ijk}/t\right)^{\alpha} I\left(E_{ijk} < t\right)$ with $N_i = \sum_j^{C_i} N_{ij}$. Here, t is the threshold for E_{ijk} under which a person/HH is considered as being "in poverty". In the standard ELL approach, a 2-level nested error regression model is considered assuming HHs at level-one and clusters at level-two as

$$y_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} + u_{ij} + \varepsilon_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)} + u_{ij} + \sigma_{\varepsilon(2)} \gamma_{ijk}, \quad (1)$$

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-specific and HH-specific random errors, $i = 1, 2, \ldots, D$; $j = 1, 2, \ldots, C_i$; $k = 1, 2, \ldots, N_{ij}$. Here, $\gamma_{ijk} \sim N(0,1)$. The subscript (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 (1) can be expressed as

$$y_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)}^{(ht)} + u_{ij} + \varepsilon_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(2)}^{(ht)} + u_{ij} + \sigma_{\varepsilon(2),ijk} \gamma_{ijk}, \quad (2)$$

where $u_{ij} \sim N\left(0, \sigma_{u(2)}^{2(ht)}\right)$; $\varepsilon_{ijk} \sim N\left(0, \sigma_{\varepsilon(2),ijk}^2\right)$and $\gamma_{ijk} \sim N(0,1)$. Here the superscript *ht* stands for heteroskedasticity. When an additional area-specific random effect is also assumed in the above two models, the corresponding 3-level models can be expressed

$$y_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(3)} + \eta_i + u_{ij} + \varepsilon_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(3)} + \eta_i + u_{ij} + \sigma_{\varepsilon(3)} \gamma_{ijk}, \quad (3)$$

with $\eta_i \sim N\left(0, \sigma_{\eta(3)}^2\right)$, $u_{ij} \sim N\left(0, \sigma_{u(3)}^2\right)$, $\varepsilon_{ijk} \sim N\left(0, \sigma_{\varepsilon(3)}^2\right)$ and $\gamma_{ijk} \sim N(0,1)$, and

$$y_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(3)}^{(ht)} + \eta_i + u_{ij} + \varepsilon_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\beta}_{(3)}^{(ht)} + \eta_i + u_{ij} + \sigma_{\varepsilon(3),ijk} \gamma_{ijk}, \quad (4)$$

with $\eta_i \sim N\left(0, \sigma_{\eta(3)}^{2(ht)}\right)$, $u_{ij} \sim N\left(0, \sigma_{u(3)}^{2(ht)}\right)$, $\varepsilon_{ijk} \sim N\left(0, \sigma_{\varepsilon(3),ijk}\right)$ and $\gamma_{ijk} \sim N(0,1)$.

If the model (3) is true but the corresponding 2-level model (1) is used to implement the standard 2-level ELL, the resulting MSE estimates will be underestimates. A similar problem will occur if the HT 2-level model (2) is used instead of HT 3-level model (4) in the ELL. Since the area homogeneity assumption of the ELL method is violated in both situations, it is necessary to capture the level-three variability in order to prevent subsequent MSE underestimation. The MELL method of Das and Chambers (2017) is designed to account for HH-level heteroskedasticity. Note that estimation of variance components is difficult when the HH random errors are assumed to be HT. Estimation methods for HM variance components and for HT error variances are first described in what follows and then the ELL based methods are discussed under both 2-level and 3-level working models assuming HT level-one errors. Finally the 2-level model-based ELL type estimators are modified to account for the ignored between area variability of a 3-level true model.

2.1. Variance component estimation under heteroskedasticity

The variance components are estimated via the method of moments (MM) approach under both homoscedasticity and heteroskedasticity. The variance components under 2-level and 3-level HM models can be easily estimated, which is not the case under heteroskedasticity. Under the 2-level HT model (2), the MM 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 $w_{ij} = n_{ij}/n$, $\hat{\bar{e}}_{ij.} = n_{ij}^{-1}\sum_k^{n_{ij}} \hat{e}_{ijk}$, $\hat{\bar{e}}_{...} = n^{-1}\sum_{jk\in s}\hat{e}_{ijk}$, $\hat{\tau}_{ij}^2 = n_{ij}^{-1}\left(n_{ij}-1\right)^{-1}\sum_k^{n_{ij}}\left(\hat{\varepsilon}_{ijk} - \hat{\bar{\varepsilon}}_{ij.}\right)^2$, $\hat{\bar{\varepsilon}}_{ij.} = n_{ij}^{-1}\sum_k^{n_{ij}}\hat{\varepsilon}_{ijk}$ and $\hat{e}_{ijk} = y_{ijk} - \hat{y}_{ijk}$. In the similar manner, the MM estimators of $\sigma_{u(3)}^{2(ht)}$ and $\sigma_{\eta(3)}^{2(ht)}$ under (3) can be obtained as

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

and

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

$$- \frac{\sum_{i\in s}\left(w_i^{-1}-1\right)\sum_{j\in s} w_{ij}^2\left\{\sum_{ij\in s} w_{ij}\left(\hat{\bar{e}}_{ij.} - \hat{\bar{e}}_{...}\right)^2 - \sum_{ij} w_{ij}\left(1-w_{ij}\right)\hat{\tau}_{ij}^2\right\}}{\left(\sum_{i\in s} w_i\left(1-w_i\right)\right)\left(\sum_{ij} w_{ij} - \sum_{i\in s} w_i^{-1}\sum_{j\in s} w_{ij}^2\right)}$$

where $w_i = n_i/n$, $\tau_i^2 = n_i^{-2}\sum_{jk\in s}\sigma_{\varepsilon(3),ijk}^2$ and $\hat{\bar{e}}_{i..} = n_i^{-1}\sum_{j\in s} n_{ij}\hat{\bar{e}}_{ij.}$. Negative values of the estimator $\hat{\sigma}_{\eta(3)}^{2(ht)}$ will be treated as zero. The derivations of these estimators with their properties are given in Appendices A.1 and A.2 respectively.

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

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

where $\hat{A} = 1.05\times$ maximum$\left\{\hat{e}_{ijk}^2\right\}$, $D_{ijk} = \exp(z_{ijk}^T\hat{\alpha})$ 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 Elbers et al. (2002) under 2-level HT working model. Under the 3-level model (4), the estimates of $\sigma_{\varepsilon(3),ijk}^2$ can be estimated by the ELL estimator using the corresponding estimated HH-level residuals.

2.2. The ELL Method

After fitting the regression model and obtaining the corresponding parameters, the second stage of the ELL method is to conduct either a parametric bootstrap (PB) or a non-parametric bootstrap (NPB) procedure to obtain the area-specific poverty estimates of interest and their corresponding estimated mean square errors (ESMEs). For both the PB or NPB procedures the basic steps are: (1) generate regression parameters $\boldsymbol{\beta}^*$ from a suitable sampling distribution, say the 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 simple random sampling with replacement (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 of interest 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 a large number of times say B=500 and then the mean and variance of the B estimates are considered as the final estimates and their MSEs respectively.

2.3. Modification of the ELL Method

The ELL approach based on a 2-level HM working model may produce underestimated MSE if the area variability is ignored. However, the approach can be modified to capture the potential area variability following the MELL approach of Das and Chambers (2017). This MELL approach is adapted here to allow for HT level-one random errors. The basis of the MELL methodology is the adjustment to the variance estimator of a 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 HM model (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 (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 expectation 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 $R = \left(n - \bar{n}^{(3)}_0\right)/\left(n - \bar{n}^{(2)}_0\right) < 1$.

An unbiased plug-in estimator of $\mathrm{Var}\left(\bar{Y}_i\right)$ is difficult to obtain under a multilevel model with HT level-1 random errors. However, a plug-in consistent estimator of this variance can be obtained if a consistent estimator of the HT error variances is available. Under the HT models (4) and (2), we have $\mathrm{Var}_{(3)}\left(\bar{Y}_i\right) = \sigma^{2(ht)}_{\eta(3)} + \sigma^{2(ht)}_{u(3)}\bar{m}^{(2)}_{Ui} + \xi^{(3)}_{ijk}$ 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 $\xi^{(2)}_{ijk} = M_i^{-2}\sum_{j=1}^{C_i}\sum_{k=1}^{N_{ij}} m^2_{ijk}\sigma^2_{\varepsilon(2),ijk}$.

