Small Area Estimation of Complex Parameters Under Unit-level Models with Skew-Normal Errors

Mamadou S. Diallo * J.N.K. Rao †

Abstract

Complex parameters, such as poverty indicators, are usually difficult to predict in small area estimation (SAE). Elbers et al. (2003) have proposed an empirical semi-parametric method for dealing with poverty indices in SAE. This method, commonly called the ELL method, consists of drawing from the empirical residuals to reconstitute the entire census. After simulating the census, any complex parameter is easily obtained. ELL method has poor MSE performance in many situations even though bias is usually small. Later, Molina and Rao (2010), proposed an empirical best predictor assuming the nested error linear regression model with normally distributed errors. As expected, this estimator can perform poorly when the model errors are not normally distributed. We relax the normality assumption by allowing the errors to follow a skew-normal distribution. Skew-normal (SN) is particularly interesting because it contains the normal distribution as a special case and at the same time it allows departure from symmetry. In this paper, empirical best (EB) estimators are derived assuming skew-normal errors and their performance in terms of MSE is studied relative to the normality-based and ELL predictors.

Key Words: complex parameters, skew-normal, empirical best (EB) estimator

1. Introduction

Increasingly policy makers are interested in local and subpopulation statistics. Limited budgets and area-specific sample sizes often result in direct estimators not being sufficiently reliable for many small areas of interest. Estimation techniques, called small area estimation (SAE), have been developed to improve the estimates where direct estimators are unstable. Rao (2003) gave a comprehensive review of the SAE techniques for estimating mainly linear functions such as means. Most of these SAE methods do not apply to complex statistics such as poverty areas measures.

Elbers et al. (2003) developed an estimation method for complex parameters which assumes the nested error regression model but does not assume any parametric distribution for the area random effects and the unit level errors (semi-parametric approach). This approach, called ELL method, is very attractive because it is free of parametric distribution assumption and very simple to implement. However, the method is not optimal for a given distribution in the sense that it does not yield the empirical best (EB) estimator. Molina and Rao (2010) addressed this issue under the nested error regression model with normally distributed errors. They provided a Monte Carlo approximation of the EB estimator. Their method can be extended to nested error models with the random errors following other distributions than the normal.

In this paper, we provide the EB estimator under the nested error model with the random errors following skew-normal (SN) distributions. SN distributions are very interesting because they relax the normality assumption by allowing asymmetrical shape and at the same time they include normal distribution as a special case. In section 2, we define the nested error model with the errors following SN distributions. In Section 3, we derive the EB estimator for complex parameters. Given the complexity of the EB estimator, we proposed

*Westat, 1600 Research Boulevard, Rockville MD 20850, U.S.A

†School of Mathematics and Statistics, Carleton University, Ottawa K1S 5B6, Canada

a simpler approach based on conditioning on the area effects. We also propose some improvements to the ELL method by correctly assigning the random area effects and using different methods to estimate the area effects without assuming any parametric distribution. In Section 4, we conducted a simulation study to compare the different estimators discussed in this paper and the direct estimator.

2. SAE Model under Skew-Normal Errors

The SN distribution allows relaxing the symmetry assumption of the randoms errors induced by the use of normal distribution. Consider a random variable $Z \sim SN(0,1,\lambda)$, referred in this paper as $SN(\lambda)$, with probability density function given by

$$f(z;\lambda) = 2\phi(z)\Phi(\lambda z) \quad z \in \mathbb{R} \tag{1}$$

where $\phi(z)$ and $\Phi(z)$ denote the $N(0,1)$ pdf and the cumulative distribution function (cdf), respectively. The parameter λ controls the asymmetry of the distribution and varies in $(-\infty,\infty)$. One attractive feature of the SN class of distributions is the fact that it includes the normal distribution as a special case. To see this, note that if $\lambda = 0$ then $f(z;0) = \phi(z)$ i.e. $Z \sim N(0,1)$. The moment generating function (mgf) of Z is $M(t) = 2\exp(t^2/2)\Phi(\delta t)$ where $\delta = \lambda/\sqrt{1+\lambda^2}$. Hence, the first and second central moments are $\mathbb{E}(Z) = \delta\sqrt{2/\pi}$ and $\mathbb{V}ar(Z) = 1 - 2\delta^2/\pi$ respectively. It can be shown that if X_0 and X_1 are independent standard normal, $N(0,1)$, variables and $\delta \in (-1,1)$ then

$$Z = \delta|X_0| + (1-\delta^2)^{1/2}X_1 \tag{2}$$

is $SN(\lambda(\delta))$. Using representation (2), we can generate easily values from $SN(\lambda)$ by generating two values x_0 and x_1 from independent $N(0,1)$ and applying the weighted average in (2). The distribution $SN(\lambda)$ can be generalized further by adding location and scale (dispersion) parameters μ and σ^2, respectively. The pdf of the random variable $Y = \mu + \sigma Z$ is

$$f(y;\mu,\sigma,\lambda) = \frac{2}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\lambda\frac{y-\mu}{\sigma}\right) \quad y \in \mathbb{R} \tag{3}$$

We denote the distribution defined by the pdf (3) by $SN(\mu,\sigma^2,\lambda)$ i.e $Y \sim SN(\mu,\sigma^2,\lambda)$. Note that if Y follows the distribution defined by (3) then $E(Y) = \mu + \delta\sigma\sqrt{2/\pi}$. Hence, the distribution $SN(-\delta\sigma\sqrt{2/\pi},\sigma^2,\lambda)$ has zero mean. Figure 1 shows the departure of the SN distribution from the normal as the parameter λ moves away from zero. The half-normal distribution is the limiting distribution as λ tends to ∞.

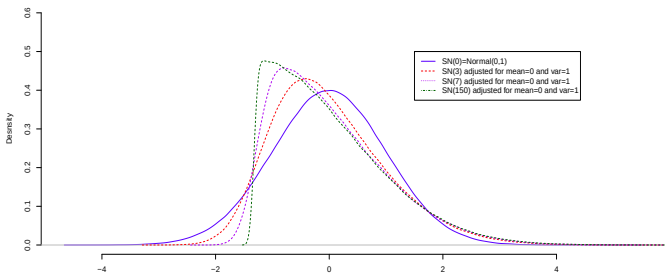

Figure 1: Probability densities of SN.

The nested error regression model was first used in the context of SAE by Battese et al. (1988) to estimate crop areas under corn and soybeans for 12 counties in north-central Iowa using farm-interview data and satellite information. They assumed that the random errors follow normal distributions. To relax the normality assumption, we consider the nested error model

$$Y_{dj} = \mathbf{x}_{dj}^T \beta + u_d + e_{dj}, \quad j = 1,\ldots,N_d, \quad d = 1,\ldots,m, \tag{4}$$

where u_d and e_{dj} are independent for any area d and element j in area d and satisfy

$$u_d \overset{iid}{\sim} SN(-\delta_u \sigma_u \sqrt{2/\pi}, \sigma_u^2, \lambda_u), \quad \text{and} \quad e_{dj} \overset{iid}{\sim} SN(-\delta_e \sigma_e \sqrt{2/\pi}, \sigma_e^2, \lambda_e). \tag{5}$$

Note that the random errors u_d and e_{dj} in (5) have zero mean. The nested error model with errors following SN distributions defined by (4)-(5) is referred in this paper as the *SN model*. When $\lambda_u = \lambda_e = 0$, the SN model reduces to the usual nested error model where the errors are normally distributed which we refer to as the *normal model*.

Since the model parameters are unknown, we need to estimate them in order to compute the "optimal" estimators developed in Section 3. The vector of model parameters to estimate is $\theta = \left(\beta, \sigma_u^2, \sigma_e^2, \lambda_u, \lambda_e\right)^T$. Let us consider the partition $\mathbf{Y}_d = \left(\mathbf{Y}_{dr}^T, \mathbf{Y}_{ds}^T\right)^T$, where $\mathbf{Y}_{dr}$ is the out-of-sample characteristic vector of interest while $\mathbf{Y}_{ds}$ is the observed vector of interest from the sample. Parameter estimation is performed using the sample data and assuming a non-informative sampling design. The first step is to determine the joint distribution of the sample vector $\mathbf{Y}_s$. Given that the random effects u_d are independent for any different values of d, the joint pdf $f(\mathbf{y}_s) = \prod_{d=1}^m f(\mathbf{y}_{ds})$. The problem reduces to finding the distribution of $\mathbf{Y}_{ds}$. We now introduce a multivariate extension of the SN distribution called the closed skew-normal (CSN) (see Chapter 2 of Genton (2004) for more details on CSN).

Definition 2.1. *Consider $p \geq 1$, $q \geq 1$, $\mu \in \mathbb{R}^p$, $\nu \in \mathbb{R}^q$, $\mathbf{D}$ an arbitrary $q \times p$ matrix, Σ and Γ positive definite matrices of dimensions $p \times p$ and $q \times q$, respectively. Then the probability density function of the CSN distribution is given by:*

$$f_{p,q}(\mathbf{y}) = C \phi_p\left(\mathbf{y}; \mu, \Sigma\right) \Phi_q\left(\mathbf{D}\left(\mathbf{y} - \mu\right); \nu, \Gamma\right), \quad \mathbf{y} \in \mathbb{R}^p, \tag{6}$$

with $C^{-1} = \Phi_q\left(\mathbf{0}; \nu, \Gamma + \mathbf{D}\Sigma\mathbf{D}^T\right)$ where ϕ_p and Φ_p are respectively the pdf and the cdf of the p-dimensional normal distribution. We denote this distribution by $\mathbf{y} \sim CSN_{p,q}\left(\mu, \Sigma, \mathbf{D}, \nu, \Gamma\right)$ and if $p = q$ we will denote it by $CSN_p\left(\mu, \Sigma, \mathbf{D}, \nu, \Gamma\right)$.

Letting $p = q = 1$, $\mu = \mu$, $\Sigma = \sigma^2$, $\nu = 0$, and $\Gamma = 1$ leads to the pdf of the CSN defined by (6) to equal the SN pdf given by (3). That is $SN(\mu, \sigma^2, \lambda) \equiv CSN(\mu, \sigma^2, \lambda, 0, 1)$. The attractive aspect of the CSN distribution is the set of closure properties similar to the multivariate normal. That is the vector resulting from linear combination, marginal and conditional operations, joint and sum of independent CSN vectors has a CSN distribution. Consider $\mathbf{Y} \sim CSN_{p,q}\left(\mu, \Sigma, \mathbf{D}, \nu, \Gamma\right)$, it follows from these closure properties that

$$\mathbf{Y}_{ds} \sim CSN_{n_d, n_d+1}\left(\mu_{Y_{ds}}, \Sigma_{Y_{ds}}, \mathbf{D}_{Y_{ds}}, \nu_{Y_{ds}}, \Gamma_{Y_{ds}}\right), \quad d = 1,\ldots,m. \tag{7}$$

where $\mu_{Y_{ds}} = \mathbf{X}_{ds}\beta + \mu_u \mathbf{1}_{n_d} + \mu_e$, $\quad \Sigma_{Y_{ds}} = \sigma_e^2 \mathbf{I}_{n_d} + \sigma_u^2 \mathbf{1}_{n_d}\mathbf{1}_{n_d}^T$, $\quad \nu_{Y_{ds}} = \mathbf{0}_{n_d+1}$,

$\mathbf{D}_{Y_{ds}} = \begin{bmatrix} \frac{\lambda_e}{\sigma_e}\left[\mathbf{I}_{n_d} - \gamma_{n_d}\mathbf{1}_{n_d}\mathbf{1}_{n_d}^T\right] \\ \frac{\lambda_u}{\sigma_u}\gamma_{n_d}\mathbf{1}_{n_d}^T \end{bmatrix}$, and $\Gamma_{Y_{ds}} = \begin{bmatrix} \mathbf{I}_{n_d} + \lambda_e^2\gamma_{n_d}\mathbf{1}_{n_d}\mathbf{1}_{n_d}^T & -\lambda_e\lambda_u\left(\frac{\sigma_e}{\sigma_u}\right)\gamma_{n_d}\mathbf{1}_{n_d} \\ -\lambda_e\lambda_u\left(\frac{\sigma_e}{\sigma_u}\right)\gamma_{n_d}\mathbf{1}_{n_d}^T & 1 + \lambda_u^2(1 - n_d\gamma_{n_d}) \end{bmatrix}$.

Up to a constant, the likelihood function associated with the model (7) is equal to:

$$l(\theta|\mathbf{y}) = -\frac{1}{2}\sum_{d=1}^{m} ln|\Sigma_{Y_{ds}}| - \frac{1}{2}\sum_{d=1}^{m}\left(\mathbf{y}_{ds}-\mu_{Y_{ds}}\right)^T\Sigma_{Y_{ds}}^{-1}\left(\mathbf{y}_{ds}-\mu_{Y_{ds}}\right) + \sum_{d=1}^{m} ln\left(\Phi_{n_d+1}\left(\mathbf{D}_{Y_{ds}}\left(\mathbf{y}_{ds}-\mu_{Y_{ds}}\right);\nu_{Y_{ds}},\Gamma_{Y_{ds}}\right)\right) \quad (8)$$

Optimization routines such as *optim* from the Statistical software R can be used to maximize the likelihood function (8) to get the maximum likelihood (ML) estimators of the model parameters.

3. Prediction of Complex Small Area Parameters

A complex small area parameter is defined as $\eta_d = h(\mathbf{y}_d)$ where h is a nonlinear function of $\mathbf{y}_d$. We study below three methods for estimating η_d under the SN model in (4)-(5). The first method, called the marginal approach, provides a Monte Carlo approximation of the best predictor. The second method, called the conditional approach, gives the best predictor of η_d conditionally on the area effect u_d. And last, we propose some improvements to the ELL method.

3.1 Marginal Approach

In SAE, the practice is to use the predictor that minimizes the MSE to estimate the small area parameter η_d. This best estimator, in terms of minimizing the MSE, is the conditional expectation

$$\hat{\eta}_d^B = \mathbb{E}(\eta_d|\mathbf{y}_{ds}) = \int h(\mathbf{y}_d) f_{dr|s}(\mathbf{y}) d\mathbf{y} \quad (9)$$

where $f_{dr|s} \equiv f_{\mathbf{Y}_{dr}|\mathbf{y}_{ds}}$ is the pdf of the distribution of $\mathbf{Y}_{dr}$ given $\mathbf{y}_{ds}$. This conditional distribution is derived under the assumption that there is no sample selection bias i.e the population model holds for the sample. The difficulty with the estimator (9) is the lack of a closed-form expression in most situations due to the complexity of the function h. Molina and Rao (2010) proposed a Monte Carlo approximation of (9) which consists of the following steps: 1) draw out-of-sample vectors $\mathbf{y}_{dr}^{(\ell)}$, $\ell = 1,...,L$ from the conditional distribution $f_{dr|s}$, 2) reconstruct the censuses $\mathbf{y}_d^{(\ell)} = \left(\mathbf{y}_{dr}^{(\ell)T}, \mathbf{y}_{ds}^T\right)^T$, and 3) from each simulated census estimate the small area parameter η_d by $\hat{\eta}_d^{(\ell)} = h(\mathbf{y}_d^{(\ell)})$. A Monte Carlo approximation of the best predictor $\hat{\eta}_d^B$, we call the quasi-best (QB) estimator, is then given by

$$\hat{\eta}_d^{QB} = \frac{1}{L}\sum_{\ell=1}^{L}\hat{\eta}_d^{(\ell)}. \quad (10)$$