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 an assumption of known HH-level error variances (see Appendices A.1 and A.2). Now suppose that the HH error variance estimators $\hat{\sigma}^2_{\varepsilon(2),ijk}$ and $\hat{\sigma}^2_{\varepsilon(3),ijk}$ are consistent estimators of $\sigma^2_{\varepsilon(2),ijk}$ and $\sigma^2_{\varepsilon(3),ijk}$ respectively. Then consistent plug-in estimators of $\mathrm{Var}_{(3)}\left(\bar{Y}_i\right)$ and $\mathrm{Var}_{(2)}\left(\bar{Y}_i\right)$ are $\hat{\mathrm{V}}_{(3)}\left(\bar{Y}_i\right) = \hat{s}^2_{h(3)} + \hat{s}^2_{u(3)}\bar{m}^{(2)}_{Ui} + \hat{x}^{(3)}_{ijk}$ and $\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right) = \hat{s}^2_{u(2)}\bar{m}^{(2)}_{Ui} + \hat{x}^{(2)}_{ijk}$ respectively 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 $E_3\left[\hat{\sigma}^{2(ht)}_{u(2)}\right] = \sigma^{2(ht)}_{u(3)} + \left(n - \bar{n}^{(3)}_0\right)\Big/\left(n - \bar{n}^{(2)}_0\right)\sigma^{2(ht)}_{\eta(3)}$. It follows that when the HT 3-level model (4) holds,

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

under the assumption of $E_3\left[\hat{\xi}^{(3)}_{ijk}\right] \approx E_3\left[\hat{\xi}^{(2)}_{ijk}\right] \approx \xi^{(3)}_{ijk}$. That is, 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. Two area-specific adjustments or robustifications of $\hat{\mathrm{V}}_{(2)}\left(\bar{Y}_i\right)$ that lead to an unbiased or approximately unbiased estimator of $\mathrm{Var}_{(3)}\left(\bar{Y}_i\right)$ are

$$\hat{\mathrm{V}}^M_{(2)}\left(\bar{Y}_i\right) = \left\{\left(1\big/\bar{m}^{(2)}_{Ui}\right)\hat{s}^2_{h(3)} + \hat{s}^2_{u(3)}\right\}\bar{m}^{(2)}_{Ui} + \hat{s}^2_{e(2)}\bar{m}^{(3)}_{Ui}$$

and

$$\hat{\mathrm{V}}^M_{(2)}\left(\bar{Y}_i\right) = \left\{\left(1\big/\bar{m}^{(2)}_{Ui}\right)\hat{s}^{2(ht)}_{h(3)} + \hat{s}^{2(ht)}_{u(3)}\right\}\bar{m}^{(2)}_{Ui} + \hat{s}^{2(ht)}_{e(2)}\bar{m}^{(3)}_{Ui},$$

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

Under the MELL approach, the main task is to adjust the cluster-level variance component in order to capture the potential area-level variability. As in Das and Chambers (2017), the adjustment factors for the cluster-variance component can then be defined as follows:

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

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

and

$$k_{3(p)} = \hat{\sigma}^{-2(ht)}_{u(2)}\left[\hat{\sigma}^{2(ht)}_{\eta(3)} D^{-1}_{(p)}\sum_{i=1}^{D_{(p)}}\left(1\big/\bar{m}^{(2)}_{Ui}\right) + \hat{\sigma}^{2(ht)}_{u(3)}\right];\;\; p = 1,\ldots,P$$

where $\bar{m}^{(2)}_{si} = m_i^2\Big/\sum_{j=1}^{C_i} m^2_{ij}$, D_s is the number of sampled areas, m_i is the total number of sampled individuals in the i^{th} sampled area and m_{ij} is the total number of sampled individuals in the j^{th} sampled cluster of the i^{th} sampled area. Note that when no household size information is available we set household sizes to one so that the adjustment factors then depend on the number of HHs. Since the adjustment factors are based on the area-specific sample and population sizes, there are no basic difference between the adjusted factors under homoscedasticity and heteroskedasticity except for the estimated variance components. The MELL 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 $\tilde{u}_{ij}$ need to be rescaled so that the ratio of bootstrap variations $\text{var}\left(\breve{u}_{ij}\right)/\text{var}\left(\tilde{u}_{ij}\right)$ approximates the ratio of corresponding estimated variance components $\hat{\sigma}_u^{2(M)}/\hat{\sigma}_{u(2)}^2$, where the scaled residuals are $\breve{u}_{ij(2)}=\tilde{u}_{ij(2)}\hat{\sigma}_u^{(M)}\delta$ and $\tilde{u}_{ij(2)}=\hat{u}_{ij(2)}\hat{\sigma}_{u(2)}\delta$ with $\delta=1/\sqrt{\left(C_s-1\right)\sum_{ij\in s}\tilde{u}_{ij(2)}^2}$. The procedure is similar for all variations of these proposed modifications. For the stratification-based MELL, the generation of cluster-specific residuals or resampling of sample cluster-specific residuals is carried out using stratum-specific adjusted cluster-variance components, say $\hat{\sigma}_{u.h}^{(M)}=k_3^{(h)}\hat{\sigma}_{u(2)}^2$, $h=1,..,H$. These bootstrap procedures are explained in more detail in the next section.

3. Implementation of ELL Method and Its Alternatives

3.1. First stage regression modelling

The first step when applying the ELL method is to fit a regression model utilizing the survey dataset. It is strongly recommended that this model incorporates a large number of explanatory variables at different levels of the data hierarchy in order to capture the potential between-cluster and between-area variabilities. However, overfitting this regression model can then become a problem. Furthermore, the inclusion of more individual level explanatory variables and contextual variables does not guarantee a significant reduction of the between-area variability. In order to avoid this overfitting problem, the logarithm of HH per capita monthly consumption expenditure is first regressed on the same 30 explanatory variables that were used in BBS-2004 study. For the first two datasets (Set-1 and Set-2), these explanatory variables are next used to fit 2-level and 3-level models. The estimated variance-components from these fits are then used to calculate the generalized least squares (GLS) estimates of the regresion model parameters and their associated estimated variance-covariance matrix. Marginal and conditional R-squared values are calculated following Nakagawa and Schielzeth (2013) in order to compare the two multilevel models. Estimated values of parameters and level-specific random errors are stored for each model for use in the bootstrap procedure. Note that these models are fitted assuming both HM and HT variances for the HH-specific random errors.

3.2. Heteroskedasticity Modelling

To examine the heteroskedasticity of the HH-level random errors, the squared least squares (LS) residuals obtained from both Set-1 and Set-2 are plotted against the corresponding predicted values (Figure 1). The plots suggest negligible monotone heteroskedasticity due to less information at the extreme tails. The “alpha” model is fitted with the explanatory variables that were used in the BBS-2004 study to avoid any extensive exploration of potential explanators of heteroskedasticity. Figure 1 shows that the HT error variances estimated by the ELL parametric approach under the 2-level model ($\hat{\sigma}_{\varepsilon(2),ijk}^{2.ELL}$) fluctuate more than those under the 3-level model ($\hat{\sigma}_{\varepsilon(3),ijk}^{2.ELL}$), which may be due to 2-level model ignoring the area variability in the survey dataset. The HT error variances are slightly higher and with less variation when estimated under a 3-level model compared with a 2-level model. The estimated area-level and cluster-level HM variance components under the assumption of HT HH-level errors are set out in Table A1 in the Annex. The variance components estimated by MOM under the HT approaches are close to those estimated under the assumption of HM HH-level errors. This suggests that level-one heteroskedasticity has a negligible effect on estimation of higher level variance components.