Under the SN model (4)-(5), the conditional distribution of $\mathbf{Y}_{dr}$ given $\mathbf{y}_{ds}$ is obtained from the closure properties of the CSN as a member of the CSN family. We have

$$\mathbf{Y}_{dr|s} \sim CSN_{N_{dr},N_d+1}\left(\mu_{dr|s}, \Sigma_{dr|s}, \mathbf{D}_{dr|s}, \nu_{dr|s}, \Gamma_{dr|s} = \Gamma_{Y_d}\right) \quad (11)$$

where $\mu_{dr|s} = \mathbf{X}_{dr}\beta + \mu_u \mathbf{1}_{N_{dr}} + \mu_{edr} + \gamma_{n_d}\left[\mathbf{1}_{n_d}^T(\mathbf{y}_{ds}-\mu_{ds})\right]\mathbf{1}_{N_{dr}}$, $\Sigma_{dr|s} = \sigma_e^2\left[\mathbf{I}_{N_{dr}} + \gamma_{n_d}\mathbf{1}_{N_{dr}}\mathbf{1}_{N_{dr}}^T\right]$,

$$\mathbf{D}_{dr|s} = \begin{bmatrix} \frac{\lambda_e}{\sigma_e}\left[\begin{bmatrix}\mathbf{I}_{N_{dr}}\\ \mathbf{0}_{(n_d\times N_{dr})}\end{bmatrix} - \gamma_{N_d}\mathbf{1}_{N_d}\mathbf{1}_{N_{dr}}^T\right] \\ \frac{\lambda_u}{\sigma_u}\gamma_{N_d}\mathbf{1}_{N_{dr}}^T \end{bmatrix}, \nu_{dr|s} = -\begin{bmatrix} \mathbf{0}_{(N_{dr}\times n_d)} \\ \frac{\lambda_e}{\sigma_e}\left[\mathbf{I}_{n_d} - \gamma_{n_d}\mathbf{1}_{n_d}\mathbf{1}_{n_d}^T\right] \\ \frac{\lambda_u}{\sigma_u}\gamma_{n_d}\mathbf{1}_{n_d}^T \end{bmatrix}(\mathbf{y}_{ds}-\mu_{ds}),$$

and $\Gamma_{Y_d} = \begin{bmatrix} \mathbf{I}_{N_d} + \lambda_e^2 \gamma_{N_d} \mathbf{1}_{N_d} \mathbf{1}_{N_d}^T & -\lambda_e \lambda_u \left(\frac{\sigma_e}{\sigma_u}\right) \gamma_{N_d} \mathbf{1}_{N_d} \\ -\lambda_e \lambda_u \left(\frac{\sigma_e}{\sigma_u}\right) \gamma_{N_d} \mathbf{1}_{N_d}^T & 1 + \lambda_u^2 (1 - N_d \gamma_{N_d}) \end{bmatrix}$.

The random vector in (11) has a dimension of $N_{dr} = N_d - n_d$ which can be very large in many applications. Hence, it is important to find a way to draw from the conditional distribution (11) in a univariate manner. First consider

$$\mathbf{V}_0 \sim LTN_q(\nu, \mathbf{0}, \Psi_0) \quad \text{and} \quad \mathbf{V}_1 \sim N_p(\mathbf{0}, \Psi_1) \tag{12}$$

where Ψ_0 and Ψ_1 are covariance matrices and the notation $LTN_q(\mathbf{c}, \mu, \Sigma)$ denotes a multivariate normal truncated below $\mathbf{c}$. Then consider the transformed random vector

$$\mathbf{V} = \mu + \mathbf{B}_0 \mathbf{V}_0 + \mathbf{B}_1 \mathbf{V}_1 \tag{13}$$

where $\mathbf{B}_0 = (\mathbf{D}\Sigma)^T \Psi_0^{-1}$ and $\mathbf{B}_1$ is a $p \times p$ matrix such that $\mathbf{B}_1 \Psi_1 \mathbf{B}_1^T = \Sigma - (\mathbf{D}\Sigma)^T \Psi_0^{-1} (\mathbf{D}\Sigma)$. Arellano-Valle and Azzalini (2006) showed that the random vector $\mathbf{V}$ has the density function:

$$f_{p,q}(\mathbf{v}) = \phi_p(\mathbf{v}; \mu, \Sigma) \frac{\Phi_q\left(\mathbf{D}(\mathbf{v} - \mu); \nu, \Psi_0 - \mathbf{D}\Sigma^{-1}\mathbf{D}^T\right)}{\Phi_q(\mathbf{0}; \nu, \Psi_0)}. \tag{14}$$

Note that if $p = N_{dr}$, $q = N_d + 1$, $\mu = \mu_{dr|s}$, $\Sigma = \Sigma_{dr|s}$, $\mathbf{D} = \mathbf{D}_{dr|s}$, $\nu = \nu_{dr|s}$, and $\Psi_0 = \Gamma_{dr|s} + \mathbf{D}_{dr|s} \Sigma_{dr|s} \mathbf{D}_{dr|s}^T$ then equation (14) defines the pdf of the conditional distribution (11). In other words, if $\mathbf{V}_{dr|s}^0$ and $\mathbf{V}_{dr|s}^1$ are independent random vectors such that:

$$\mathbf{V}_{dr|s}^0 \sim LTN_{N_d+1}\left(\nu_{dr|s}; \mathbf{0}, \Psi_{dr|s}^0 = \Gamma_{dr|s} + \mathbf{D}_{dr|s} \Sigma_{dr|s} \mathbf{D}_{dr|s}^T\right) \tag{15}$$

$$\mathbf{V}_{dr|s}^1 \sim N_{N_{dr}}\left(\mathbf{0}, \Psi_{dr|s}^1 = \Sigma_{dr|s} - \left(\mathbf{D}_{dr|s} \Sigma_{dr|s}\right)^T \left(\Psi_{dr|s}^0\right)^{-1} \mathbf{D}_{dr|s} \Sigma_{dr|s}\right), \tag{16}$$

then the random vector $\mathbf{Y}_{dr|s}$ defined in (11) can be obtained using the transformation:

$$\mathbf{Y}_{dr|s} = \mu_{dr|s} + \mathbf{B}_{dr|s}^0 \mathbf{V}_{dr|s}^0 + \mathbf{V}_{dr|s}^1, \tag{17}$$

where the matrix $\mathbf{B}_{dr|s}^0$ is defined as follows

$$\mathbf{B}_{dr|s}^0 = (\mathbf{D}_{dr|s} \Sigma_{dr|s})^T \Psi_0^{-1} = \left[\frac{\lambda_e \sigma_e}{1 + \lambda_e^2} \mathbf{I}_{N_{dr}} \quad b_{01} \mathbf{1}_{N_{dr}} \mathbf{1}_{n_d}^T \quad b_{02} \mathbf{1}_{N_{dr}}\right], \tag{18}$$

with

$$b_{01} = \frac{-\gamma_{n_d} \lambda_e \sigma_e \sigma_u^2}{(1 + n_d \gamma_{n_d} \lambda_e^2)\sigma_u^2 + \gamma_{n_d} \lambda_u^2 \sigma_e^2} \quad \text{and} \quad b_{02} = \frac{\gamma_{n_d} \lambda_u \sigma_u \sigma_e^2}{(1 + n_d \gamma_{n_d} \lambda_e^2)\sigma_u^2 + \gamma_{n_d} \lambda_u^2 \sigma_e^2} \tag{19}$$