3.3. Bootstrapping

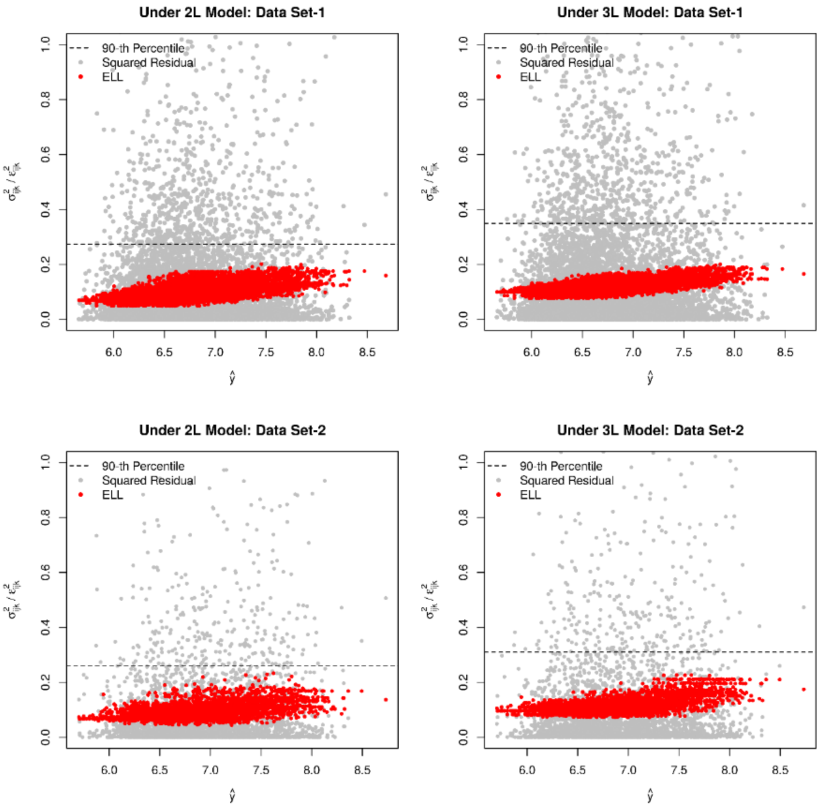
Both the PB and NPB bootstrap procedures have been used to calculate the FGT poverty estimates and their MSEs. Under the HT model, the NPB procedure is a semi-parametric bootstrap (SPB) since the HH-level error variances are generated on the basis of a parametric specification for the alpha parameters. In both types of bootstrap procedure, the regression parameters are generated from a multivariate $N\left(\boldsymbol{\beta}^{gls}_{(.)}, \hat{v}\left(\boldsymbol{\beta}^{gls}_{(.)}\right)\right)$ distribution. For the PB procedure, level specific random errors are generated from a normal distribution with zero mean and the corresponding estimated variances as variance. In the NPB and SPB procedures, the LS raw residuals are used to generate level-specific random errors via 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 need to be re-scaled. We therefore use scaled residuals in all bootstrapping under both the 2-level and 3-level models.

Under HH-level heteroskedasticity there is no recommended way to re-scale the HH-level residuals when they are used for bootstrapping. In this paper we therefore scale the HH-specific HT residuals by the HH-level variance component, estimated by $\hat{\sigma}_e^2 - \hat{\sigma}_{u(2)}^{2(ht)}$ under a 2-level model, where $\hat{\sigma}_e^2$ is the MSE of the initial single-level linear model fitted by the LS method. The mean of the estimated HH-level residual variances (say, $n^{-1}\sum \hat{\sigma}^{2.ELL}_{\varepsilon(2),ijk}$) or the HM variance component $\hat{\sigma}^2_{\varepsilon(2)}$ can also be used for scaling. SRSWR resampling of the level-specific residuals can be implemented either unconditionally or conditionally when implementing the ELL (2002) method. Under unconditional sampling, the level-specific errors are assigned to census units from the full set of sample residuals. Under conditional sampling, the level-specific residuals are drawn following a nested approach. In this case, suppose a cluster-specific sample residual is randomly assigned to a census cluster. Then census HHs nested within that census cluster are assigned HH-level random errors from the subset 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 our study, we have followed a conditional approach under a 2-level model and an unconditional approach under a 3-level model. We note that both the unconditional and conditional approaches behave similarly under a 2-level model, but the conditional approach produces unstable results under a 3-level model. Since ELL-based methods can be implemented in different ways based on (i) the adopted bootstrap procedure (PB/NPB/SPB), (ii) heteroskedasticity vs, homoskedasticity of HH-level random errors (HM/HT), and (ii) the assumed working model (2L/3L); the resulting estimators are denoted differently in Table A2 in the Annex. For example “PELL.HM.2L” stands for PB based ELL estimator under a 2-level HM working model.

3.4. Application of modified ELL methods

Das and Chambers (2017) found that the stratification-based adjustment (Adjustment 3) works well, and so this approach has been used when implementing the MELL methods under both HM and HT level-one errors. Sub-districts are grouped into six strata according to 20^{th}, 35^{th}, 50^{th}, 65^{th}, and 80^{th} percentile of the distribution of their population size. The cluster-level variance component $\hat{s}^{2}_{u(2)}$ is adjusted within each stratum according to $\hat{s}^{2(h)}_{u(2)} = k_3^{(h)} \hat{s}^{2}_{u(2)}$, $h = 1,\ldots,6$. Note that the same stratification has been used under both the HM and HT working models. As a consequence, stratum specific bootstrap procedures based on the stratum-specific adjusted cluster-level variance component are implemented when applying the ELL method.

Figure 1: ELL-based estimates of unit-level heteroskedastic error variances under 2-level and 3-level models



4. Results and Discussion

In the HIES 2000, lower (LPL) and upper (UPL) poverty lines were developed in order to calculate FGT poverty indices. These poverty lines were defined on the basis of the 16 strata that were created for the survey sampling design. Since the HIES 2000 was based on the Bangladesh Population and Housing Census 1991, these poverty lines were therefore mapped to the 2001 Census in the poverty mapping study based on BBS-2004. However, it should also be noted that the poverty lines used in the BBS study are not exactly maintained in the study reported here because some sub-districts lacked information. Sub-district level poverty incidences in BBS-2004 were calculated using the ELL method with SPB via a conditional approach under a 2-level working model. In this study the similar approach is followed using scaled residuals to estimate poverty indicators under a 2-level model (SPELL.HT.2L). The consistency of the resulting estimated FGT poverty indices was checked by comparing summary statistics for the estimated poverty measures and their MSEs with those obtained via BBS and UNWFP (2004, page 34). Table 3 shows that means of the estimated FGT measures are slightly lower with slightly higher standard deviation (SD) compared to those obtained in the BBS study, while means and SDs of the estimated MSEs are marginally smaller. The reasons of these differences may be (i) a different procedure was used to fit the regression model, (ii) a different approach to heteroskedasticity modelling, (iii) a different bootstrap procedure, (iv) slightly different poverty lines for some sub-districts, and (v) the number of resamples in the bootstrap procedure (500 instead of 100 bootstrap). We also noticed that two of the model covariates that were created for our study behaved differently in the fitted regression model. Summary values of the estimated MSEs generated by the 3-level model-based ELL (SPELL.HT.3L), and the modified ELL (MSPELL.HT) estimators are also shown in Table 3. It can be seen that the SPELL.HT.2L estimator produces underestimated MSEs in comparison to the SPELL.HT.3L estimator, while the MSPELL.HT estimator corrects this downward bias under an assumption of between area variability. We therefore conclude that the SPELL.HT.2L estimator might provide better accuracy if real between area variability is absent.