Using expression (17), univariate draws from $\mathbf{Y}_{dr|s}$ reduce to univariate generations from $\mathbf{V}_{dr|s}^0$ and $\mathbf{V}_{dr|s}^1$. The nested error model assumption leads to special forms of the matrices involved in the best predictor (17). Under this model we have

$$\Psi_{dr|s}^0 = \begin{bmatrix} (1 + \lambda_e^2)\mathbf{I}_{N_{dr}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_d} + \lambda_e^2 \gamma_{n_d} \mathbf{1}_{n_d} \mathbf{1}_{n_d}^T & -\lambda_e \lambda_u \left(\frac{\sigma_e}{\sigma_u}\right) \gamma_{n_d} \mathbf{1}_{n_d} \\ \mathbf{0} & -\lambda_e \lambda_u \left(\frac{\sigma_e}{\sigma_u}\right) \gamma_{n_d} \mathbf{1}_{n_d}^T & 1 + \lambda_u^2 \left(\frac{\sigma_e}{\sigma_u}\right)^2 \gamma_{n_d} \end{bmatrix}, \tag{20}$$

and

$$\Psi_{dr|s}^1 = \alpha \mathbf{I}_{N_{dr}} + \beta \mathbf{1}_{N_{dr}} \mathbf{1}_{N_{dr}}^T \tag{21}$$

with α and β defined as follows

$$\alpha = \sigma_e^2\left(1-\frac{\lambda_e^2}{1+\lambda_e^2}\right) = \sigma_e^2\left(1-\delta_e^2\right), \quad \text{and} \quad \beta = \frac{\gamma_{n_d}\sigma_u^2\sigma_e^2}{\left(1+n_d\gamma_{n_d}\lambda_e^2\right)\sigma_u^2+\gamma_{n_d}\lambda_u^2\sigma_e^2} \tag{22}$$

The random vector $\mathbf{V}_{dr|s}^1$ is easily generated in a univariate manner by

$$\left(\mathbf{V}_{dr|s}^1\right)_j = v_{dj}^{11}+v_d^{10}, \quad j=1,...,n_d, d=1,...,m,$$

where $v_{dj}^{11} \overset{iid}{\sim} N(0,\alpha)$ is independent of $v_d^{10} \overset{ind}{\sim} N(0,\beta)$ and α and β are defined as in (22). Note that, for each small area d, we only generate one value v_d^{10} and n_d values v_{dj}^{11}, $j=1,...,n_d$. The random vector $\mathbf{V}_{dr|s}^0$ can be decomposed into two components $\mathbf{V}_{dr|s}^{0r}$ and $\mathbf{V}_{dr|s}^{0s}$ where

$$\mathbf{V}_{dr|s}^0 = \begin{pmatrix} \mathbf{V}_{dr|s}^{0r} \\ \mathbf{V}_{dr|s}^{0s} \end{pmatrix} \sim LTN_{N_d+1}\left(\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\nu}_{dr|s}^{0s} \end{pmatrix}; \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{bmatrix} \Psi_{dr|s}^{0r} & \mathbf{0} \\ \mathbf{0} & \Psi_{dr|s}^{0s} \end{bmatrix}\right) \tag{23}$$

and the components of (23) are defined as follows:

$$\boldsymbol{\nu}_{dr|s}^{0s} = -\begin{bmatrix} \frac{\lambda_e}{\sigma_e}\left[\mathbf{I}_{n_d}-\gamma_{n_d}\mathbf{1}_{n_d}\mathbf{1}{n_d}^T\right] \\ \frac{\lambda_u}{\sigma_u}\gamma_{n_d}\mathbf{1}{n_d}^T \end{bmatrix}\left(\mathbf{y}_{d_s}-\boldsymbol{\mu}_{d_s}\right) \tag{24}$$

$$\Psi_{dr|s}^{0r} = (1+\lambda_e^2)\mathbf{I}_{N_{dr}} \tag{25}$$

$$\Psi_{dr|s}^{0s} = \begin{bmatrix} \mathbf{I}_{N_{dr}}+\lambda_e^2\gamma_{n_d}\mathbf{1}_{n_d}\mathbf{1}_{n_d}^T & -\lambda_e\lambda_u\left(\frac{\sigma_e}{\sigma_u}\right)\gamma_{n_d}\mathbf{1}_{n_d} \\ -\lambda_e\lambda_u\left(\frac{\sigma_e}{\sigma_u}\right)\gamma_{n_d}\mathbf{1}_{n_d}^T & 1+\lambda_u^2\left(\frac{\sigma_e}{\sigma_u}\right)^2\gamma_{n_d} \end{bmatrix} \tag{26}$$

$\mathbf{V}_{dr|s}^{0r}$ is simply a vector of uncorrelated half-normal variables since $\Psi_{dr|s}^{0r} = (1+\lambda_e^2)\mathbf{I}_{N_{dr}}$ is a diagonal covariance matrix. Therefore, to obtain a realization of $\mathbf{V}_{dr|s}^{0r}$, it is sufficient to generate univariate normally distributed values and take their absolute values. In other words,

$$\left(\mathbf{V}_{dr|s}^{0r}\right)_j = |v_{dj}^{0r}|, \quad j=1,...,N_{dr}, d=1,...,m, \tag{27}$$

where $v_{dj}^{0r} \overset{iid}{\sim} N\left(0,1+\lambda_e^2\right)$. The only vector left to be generated is $\mathbf{V}_{dr|s}^{0s}$. Unfortunately there is no reasonable known approach for generating the vector $\mathbf{V}_{dr|s}^{0s}$ in a univariate manner because $\boldsymbol{\nu}_{dr|s}^{0s}$ is not equal to the zero vector and the covariance matrix $\Psi_{dr|s}^{0s}$ is not diagonal. However the length of this vector is n_d+1 which is easily manageable in a multivariate way especially given the small sample sizes, n_d, involved in SAE. The vector $\mathbf{V}_{dr|s}^{0s}$ will therefore be generated using the multivariate truncated normal. Most of the major statistical software packages offer procedures for generating univariate truncated normals and some of them have routines for generating from multivariate truncated distributions. For the simulation study presented in Section 4, the R-package tmvtnorm (Truncated Multivariate Normal and Student t Distribution) was used to generate data from multivariate truncated normal distributions. In summary the quasi-univariate procedure below shows how to generate the predicted values of the elements of $\mathbf{y}_{dr|s}$ for the non-sampled units in small area d:

1. Estimate the unknown vector of model parameters $\theta = \left(\beta, \sigma_u^2, \sigma_e^2, \lambda_u, \lambda_e\right)$ using the ML method and the sample data $\mathbf{y}_s$.

2. Generate values v_{dj}^{11} from $N(0,\alpha)$ and v_d^{10} from $N(0,\beta)$ independently, where α and β are defined in (22) with θ replaced by its estimator $\hat{\theta}$. Note that only one v_d^{10} is generated for each small area d but a different v_{dj}^{11} is draw for each unit j in small area d.

3. Generate value v_{dj}^{0r} from $N(0, 1+\hat{\lambda}_e^2)$.

4. Generate vector $\mathbf{v}_{dr|s}^{0s}$ from $LTN_{n_d+1}\left(\nu_{dr|s}^{0s}; \mathbf{0}, \Psi_{dr|s}^{0s}\right)$, where $\nu_{dr|s}^{0s}$ and $\Psi_{dr|s}^{0s}$ are respectively defined in (24) and (26) with θ replaced by $\hat{\theta}$.

5. Create the vector $\mathbf{w}_{dr|s}^{0s} = \left(b_{01}\mathbf{1}_{N_{dr}}\mathbf{1}_{n_d}^T \quad b_{02}\mathbf{1}_{N_{dr}}\right)\mathbf{v}_{dr|s}^{0s}$, where b_{01} and b_{02} are defined in (19), with the parameters replaced by their estimators.