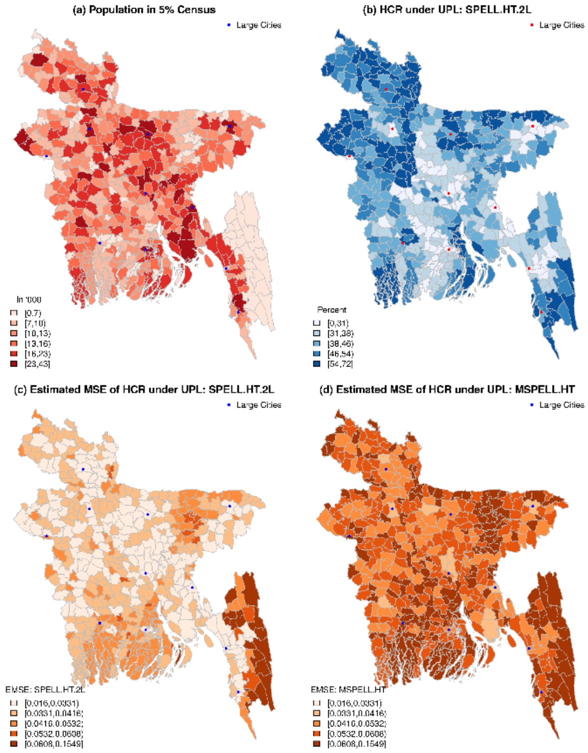
In the BBS-2004 poverty mapping study, MSEs were calculated for the FGT estimates as well as the number of households. For examining the relationship between these quantities, sub-district specific population sizes, estimated HCR at UPL, and estimated MSE (SPELL.HT.2L and MSPELL.HT) are plotted in Figure 2. The first three maps of Figure 2 suggest that sub-districts with large populations and closer to the capital (Dhaka) or port cities (Chittagong) have lower HCRs with lower MSEs. Conversely, sub-districts with smaller populations in coastal and hilly regions have higher HCRs with higher MSEs. Some sub-districts in the north-western part of Bangladesh with large populations (Rangpur and Rajshahi divisions) have relatively higher HCRs with lower MSEs. The maps (c) and (d) in Figure 2 show that the values of the SPELL.HT.2L MSE estimator are generally lower than comparable MSE values generated by the MSPELL.HT estimator. In general, this good performance of SPELL.HT.2L is 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. Ignoring between area variability is the main reason why the SPELL.HT.2L estimator has low MSEs. If real between area variability exists, this false accuracy will mislead policy makers.

Under the assumption of between area variability, it is clear that the SPELL.HT.2L estimator leads to underestimated MSEs particularly for large cities (e.g., blue points) because it ignores between area variability while the MSPELL.HT estimator corrects for this underestimation by allowing for potential between area variability. The SPELL.HT.2L and MSPELL.HT estimators lead to similar MSEs only for the significantly smaller sub-districts, particularly those in the south-eastern regions. The variation in the estimated MSEs is more explicit when Map (c) is compared to Map (d). The reason for this may be the tendency of lower MSEs for sub-districts with a modest population sizes (more than 13,000) and highest MSEs for the smaller sub-districts when using the SPELL.HT.2L estimator, compared with the MSPELL.HT estimator, which provides comparatively higher weight to the smaller sub-districts and comparatively lower weight to the larger sub-districts when accounting for between area variability. We conclude that when there is negligible between area variability, SPELL.HT.2L is highly optimistic while MSPELL.HT is reasonably conservative. More detailed comparisons are given in the following paragraphs.

Table 3: Summary statistics of the estimated HCR, PG, and PS at lower (LPL) and upper (UPL) 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.	Max.	Mean	SD
HCR	LPL	BBS-2004	0.0014	0.5531	0.2930	0.1063
		SPELL.HT.2L	0.0089	0.5898	0.2810	0.1174
	UPL	BBS-2004	0.0049	0.7081	0.4212	0.1238
		SPELL.HT.2L	0.0453	0.7239	0.4164	0.1280
PG	LPL	BBS-2004	0.0001	0.1638	0.0679	0.0297
		SPELL.HT.2L	0.0012	0.1732	0.0622	0.0325
	UPL	BBS-2004	0.0006	0.2441	0.1142	0.0421
		SPELL.HT.2L	0.0073	0.2462	0.1082	0.0443
PS	LPL	BBS-2004	0.0000	0.0650	0.0228	0.0112
		SPELL.HT.2L	0.0003	0.0676	0.0201	0.0121
	UPL	BBS-2004	0.0001	0.1091	0.0429	0.0182
		SPELL.HT.2L	0.0019	0.1079	0.0392	0.0190
Estimated MSE of HCR	LPL	BBS-2004	0.0025	0.1145	0.0388	0.0146
		SPELL.HT.2L	0.0047	0.1047	0.0340	0.0122
		SPELL.HT.3L	0.0070	0.1152	0.0617	0.0159
		MSPELL.HT	0.0146	0.1533	0.0553	0.0146
	UPL	BBS-2004	0.0062	0.1081	0.0416	0.0143
		SPELL.HT.2L	0.0127	0.1027	0.0371	0.0115
		SPELL.HT.3L	0.0226	0.1147	0.0680	0.0118
		MSPELL.HT	0.0304	0.1549	0.0596	0.0135
Estimated MSE of PG	LPL	BBS-2004	0.0003	0.0436	0.0127	0.0054
		SPELL.HT.2L	0.0008	0.0313	0.0106	0.0045
		SPELL.HT.3L	0.0012	0.0401	0.0196	0.0071
		MSPELL.HT	0.0037	0.0549	0.0216	0.0061
	UPL	BBS-2004	0.0009	0.0538	0.0169	0.0066
		SPELL.HT.2L	0.0027	0.0434	0.0144	0.0051
		SPELL.HT.3L	0.0045	0.0538	0.0269	0.0073
		MSPELL.HT	0.0085	0.0730	0.0270	0.0066
Estimated MSE of PS	LPL	BBS-2004	0.0001	0.0203	0.0055	0.0026
		SPELL.HT.2L	0.0003	0.0125	0.0044	0.0021
		SPELL.HT.3L	0.0003	0.0210	0.0081	0.0035
		MSPELL.HT	0.0013	0.0250	0.0110	0.0035
	UPL	BBS-2004	0.0002	0.0300	0.0082	0.0036
		SPELL.HT.2L	0.0009	0.0209	0.0068	0.0028
		SPELL.HT.3L	0.0014	0.0268	0.0126	0.0044
		MSPELL.HT	0.0033	0.0386	0.0149	0.0041

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



In what follows FGT poverty indicators at UPL are estimated assuming both HM and HT HH-level errors under 2-level and 3-level working models for data Set-1. The HCRs estimated by different ELL estimators are plotted against the HCRs estimated by the standard 2-level model-based ELL estimator with PB procedure (PELL.HM.2L). Figure 3 shows that the NPB-based ELL estimators of HCR under a 2-level working model (NPELL.HM.2L) lead to almost same results as PELL.HM.2L estimator under homoskedasticity. The HT estimator SPELL.HT.2L leads to slightly larger HCRs compared to the PELL.HM.2L estimator particularly for the areas with higher HCRs. These values are most likely overestimates when 3-level model is taken to be the working model. However, under homoscedasticity, the 3-level model-based estimator performs similarly to the 2-level model-based estimator with some fluctuations. This suggests that the 3-level model-based estimators provide comparable unbiased poverty estimates to the 2-level model-based estimators. The observed differences in the performances of the estimators under homoscedasticity and heteroskedasticity indicate that the observed heteroskedasticity might have influence on the estimated poverty measures. This influence of level-one heteroskedasticity has already been noted in the MSE estimates displayed in Figure 4. Under both the 2-level and the 3-level model based approaches, estimators with heteroskedastic level-1 errors have slightly higher MSEs than their homoscedastic counterparts.

Figure 3: NPELL and SPELL estimates of HCR at UPL under 2-level and 3-level working model with homoskedastic (HM) and heteroskedastic (HT) level-one errors

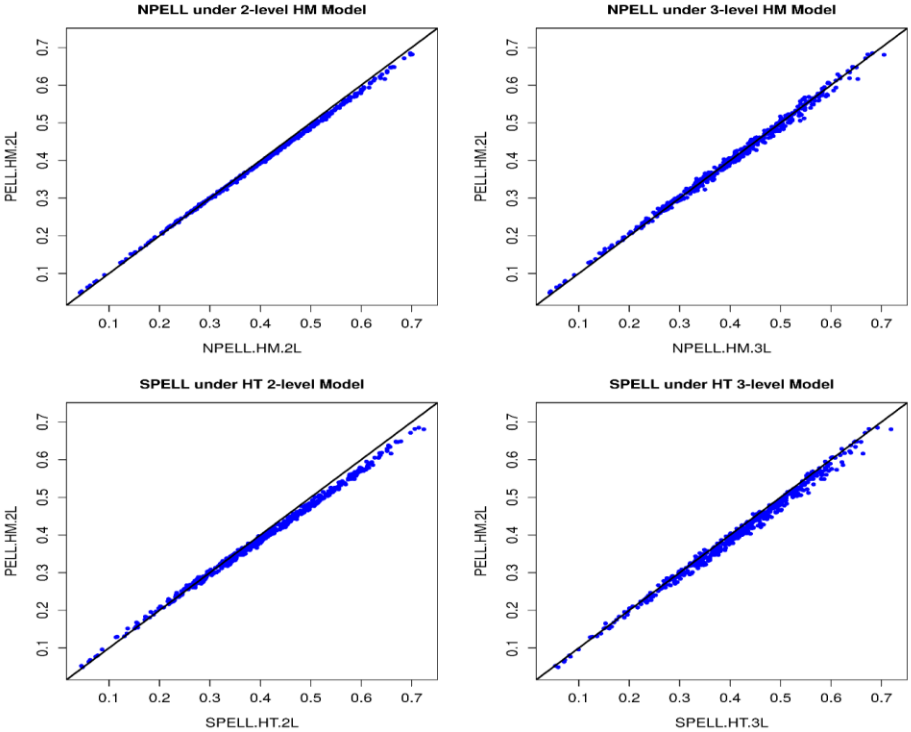
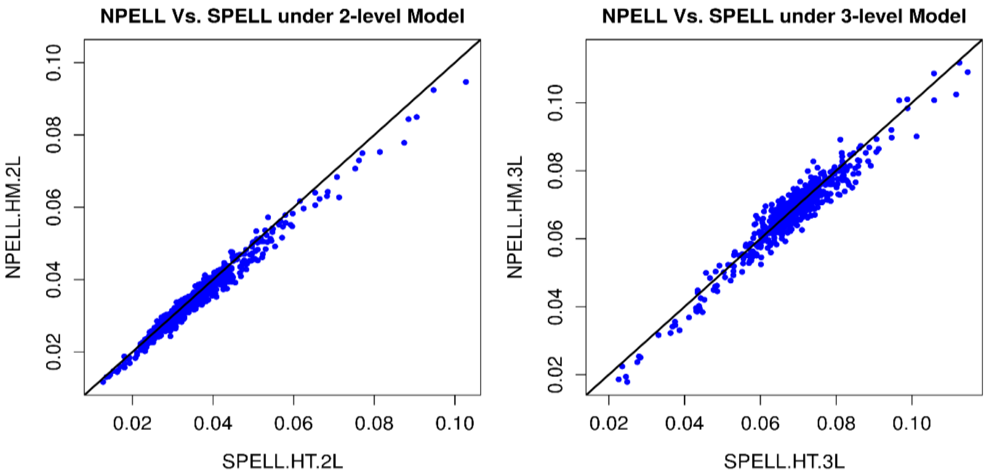
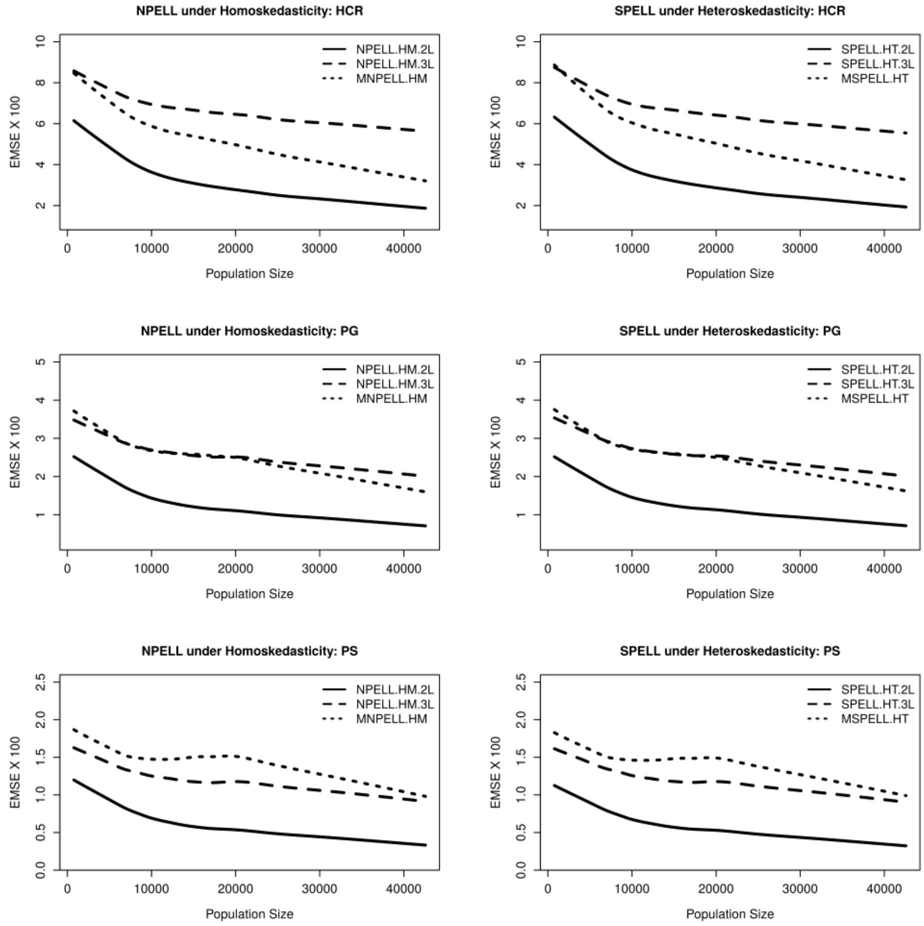


Figure 4: Comparison of estimated MSEs (EMSE) of the estimated HCRs at UPL under 2-level and 3-level working model with homoskedastic (HM) and heteroskedastic (HT) level-one errors for the NPELL and SPELL estimates