6. The element j of the vector $y_{dr|s}$ is:

$$(y_{dr|s})_j = (\hat{\mu}_{dr|s})_j + v_{dj}^{11} + v_d^{10} + \frac{\hat{\lambda}_e\hat{\sigma}_e}{1+\hat{\lambda}_e^2}|v_{dj}^{0r}| + (w_{dr|s}^{0s})_{dj},$$

where $(\hat{\mu}_{dr|s})_j = \mathbf{X}_{drj}\hat{\beta} + \hat{\mu}_u + \hat{\mu}_e + \hat{\gamma}_{n_d}\left[\mathbf{1}_{n_d}^T(\mathbf{y}_{ds} - \hat{\mu}_{ds})\right]$

3.2 Conditional Approach

In SAE, the goal is to estimate the conditional parameter or area specific $\eta_d(u_d) = \mathbb{E}(h(\mathbf{Y}_d)|\mathbf{y}_{ds}, u_d, \theta)$ because only one population is "realized" and prediction is conducted conditionally on the given fixed population. The conditional distribution of $\mathbf{Y}_d$ given u_d is:

$$\mathbf{Y}_d|u_d \sim CSN_{N_d,N_d}\left(\mathbf{X}_d\beta + u_d\mathbf{1}_{N_d} + \mu_{ed}, \sigma_e^2\mathbf{I}_{N_d}, \frac{\lambda_e}{\sigma_e}\mathbf{I}_{N_d}, \mathbf{0}_{N_d}, \mathbf{I}_{N_d}\right) \tag{28}$$

Note that the element of the vector $\mathbf{Y}_d|u_d$ are independent and we may write

$$(\mathbf{Y}_d|u_d)_j \sim SN\left(\mathbf{X}_{dj}\beta + u_d + \mu_e, \sigma_e^2, \frac{\lambda_e}{\sigma_e}\right), \quad j \in s_d^c \tag{29}$$

where s_d^c indicates the out-of-sample units. The distribution in (29) is the distribution of the unit level error e_{dj} with a different location parameter. Therefore, generating the out-of-sample values $(\mathbf{y}_{dr|s})_j = (\mathbf{y}_d|u_d)_j$, where $j \in s_d^c$, for the simulated census reduces to drawing from a univariate SN distribution. Unfortunately u_d is not observed. Hence u_d is predicted by the best estimator $\hat{u}_d^B$ and used to obtain prediction of $\eta(u_d)$ as $\eta(\hat{u}_d^B)$ or the BLUP estimator $\hat{u}_d^{BLUP}$ may be used to get $\eta(\hat{u}_d^{BLUP})$.

Only the mean and the covariance matrix of the joint distribution of u_d and $\mathbf{y}_{ds}$ is needed to define the BLUP of u_d as

$$\hat{u}_d^{BLUP} = \mathbb{E}(u_d) + C_d\Sigma_d^{-1}(\mathbf{Y}_{ds} - \mathbb{E}(\mathbf{Y}_{ds})) \tag{30}$$

where $C_d = \mathbb{C}ov(u_d, \mathbf{Y}_{ds})$ and $\Sigma_d^{-1} = [\mathbb{C}ov(\mathbf{Y}_{ds})]^{-1}$. Noting that $\mathbb{E}(u_d) = 0$, $C_d = \sigma_u^2(1 - \frac{2}{\pi}\delta_u^2)\mathbf{1}_{Nd}^T$, $\Sigma_d^{-1} = \frac{1}{\sigma_e^2(1-\frac{2}{\pi}\delta_e^2)}(\mathbf{I}_{n_d} - \frac{\sigma_u^2(1-\frac{2}{\pi}\delta_u^2)}{\sigma_e^2(1-\frac{2}{\pi}\delta_e^2)+n_d\sigma_u^2(1-\frac{2}{\pi}\delta_u^2)}\mathbf{1}_{n_d}\mathbf{1}_{n_d}^T)$, and $\mu_{Y_d} = \mathbf{X}_d\beta$ leads to

$$\hat{u}_d^{BLUP} = \frac{\sigma_u^2(1-\frac{2}{\pi}\delta_u^2)}{\sigma_e^2(1-\frac{2}{\pi}\delta_e^2) + n_d\sigma_u^2(1-\frac{2}{\pi}\delta_u^2)}\sum_{j\in s_d}(y_{dj} - \mathbf{x}_{dj}^T\tilde{\beta}) \tag{31}$$

Note that the errors are adjusted so that their means are equal to zero. On the other hand, the best estimator $\hat{u}_d^B$ requires the joint distribution of u_d and $\mathbf{Y}_{ds}$. The best predictor of the small area random effect, u_d, is now obtained, using Theorem 3.1.

Theorem 3.1. *For the SN nested error model (4) with adjusted errors so that* $\mathbb{E}(u_d) = 0$ *and* $\mathbb{E}(\mathbf{e}_d) = \mathbf{0}$*, the best predictor of* u_d *is:*

$$\hat{u}_d^B = \mu_u + \gamma_{n_d} \sum_{j \in s_d} \{y_{dj} - (\mathbf{x}_{dj}^T \beta + \mu_u + \mu_e)\} + \frac{\Phi_{n_d+1}^{(1)}(\mathbf{0}; \nu_A, \Gamma)}{\Phi_{n_d+1}(\mathbf{0}; \nu_A, \Gamma)} \tag{32}$$

where

$$\Gamma = \begin{bmatrix} \mathbf{I}_{n_d} + \lambda_e^2 \gamma_{n_d} \mathbf{1}_{n_d} \mathbf{1}_{n_d}^T & -\lambda_e \lambda_u \left(\frac{\sigma_e}{\sigma_u}\right) \gamma_{n_d} \mathbf{1}_{n_d} \\ -\lambda_e \lambda_u \left(\frac{\sigma_e}{\sigma_u}\right) \gamma_{n_d} \mathbf{1}_{n_d}^T & 1 + \lambda_u^2 (1 - n_d \gamma_{n_d}) \end{bmatrix}, \text{ and}$$

$\Phi_{n_d+1}^{(1)}(\mathbf{0}; \nu_A, \Gamma) = \frac{\partial}{\partial t} \Phi_{n_d+1}(\mathbf{D}_A \sigma_A t; \nu_A, \Gamma)_{t=0}$, *with*

$\mu_A = \mu_u + \gamma_{n_d} \mathbf{1}_{n_d}^T (\mathbf{y}_{n_d} - (\mathbf{X}_d \beta + \mu_u \mathbf{1}_{n_d} + \mu_{eds}))$, $\sigma_A = \sigma_e^2 \gamma_{n_d}$, $\mathbf{D}_A = \begin{pmatrix} -\frac{\lambda_e}{\sigma_e} \mathbf{1}_{n_d} \\ \frac{\lambda_u}{\sigma_u} \end{pmatrix}$,

$\nu_A = -\begin{bmatrix} \frac{\lambda_e}{\sigma_e} [\mathbf{I}_{n_d} - \gamma_{n_d} \mathbf{1}_{n_d} \mathbf{1}_{n_d}^T] \\ \frac{\lambda_u}{\sigma_u} \gamma_{n_d} \mathbf{1}_{n_d}^T \end{bmatrix} (\mathbf{y}_{n_d} - (\mathbf{X}_d \beta + \mu_u \mathbf{1}_{n_d} + \mu_{eds}))$, *and* $\Gamma_A = \mathbf{I}_{n_d+1}$.

Proof. Let $\mathbf{T} = \begin{pmatrix} \mathbf{Y}_{ds} \\ u_d \end{pmatrix} \sim CSN_{n_d+1, n_d+1}(\mu_T, \Sigma_T, \mathbf{D}_T, \nu_T, \Gamma_T)$ where

$$\mu_T = \begin{pmatrix} \mu_{Y_d} = \mathbf{X}_d \beta + \mu_u \mathbf{1}_{n_d} + \mu_{eds} \\ \mu_u \end{pmatrix}, \quad \Sigma_T = \begin{bmatrix} \sigma_e^2 \mathbf{I}_{n_d} + \sigma_u^2 \mathbf{1}_{n_d} \mathbf{1}_{n_d}^T & \sigma_u^2 \mathbf{1}_{n_d} \\ \sigma_u^2 \mathbf{1}_{n_d}^T & \sigma_u^2 \end{bmatrix},$$

$$\mathbf{D}_T = \begin{bmatrix} -\frac{\lambda_e}{\sigma_e} \mathbf{1}_{n_d} & \frac{\lambda_e}{\sigma_e} \mathbf{I}_{n_d} \\ \frac{\lambda_u}{\sigma_u} & \mathbf{0} \end{bmatrix}, \quad \nu_T = \mathbf{0}, \quad \text{and} \quad \Gamma_T = \mathbf{I}_{n_d+1}$$