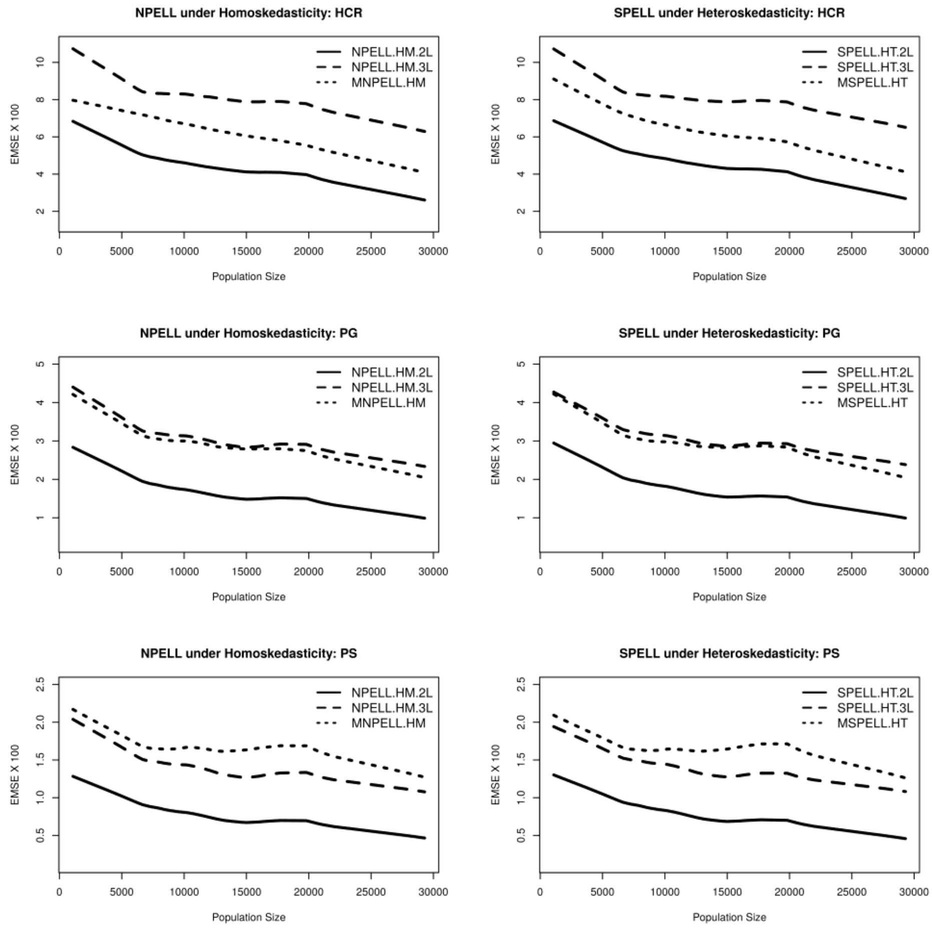
The estimated MSEs of the estimated FGT indicators (HCR, PG, PS) obtained via the ELL estimators and their modified versions under the assumption of both HM and HT level-one errors are plotted against area-specific population sizes in Figure 5 for the data Set-1. This shows a declining trend of estimated MSE as population size increases. As expected the 2-level model-based ELL MSE estimators are lower than the 3-level model-based MSE estimators, with the modified 2-level MSE estimators fixing this underestimation problem for all three indicators. For HCR, the modified estimators have estimated MSEs close to the naïve 3-level estimator for areas with smaller population, but have lower estimated MSEs for areas with larger population. The difference between the estimated MSEs calculated by the 2-level and 3-level model-based estimators increases with population size. For PG and PS, the performance of the modified estimators does not vary as much. Figure 5 shows that the modified estimators lead to estimated MSEs very close to those generated by the 3-level model-based estimators for PG and estimated MSEs that are slightly higher for PS.

Figure 5: Estimated MSE (EMSE) of HCR at UPL under 2-level and 3-level models with homoskedastic (HM) and heteroskedastic (HT) level-one errors by ELL estimators with their modified versions for full data (set-1)



When between area variability is clear in the dataset Set-2, the modified estimators based on the PB procedure behave in the same way as in the dataset Set-1. See Figure 6. In particular, the modified ELL estimators of HCR exhibit some downward bias in estimated MSE when compared to the 3-level ELL estimator but still perform much better than the naïve 2-level ELL estimators. This downward bias disappears when the modified procedures are used to estimate PG and PS.

Figure 6: Estimated MSE (EMSE) of HCR at UPL 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



5. Concluding Remarks

The empirical evidence presented in this paper shows that the ELL estimators based on a 2-level working model fail to capture the true MSE if the working regression model violates the area homogeneity assumption. In the Bangladesh dataset that is our focus, the area homogeneity assumption is clearly violated in urban small areas (sub-districts). Since use of ELL estimators based on a 3-level model is not realistic for this dataset (most sampled sub-districts have single sampled clusters), use of 2-level model-based ELL estimator leads to stable FGT estimates but underestimated MSEs. The proposed modified version of the 2-level ELL estimator overcomes this underestimation problem, and seems to work well under both HM and HT level-one errors. We explore the empirical behavior of this modified 2-level approach where between area variability is obvious (set-2) and when it is negligible (set-1). In both situations, the modified estimators performed better than the naive 2-level ELL estimators. When used to estimate the HCR, the modified estimators appear to be slightly downward biased in terms of estimated MSE when compared to the 3-level estimators but much less biased than the naïve 2-level estimators. In particular, we noted that when there is significant between area variability, as occurs in the urban areas of Bangladesh, the 2-level model-based naïve MSE estimators provide an inaccurate impression of accuracy as far as estimation of FGT measures is concerned. In contrast, although the modified MSE estimators seem conservative because they allow for between area variability, they also considerably reduce this problem of MSE underestimation. This is in accord with simulation results reported in Das and Chambers (2017). Since the cost of underestimating the MSE (i.e. incorrectly claiming higher precision) may be much higher than the premium for obtaining conservative precision (i.e. a slightly overestimated MSE), proper care and appropriate investigation should be undertaken before using a naïve 2-level model-based MSE estimator to obtain FGT estimates.

References

BBS (2001). 2001 Popultion Census: Priliminary Report. Ministry Planning, Government of Bangladesh, Statistics Division. 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.

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

Das, S. (2016). Robust inference in poverty mapping. *Doctor of Philosophy thesis*. University of Wollongong, Australia.

Das, S. and Chambers, R. (2017). Robust mean-squared error estimation for poverty estimates based on the method of Elbers, Lanjouw and Lanjouw. *Journal of Royal Statistical Society, Series A*, **180**, 1137–1161.

Elbers, C., Lanjouw, J.O., and 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. and Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica,* **71(1)**, 355–364.

Elbers, C., Lanjouw, P. and 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.

Nakagawa, S. and 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.

Tarozzi, A. and 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.

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

World Bank (2013). *Nepal Small Area Estimation of Poverty 2011. Volume 1.* Washington DC: The World Bank.

World Bank (2015). Poverty and Equality Databank. http://povertydata.worldbank.org.

Annexure 1

Table A1: 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 A2: ELL 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
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

Annexure 2

A.1 Variance Component Estimation under 2-level Model with HT Level-1 Errors

Under the population model (2), the 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, it can be shown that

$$\mathrm{Var}_2\left(\hat{\bar{e}}_{ij.}\right) = \sigma_u^{2(ht)} + n_{ij}^{-2}\sum_{k\in s}\sigma^2_{\varepsilon(2),ijk}\ , \mathrm{Cov}_2\left(\hat{u}_{ij},\hat{u}_{i'j'}\right) = \left(\sigma_u^{2(ht)} + n_{ij}^{-2}\sum_{k\in s}\sigma^2_{\varepsilon(2),ijk}\right) I\left\{ij = i'j'\right\},$$

$$\mathrm{Var}_2\left(\hat{\bar{e}}_{...}\right) = n^{-2}\sum_{ij} n_{ij}^2\sigma_u^{2(ht)} + n^{-2}\sum_{ijk\in s}\sigma^2_{\varepsilon(2),ijk}\ , E_2\left(\sum_{ijk\in s}\hat{e}_{ijk}^2\right) = n\sigma_u^{2(ht)} + \sum_{ijk\in s}\sigma^2_{\varepsilon(2),ijk}\ ,$$

$$E_2\left(\sum_{ijk\in s}\hat{e}_{ijk}\right)^2 = \sigma_u^{2(ht)}\sum_{ij\in s} n_{ij}^2 + \sum_{ijk\in s}\sigma^2_{\varepsilon(2),ijk} \text{ and } E_2\left(\sum_{ij\in s} n_{ij}\hat{\bar{e}}_{ij.}^2\right) = n\sigma_u^{2(ht)} + \sum_{ij\in s} n_{ij}^{-1}\sum_k \sigma^2_{\varepsilon(2),ijk}\ .$$

Then the expectation of HH and cluster level sample residual variances $s^{(1)}$ and $s^{(2)}$ can be written as

$$E_2\left(s^{(1)}\right) = \frac{1}{n-1}\sum_{ijk\in s} E_2\left(\hat{e}_{ijk} - \hat{\bar{e}}_{...}\right)^2 = \frac{n-\bar{n}_0^{(2)}}{n-1}\sigma_u^{2(ht)} + \frac{1}{n}\sum_{ijk\in s}\sigma^2_{\varepsilon(2),ijk} \text{ where } \bar{n}_0^{(2)} = n^{-1}\sum_{ij\in s} n_{ij}^2$$

$$\text{and } E_2\left(s^{(2)}\right) = \frac{1}{C_s - 1}\sum_{ij\in s} n_{ij} E_2\left(\hat{\bar{e}}_{ij.} - \hat{\bar{e}}_{...}\right)^2 = \frac{n-\bar{n}_0^{(2)}}{C_s - 1}\sigma_u^{2(ht)} + \frac{1}{C_s - 1}\sum_{ij\in s}\left(\frac{1}{n_{ij}} - \frac{1}{n}\right)\sum_k \sigma^2_{\varepsilon(2),ijk}\ .$$