then, applying the closure properties of the CSN, we have
$u_d \sim SN(\mu_u, \sigma_u^2, \lambda_u) \equiv CSN_1(\mu_u, \sigma_u^2, \lambda_u, 0, 1)$ and $\mathbf{Y}_{ds}$ as in (7). Also, the closure properties give the distribution of $u_d | \mathbf{y}_{ds}$ as a CSN, precisely, we have

$$u_d | \mathbf{y}_{ds} \sim CSN_{1, n_d+1}(\mu_A, \sigma_A, \mathbf{D}_A, \nu_A, \Gamma_A)$$

where μ_A, σ_A, $\mathbf{D}_A$, ν_A defined as in Theorem 3.1. Hence, the mgf is
$M_{u_d|\mathbf{y}_{ds}}(t) = \frac{\Phi_{n_d+1}(\mathbf{D}_A \sigma_A t; \nu_A, \Gamma_A + \sigma_A \mathbf{D}_A \mathbf{D}_A^T)}{\Phi_{n_d+1}(\mathbf{0}; \nu_A, \Gamma_A + \sigma_A \mathbf{D}_A \mathbf{D}_A^T)} e^{t\mu_A + \frac{1}{2}\sigma_A t^T t}$ and the best predictor is obtained as the expectation using $\mathbb{E}(X) = \frac{\partial}{\partial t} M(t)_{t=0}$. □

In practice, the parameters of the model are estimated using a suitable method to obtain EB estimator $\hat{u}_d^{EB}$ and EBLUP estimator $\hat{u}_d^{EBLUP}$. This leads to the empirical predictors $\hat{\eta}(\hat{u}_d^{EB})$ and $\hat{\eta}(\hat{u}_d^{EBLUP})$. We refer to these latter estimators as respectively $\hat{\eta}_d^{C-EQB}$ and $\hat{\eta}_d^{C-EQBLUP}$, where C refers to conditional. The superscript $C-EQBLUP$ is a little confusing because $\hat{\eta}_d^{C-EBLUP}$ is not linear nor unbiased; the subscript just refers to the way the area effect was estimated to distinguish between the two estimators. In practice, $\hat{\eta}_d^{C-EQB}$ should be preferred between the two because it approximates the conditional best predictor. The univariate generation of the predicted values $(\mathbf{y}_{dr|s})_j$ for the non-sampled units j in small area d can be summarized as follows:

1. Estimate the unknown vector of model parameters $\theta = (\beta, \sigma_u^2, \sigma_e^2, \lambda_u, \lambda_e)$ using a suitable method and the sample data $\mathbf{y}_s$.

2. Predict the random effect u_d, say $\hat{u}_d$, using the sample data $\mathbf{y}_s$ with θ replaced by its estimator $\hat{\theta}$.

3. Generate independently values v_{dj}^{1} and v_{dj}^{0r} from $N(0,1)$.

4. The element j of the predictor $\mathbf{y}_{dr|s}$ is:

$$(\mathbf{y}_{dr|s})_{dj} = \mathbf{X}_{drj}\hat{\beta} + \hat{u}_d + \hat{\mu}_e + \hat{\sigma}_e(1-\hat{\delta}_e^2)^{1/2}v_{dj}^{1} + \hat{\sigma}_e\hat{\delta}_e|v_{dj}^{0r}| \tag{33}$$

To get $\hat{\eta}_d^{C-EQB}$ (respectively $\hat{\eta}_d^{C-EQBLUP}$), predict u_d using $\hat{u}_d^{EB}$ (respectively $\hat{u}_d^{EBLUP}$). Expression of $\hat{u}_d^{EB}$ is provided by Theorem 3.1 and $\hat{u}_d^{EBLUP}$ is given by (31).

3.3 ELL Method

The approach consists of drawing from the empirical area and unit level residuals to create a simulated census. The steps of the ELL method can be summarized as follows:

1. From the nested error model (4)-(5), calculate the total residuals $\hat{r}_{dj} = y_{dj} - \mathbf{x}_{dj}^T\hat{\beta}_{OLS}$ where $\hat{\beta}_{OLS}$ is the ordinary least square (OLS) estimate of β.
2. The effect of small area d, u_d, is estimated as the empirical mean value of the total residuals $\hat{r}_{dj}$ over all the observations from the small area d:

$$\hat{u}_d = \frac{1}{n_d}\sum_{j=1}^{n_d}\hat{r}_{dj} \tag{34}$$

3. The unit level residuals e_{dj} are estimated as:

$$\hat{e}_{dj} = \hat{r}_{dj} - \hat{u}_d. \tag{35}$$

   These residuals are then mean-corrected to sum to zero across the small area d.
4. Draw $\hat{\beta}^{(\ell)}$, $\hat{u}_d^{(\ell)}$, and $\hat{e}_{dj}^{(\ell)}$, $\ell = 1,...,L$ from respectively $N\left(\hat{\beta}_{OLS}, \mathbb{C}ov(\hat{\beta}_{OLS})\right)$, the empirical distributions of $\hat{u}_d$, and $\hat{e}_{dj}$.
5. Construct L predictors $y_{dj}^{(\ell)}$ as follows:

$$y_{dj}^{(\ell)} = \mathbf{x}_{dj}^T\hat{\beta}^{(\ell)} + \hat{u}_d^{(\ell)} + \hat{e}_{dj}^{(\ell)} \tag{36}$$

6. Get an estimate of the complex parameter as follows:

$$\hat{\eta}_d^{ELL-TRAD} = \frac{1}{L}\sum_{\ell=1}^{L}\hat{\eta}_d^{(\ell)} = \frac{1}{L}\sum_{\ell=1}^{L}h(\mathbf{y}_d^{(\ell)}) \tag{37}$$

We refer to the estimator (37) as the traditional ELL (ELL-TRAD) predictor. Unfortunately, this traditional ELL method does not provide a correct prediction of the area effects u_d across the L censuses since each draw l selects a different value $\hat{u}_d^{(\ell)}$ from the empirical area level residuals $\hat{u}_d$ with a mean approximately equals to 0. Across the L prediction cycles, the traditional ELL method does not attach a specific estimated random effect (area level residual) to the small area d. Instead, the traditional ELL method uses a combination of estimated random effects from other areas to predict the complex parameter for small area d. Empirical studies of the traditional ELL method by Molina and Rao (2010) have revealed very high MSEs compared to the best estimator and even higher than the MSE of the direct estimator under the normal model. In the simulation study in section 4, similar results as in Molina and Rao (2010) are observed. In order to reduce the MSE, we propose two nonparametric adjustments to the original ELL method. The traditional ELL approach has two main problems:

1. The area effects are wrongly assigned to the small areas across the L censuses. In fact, the random effects predictions are approximately equal to zero for linear parameters since the empirical average converges to zero ($\mathbb{E}(u_d) = 0$).
2. The variability of the empirical ELL is increased by drawing L different values from the empirical distributions of $N\left(\hat{\beta}_{OLS}, \mathbb{C}ov(\hat{\beta}_{OLS})\right)$ and the area effects residuals $\hat{u}_d^{(\ell)}$, $\ell = 1, ..., L$, for the same area, given a fixed sample.

Therefore, to address those two issues, we estimate the fixed effects and the random area effects using the sample data. Then, for the given sample, the intra-area distribution (conditional on the area effects) is estimated by drawing from the unit level residuals. The fixed effects are estimated using OLS as previously and two different nonparametric methods are used to obtain the area effect estimates.

The first method for estimating the area effects is the same as in the traditional ELL approach. For a given sample, we do not bootstrap the area level residuals. The steps of the algorithm are as follows:

1. From the nested error model (4)-(5), estimate the fixed effects β using OLS.
2. Estimate $\hat{u}_d$ and $\hat{e}_{dj}$ as in the traditional ELL method.
3. Draw $\hat{e}_{dj}^{(\ell)}$, $\ell = 1, ..., L$ from the empirical distribution of the $\hat{e}_{dj}$.
4. Construct L predictors $y_{dj}^{(\ell)}$ as follows:

$$y_{dj}^{(\ell)} = \mathbf{x}_{dj}^T \hat{\beta}_{OLS} + \hat{u}_d + \hat{e}_{dj}^{(\ell)}. \tag{38}$$

   Note that the same area level residual $\hat{u}_d$ is used for a given area d.
5. Obtain an estimate of the complex parameter as follows:

$$\hat{\eta}_d^{ELL-RES} = \frac{1}{L}\sum_{\ell=1}^{L} \hat{\eta}_d^{(\ell)} = \frac{1}{L}\sum_{\ell=1}^{L} h(\mathbf{y}_d^{(\ell)}) \tag{39}$$

The second method for estimating the area effects consists of using a combination of OLS and the method of moments to get estimators of σ_u^2 and σ_e^2 (Fuller and Battese (1973)). Then use the estimated variance components to get predictors of the area effects. Note that, as for the first method, no distributional assumption is needed. The steps of the algorithm are as follows:

1. Obtain an estimate of β using OLS.
2. Estimate σ_e^2 and σ_u^2 as follows:

$$\hat{\sigma}_e^2 = \frac{SSE(1)}{n - m - p_1}, \quad p_1 = \text{number of non-zero X-derivation} \tag{40}$$

   where $SSE(1)$ is the residual sum of squares obtain by regressing $(y_{dj} - \bar{y}_{d.})$ on the non-zero X-derivations $(x_{dj} - \bar{x}_{d.})$ for areas with $n_d > 1$ with $\bar{y}_{d.} = \sum_{j=1}^{n_d} y_{dj}/n_d$ and $\bar{\mathbf{x}}_{d.} = \sum_{j=1}^{n_d} \mathbf{x}_{dj}/n_d$. Also, we have

$$\hat{\sigma}_{um}^2 = \frac{SSE(2) - (n-p)\hat{\sigma}_e^2}{\sum_{j=1}^{n_d} n_d\{1 - n_d \bar{x}_{d.}\left(\sum_{d=1}^{m}\sum_{j=1}^{n_d} x_{dj}x_{dj}^T\right)^{-1}\bar{x}_{d.}^T\}}, \tag{41}$$

   where $SSE(2)$ is the residual sum of squares obtain by regressing y_{dj} on the non-zero x-derivations x_{dj}. Because the estimates can be negative, we truncate to get:

$$\hat{\sigma}_u^2 = max(\hat{\sigma}_{um}^2, 0). \tag{42}$$

3. Compute the estimated small area effect $\hat{u}_d$ as follows:

$$\hat{u}_d = \hat{\sigma}_u^2 \mathbf{1}_{n_d}^T \hat{\mathbf{V}}_d^{-1} \left(\mathbf{y}_d - \mathbf{X}_d \hat{\beta}_{OLS} \right) \tag{43}$$

where $\hat{\mathbf{V}}_d = \left(\hat{\sigma}_e^2 \mathbf{I}_{n_d} + \hat{\sigma}_u^2 \mathbf{1}_{n_d} \mathbf{1}_{n_d}^T \right)$.

4. Obtain unit level residuals as follows:

$$\tilde{e}_{dj} = y_{dj} - \mathbf{x}_{dj}^T \hat{\beta}_{OLS} - \hat{u}_d \tag{44}$$

Adjust these residuals to sum to zero and obtain $\hat{e}_{dj}$. Draw $\hat{e}_{dj}^{(\ell)}$ from the empirical distribution of the $\hat{e}_{dj}$.

5. Construct L predictors $\hat{y}_{dj}^{(\ell)}$ as follows:

$$\hat{y}_{dj}^{(\ell)} = \mathbf{x}^T \hat{\beta}_{OLS} + \hat{u}_d + \hat{e}_{dj}^{(\ell)}, \quad b = 1, ..., B. \tag{45}$$

Note that the same area level residual $\hat{u}_d$ is used for a given area d.

6. Obtain an estimate of the complex parameter as follows:

$$\hat{\eta}_d^{ELL-MOM} = \frac{1}{L} \sum_{\ell=1}^{L} \hat{\eta}_d^{(\ell)} = \frac{1}{L} \sum_{\ell=1}^{L} h(\mathbf{y}_d^{(\ell)}) \tag{46}$$

4. Simulation Results

The parameters of interest considered in this simulation study are the FGT poverty measures introduced by Foster et al. (1984). The FGT class is defined, for domain d, as:

$$F_{\alpha d} = \frac{1}{N_d} \sum_{j=1} N_d \left(\frac{z - E_{dj}}{z} \right)^{\alpha} I\left(E_j < z\right), \quad j = 1, ..., N_d, \alpha \geq 0, \tag{47}$$

where z is the poverty line, E_{dj} is a quantitative measure of welfare such as income or expenditure associated with individual j from domain d, and I is an indicator function. $I\left(E_j < z\right) = 1$ if $E_j < z$ meaning that the person j from area d is considered to be in poverty (welfare measure under poverty line) and similarly $I\left(E_j < z\right)$ is equal to 0 if $E_j \geq z$ (person j is not in poverty). The choice $\alpha = 0$ yields the proportion of people in poverty for domain d and F_{0d} is called the poverty incidence. The choice $\alpha = 1$, called the poverty gap, uses the normalized gap $\frac{z - E_{dj}}{z}$ to differentiate among the poor. The choice $\alpha = 2$, called poverty severity, squares the normalized gap.

A setup similar to Molina and Rao (2010) is used for these simulations: $m = 80$ small areas, $N_d = 250$, $n_d = 50$, $\beta = (3, 0.03, -0.04)^T$. Further, we set $\lambda_u = 1$, and $\lambda_e = 3$. The skewness parameter $\lambda_u = 1$ results in a minimal asymmetry and departure from the normal distribution with the same mean and variance. The skewness of the unit level distribution e_{dj} is moderate with $\lambda_e = 3$. The scale parameters σ_u^2 and σ_e^2 were chosen to ensure that $\mathbb{V}ar(u_d) = 0.15^2$ and $\mathbb{V}ar(e_{dj}) = 0.50^2$. There are two auxiliary variables $X_1 \in \{0, 1\}$ and $X_2 \in \{0, 1\}$ plus an intercept X_0. The values of the two dummies X_1 and X_2 are generated from Bernoulli distributions with

$$P(X_1 = 1) = 0.3 + \frac{0.5d}{m}, \quad P(X_2 = 1) = 0.2, \quad d = 1, \ldots, m = 80. \tag{48}$$

In each small area d, a sample s_d of 50 units is selected using simple random sampling. The total sample size is $n = 4,000$ selected from a total population size of $N = 20,000$. The Monte Carlo simulation consists of generating $I = 5,000$ populations, then for each

generated population the SAE methods described in the previous sections (empirical best, conditional empirical best, Molina-Rao (MR) normality-based, and ELL) are applied to obtain estimates of the complex parameter for the small areas. A simple direct estimator $\hat{F}_{\alpha d}^{srs} = \frac{1}{n_d}\sum_{j=1}^{n_d} F_{\alpha dj}, j = 1,...,n_d, \alpha \geq 0$ of $F_{\alpha d}$ is also included.

We fitted the nested error model with $y_{dj} = log(E_{dj})$ assuming normal distribution and used the estimate of β, $\mathbb{V}ar(u_d)$, and $\mathbb{V}ar(e_{dj})$ as the initial values of β, σ_u^2, and σ_e^2 respectively. The initial values of λ_u and λ_e were both chosen to be equal to 0.5. Empirical $MSE(\hat{\eta}_d)$ of an estimator $e\hat{t}a_d$ is calculated as

$$MSE(\hat{\eta}_d) = \frac{1}{5000}\sum_{i=1}^{5000}(\hat{\eta}_d^{(i)} - \eta_d^{(i)})^2 \tag{49}$$

where $\hat{\eta}_d^{(i)}$ is the estimate of the parameter $\eta_d^{(i)}$ for the i^{th} simulated population.