Under the assumption of known $\sigma^2_{\varepsilon(2),ijk}$, the estimator $\hat{\sigma}_{u(2)}^{2(ht)}$ can be obtained from $E_2\left(s^{(2)}\right)$ as

$$\hat{\sigma}_{u(2)}^{2(ht)} = \frac{1}{n - \bar{n}_0^{(2)}}\left\{\left(C_s - 1\right)s^{(2)} - \sum_{ij\in s}\left(\frac{1}{n_{ij}} - \frac{1}{n}\right)\sum_k \sigma^2_{\varepsilon(2),ijk}\right\}.$$

Putting $w_{ij} = n_{ij}n^{-1}$, the terms involved in $\hat{\sigma}_{u(2)}^{2(ht)}$ can be expressed as $n - \bar{n}_0^{(2)} = n\sum_{ij} n_{ij}n^{-1}\left(1 - n_{ij}n^{-1}\right) = n\sum_{ij} w_{ij}\left(1 - w_{ij}\right)$, $\sum_{ij\in s}\left(n_{ij}^{-1} - n^{-1}\right) = n\sum_{ij\in s} w_{ij}\left(1 - w_{ij}\right)n_{ij}^{-2}$, 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

$$\hat{\sigma}_{u(2)}^{2(ht)} = \left\{1\Big/\sum_{ij} w_{ij}\left(1 - w_{ij}\right)\right\}\left\{\sum_{ij\in s} w_{ij}\left(\hat{\bar{e}}_{ij.} - \hat{\bar{e}}_{...}\right)^2 - \sum_{ij\in s} w_{ij}\left(1 - w_{ij}\right)n_{ij}^{-2}\tau_{ij}^2\right\}$$

where $\tau_{ij}^2 = n_{ij}^{-2}\sum_{k\in s}\sigma^2_{\varepsilon(2),ijk}$. 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 $\sigma^2_{\varepsilon(2),ijk}$. An unbiased estimator of τ_{ij}^2 is $\hat{\tau}_{ij}^2 = n_{ij}^{-1}\left(n_{ij} - 1\right)^{-1}\sum_{k\in s}\left(\hat{\varepsilon}_{ijk} - \hat{\bar{\varepsilon}}_{ij.}\right)^2 = n_{ij}^{-1}\left(n_{ij} - 1\right)^{-1}\sum_{k\in s}\left(\hat{e}_{ijk} - \hat{\bar{e}}_{ij.}\right)^2$ where

$\hat{\varepsilon}_{ij.} = n_{ij}^{-1}\sum_{k\in s}\hat{\varepsilon}_{ijk}$. It is easy to show that $E_2\left(\hat{\tau}_{ij}^2\right) = n_{ij}^{-2}\sum_{k\in s}\sigma^2_{\varepsilon(2),ijk} = \tau_{ij}^2$ since $E_2\left[\sum_{k\in s}\left(\hat{e}_{ijk} - \hat{\bar{e}}_{ij.}\right)^2\right] = \left(n_{ij}-1\right)n_{ij}^{-1}\sum_{k\in s}\sigma^2_{\varepsilon(2),ijk}$.

Thus the ultimate plug-in estimator of $\sigma^2_{u(2)}$ considering heteroskedasticity at the level-one is

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

A.2 Variance Component Estimation: 3-level Model with HT Level-1 Errors

Under the population model (4), the moment-based estimates of area, cluster and HH level random effects are calculated 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 $\hat{\varepsilon}_{ijk} = \hat{e}_{ijk} - \hat{\eta}_i - \hat{u}_{ij}$. Under (4), it can be shown that

$$\mathrm{Var}_3\left(\hat{\bar{e}}_{ij.}\right) = \sigma^{2(ht)}_{\eta(3)} + \sigma^{2(ht)}_{u(3)} + n_{ij}^{-2}\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk},\ \mathrm{Var}_3\left(\hat{\bar{e}}_{i..}\right) = \sigma^{2(ht)}_{\eta(3)} + n_i^{-2}\sum_{j\in s} n_{ij}^2\sigma^{2(ht)}_{u(3)} + n_i^{-2}\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}$$

$$\mathrm{Var}_3\left(\hat{\bar{e}}_{...}\right) = n^{-2}\left(\sum_{i\in s} n_i^2\sigma^{2(ht)}_{\eta(3)} + \sum_{ij\in s} n_{ij}^2\sigma^{2(ht)}_{u(3)} + \sum_{ijk\in s}\sigma^2_{\varepsilon(3),ijk}\right), E_3\left(\sum_{ijk\in s}\hat{e}^2_{ijk}\right) = n\left(\sigma^{2(ht)}_{\eta(3)} + \sigma^{2(ht)}_{u(3)}\right) + \sum_{ijk\in s}\sigma^2_{\varepsilon(3),ijk},$$

$$E_3\left(\sum_{ij\in s} n_{ij}\hat{\bar{e}}^2_{ij.}\right) = n\left(\sigma^{2(ht)}_{\eta(3)} + \sigma^{2(ht)}_{u(3)}\right) + \sum_{ij\in s}\frac{1}{n_{ij}}\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk},$$

$$E_3\left(\sum_{i\in s} n_i\hat{\bar{e}}^2_{i..}\right) = n\sigma^{2(ht)}_{\eta(3)} + \sum_{i\in s}\bar{n}^{(2)}_{0i}\sigma^{2(ht)}_{u(3)} + \sum_{i\in s} n_i^{-1}\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}, \text{ and}$$

$$E_3\left(\sum_{ijk\in s}\hat{e}_{ijk}\right)^2 = \sum_{ijk\in s}\sigma^2_{\varepsilon(3),ijk} + \sigma^{2(ht)}_{u(3)}\sum_{ij\in s} n_{ij}^2 + \sigma^{2(ht)}_{\eta(3)}\sum_{i\in s} n_i^2.$$

Now the expectation of the sample residual variances become

$$E_3\left[s^{(1)}\right] = n^{-1}\sum_{ijk\in s}\sigma^2_{\varepsilon(3),ijk} + \frac{n-\bar{n}_0^{(2)}}{n-1}\sigma^{2(ht)}_{u(3)} + \frac{n-\bar{n}_0^{(3)}}{n-1}\sigma^{2(ht)}_{\eta(3)} = a_{11} + a_{12}\sigma^{2(ht)}_{u(3)} + a_{13}\sigma^{2(ht)}_{\eta(3)},$$

$$E_3\left[s^{(2)}\right] = \left(C_s-1\right)^{-1}\left[\sum_{ij\in s}\left(n_{ij}^{-1} - n^{-1}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}\right] + \left(C_s-1\right)^{-1}\left(n-\bar{n}_0^{(2)}\right)\sigma^{2(ht)}_{u(3)} + \left(C_s-1\right)^{-1}\left(n-\bar{n}_0^{(3)}\right)\sigma^{2(ht)}_{\eta(3)}$$
$$= a_{21} + a_{22}\sigma^{2(ht)}_{u(3)} + a_{23}\sigma^{2(ht)}_{\eta(3)}, \text{ and}$$

$$E_3\left[s^{(3)}\right] = \left(D-1\right)^{-1}\sum_{i\in s}\left(n_i^{-1} - n^{-1}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk} + \frac{\sum_{i\in s}\bar{n}^{(2)}_{0i} - \bar{n}^{(2)}_0}{D-1}\sigma^{2(ht)}_{u(3)} + \frac{n-\bar{n}^{(3)}_0}{D-1}\sigma^{2(ht)}_{\eta(3)}$$
$$= a_{31} + a_{32}\sigma^{2(ht)}_{u(3)} + a_{33}\sigma^{2(ht)}_{\eta(3)}$$

where $\bar{n}^{(3)}_0 = n^{-1}\sum_{i\in s} n_i^2$.

The estimators $\hat{\sigma}^{2(ht)}_{u(3)}$ and $\hat{\sigma}^{2(ht)}_{\eta(3)}$ can be derived directly from the second and third simultaneous equations, as $\hat{\sigma}^{2(ht)}_{u(3)} = \left[s^{(2)} - a_{21} - a_{23}\hat{\sigma}^{2(ht)}_{\eta(3)}\right]\Big/a_{22}$ and $\hat{\sigma}^{2(ht)}_{\eta(3)} = \left[s^{(3)} - a_{31} - a_{32}\hat{\sigma}^{2(ht)}_{u(3)}\right]\Big/a_{33}$ respectively. Putting $\hat{\sigma}^{2(ht)}_{u(3)}$ into $\hat{\sigma}^{2(ht)}_{\eta(3)}$, we can obtain

$$\hat{\sigma}^{2(ht)}_{\eta(3)} = \left[\left(a_{22}a_{33} - a_{32}a_{23}\right)\right]^{-1}\left[a_{22}s^{(3)} - a_{32}s^{(2)} - \left(a_{31}a_{22} - a_{21}a_{32}\right)\right].$$

In similar way, putting new $\hat{\sigma}^{2(ht)}_{\eta(3)}$ in $\hat{\sigma}^{2(ht)}_{u(3)}$ we can obtain

$$\hat{\sigma}^{2(ht)}_{u(3)} = \left(a_{22}a_{33} - a_{32}a_{23}\right)^{-1} \times \left[a_{33}s^{(2)} - a_{23}s^{(3)} - \left(a_{21}a_{33} - a_{31}a_{23}\right)\right].$$

The terms involved in $\hat{\sigma}^{2(ht)}_{\eta(3)}$ and $\hat{\sigma}^{2(ht)}_{u(3)}$ can be simplified as

$$a_{22}a_{33} - a_{32}a_{23} = \frac{\left(n - \overline{n}_0^{(2)}\right)\left(n - \overline{n}_0^{(3)}\right)}{\delta_n} - \frac{\left(n - \overline{n}_0^{(3)}\right)\left(\sum_{i\in s}\overline{n}_{0i}^{(2)} - \overline{n}_0^{(2)}\right)}{\delta_n} = \frac{\left(n - \overline{n}_0^{(3)}\right)\left(n - \sum_{i\in s}\overline{n}_{0i}^{(2)}\right)}{\delta_n},$$

$$a_{31}a_{22} - a_{21}a_{32} = \frac{1}{\delta_n}\left\{\left(n - \overline{n}_0^{(2)}\right)\sum_{i\in s}\left(\frac{1}{n_i} - \frac{1}{n}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk} - \left(\sum_{i\in s}\overline{n}_{0i}^{(2)} - \overline{n}_0^{(2)}\right)\sum_{ij\in s}\left(\frac{1}{n_{ij}} - \frac{1}{n}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk}\right\},$$

$$a_{21}a_{33} - a_{31}a_{23} = \frac{n - \overline{n}_0^{(3)}}{\delta_n}\left\{\sum_{ij\in s}\left(\frac{1}{n_{ij}} - \frac{1}{n}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk} - \sum_{i\in s}\left(\frac{1}{n_i} - \frac{1}{n}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}\right\}.$$

where $\delta_n = \left(C_s - 1\right)\left(D - 1\right)$. Then the ultimate estimators of cluster and area level variance components become

$$\hat{\sigma}^{2(ht)}_{u(3)} = \frac{C_s - 1}{n - \overline{n}_0^{(3)}}\frac{D-1}{n - \sum_{i\in s}\overline{n}_{0i}^{(2)}}\left[\begin{array}{l}(D-1)^{-1}\left(n - \overline{n}_0^{(3)}\right)s^{(2)} - (C_s - 1)^{-1}\left(n - \overline{n}_0^{(3)}\right)s^{(3)} - \\ \dfrac{n - \overline{n}_0^{(3)}}{(C_s - 1)(D-1)}\left\{\sum_{ij\in s}\left(\frac{1}{n_{ij}} - \frac{1}{n}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk} - \sum_{i\in s}\left(\frac{1}{n_i} - \frac{1}{n}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk}\right\}\end{array}\right]$$

$$= \frac{1}{n - \sum_{i\in s}\overline{n}_{0i}^{(2)}}\left[-\sum_{ij\in s}\left(\frac{1}{n_{ij}} - \frac{1}{n}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk} + \sum_{i\in s}\left(\frac{1}{n_i} - \frac{1}{n}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk} + (C_s - 1)s^{(2)} - (D-1)s^{(3)}\right],$$

$$\hat{\sigma}^{2(ht)}_{\eta(3)} = \frac{\left(n - \overline{n}_0^{(3)}\right)^{-1}}{n - \sum_{i\in s}\overline{n}_{0i}^{(2)}}\left[\begin{array}{l}\left(\sum_{i\in s}\overline{n}_{0i}^{(2)} - \overline{n}_0^{(2)}\right)\sum_{ij\in s}\left(\frac{1}{n_{ij}} - \frac{1}{n}\right)\sum_{k\in s}\sigma^2_{\varepsilon(3),ijk} - \left(n - \overline{n}_0^{(2)}\right)\sum_{i\in s}\left(\frac{1}{n_i} - \frac{1}{n}\right)\sum_{jk\in s}\sigma^2_{\varepsilon(3),ijk} \\ -(C_s - 1)\left(\sum_{i\in s}\overline{n}_{0i}^{(2)} - \overline{n}_0^{(2)}\right)s^{(2)} + (D-1)\left(n - \overline{n}_0^{(2)}\right)s^{(3)}\end{array}\right]$$

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

$$n - \sum_{i\in s}\overline{n}_{0i}^{(2)} = n\left(\sum_{ij}w_{ij} - \sum_{i\in s}w_i^{-1}\sum_{j\in s}w_{ij}^2\right),\ (C-1)s^{(2)} = \sum_{ij\in s}n_{ij}\left(\hat{\bar{e}}_{ij.} - \hat{\bar{e}}_{...}\right)^2 = n\sum_{ij\in s}w_{ij}\left(\hat{\bar{e}}_{ij.} - \hat{\bar{e}}_{...}\right)^2,$$

$$(D-1)s^{(3)} = \sum_{i\in s}n_i\left(\hat{\bar{e}}_{i..} - \hat{\bar{e}}_{...}\right)^2 = n\sum_{i\in s}w_i\left(\hat{\bar{e}}_{i..} - \hat{\bar{e}}_{...}\right)^2, \sum_{ij\in s}\left(\frac{1}{n_{ij}} - \frac{1}{n}\right) = n\sum_{ij\in s}w_{ij}\left(1 - w_{ij}\right)n_{ij}^{-2},$$

$$\sum_{i\in s}\left(n_i^{-1} - n^{-1}\right) = n\sum_{i\in s}w_i\left(1 - w_i\right)n_i^{-2},\ n - \overline{n}_0^{(3)} = n\sum_{i\in s}w_i\left(1 - w_i\right), \qquad n - \overline{n}_0^{(2)} = n\sum_{ij\in s}w_{ij}\left(1 - w_{ij}\right),$$

and

$$\sum_{i\in s}\overline{n}_{0i}^{(2)} - \overline{n}_0^{(2)} = n\sum_{i\in s}\left(\frac{1}{w_i} - 1\right)\sum_{j\in s}w_{ij}^2.$$

The estimators $\hat{\sigma}_{u(3)}^{2(ht)}$ and $\hat{\sigma}_{\eta(3)}^{2(ht)}$ shown in Section 2.1 can be obtained by replacing the complex terms with these expressions. The detail derivations of these two indices are given in Das (2016).