Figure 2 shows the bias of the MR normality-based estimator of the poverty gap under the three SN models: both u_d and e_{dj} follow SN distributions, only e_{dj} follows SN distribution, and only u_d follows SN distribution. When only u_d follows SN distribution, the MR normality-based estimator is essentially unbiased. Hence, misspecification of the distribution of u_d has no significant effect on the bias. On the other hand, bias of MR normality-based estimator is substantial for the other two cases. Similar result was found on the MSE.

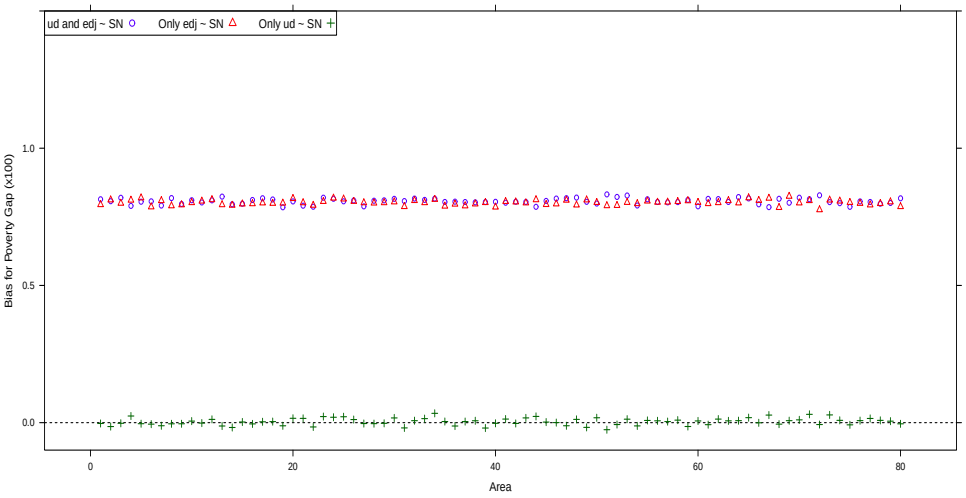

Figure 2: Bias of the MR normality-based estimator (MR-N) when at least one random error follows SN distribution (poverty gap, α=1).

Figure 3 shows a comparison of the empirical quasi-best (EQB) and the conditional empirical predictors (C-EQB and C-EBLUP). The two estimators EQB and C-EQB are equivalent in terms of MSE. As expected C-EQBLUP shows higher MSE than the quasi-best predictors EQB and C-EQB. It was also found that the estimator EQB is nearly unbiased while C-EQB and C-EQBLUP show a small bias.

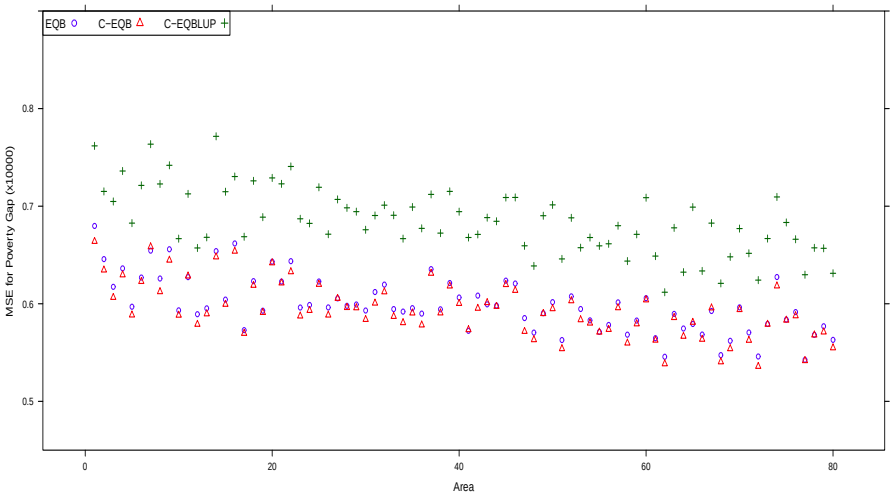

Figure 3: Comparison of the empirical quasi-best predictor (EQB) to the conditional predictors C-EQB and C-EQBLUP in terms of MSE (poverty gap, α=1).

Turning to ELL, the MSEs, as shown in Figure 4, are very high for the original ELL, $\hat{\eta}_d^{ELL-TRAD}$, compared to the two proposed alternative methods. The results show a very high gain for both alternative methods with $\hat{\eta}_d^{ELL-MOM}$ achieving nearly 80% improvement over the original ELL.

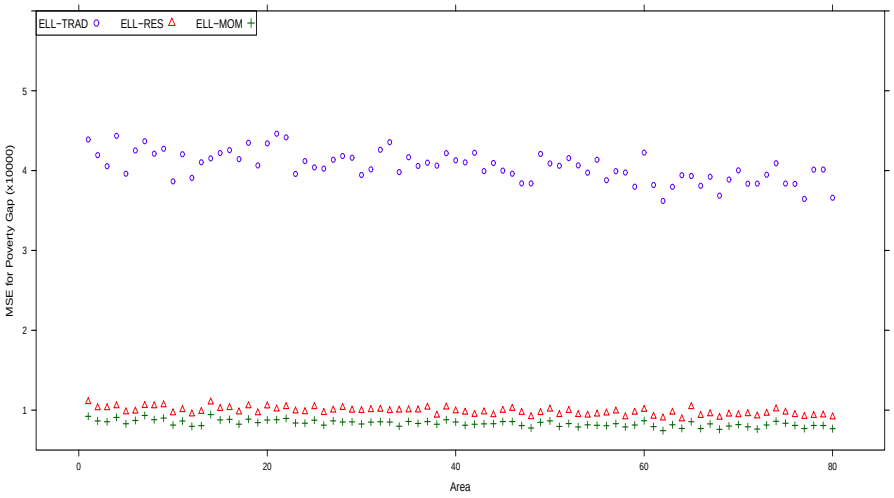

Figure 4: MSE of the three different ELL methods when both random errors follow SN distribution (poverty gap, α=1).

For each of the four estimators, EQB, C-EQB, ELL-MOM, and MR-N, the ratio of its MSE over the MSE of the simple direct estimator is computed. Values of the ratio under 1 show gain over the direct estimator while values over 1 indicate that the direct estimator is better in term of MSE. Results on poverty incidence ($\alpha = 0$), from Figure 5, show that all four estimators do better than the direct estimator. Among the four estimators, the marginal and conditional quasi-best estimators are equivalently the best and the modified ELL and the MR normality-based estimators are equivalent.

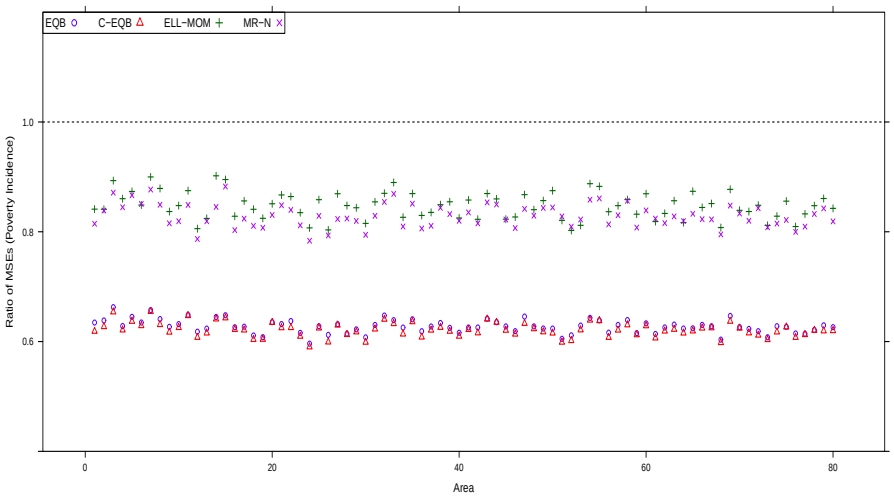

Figure 5: Ratio of the MSEs of EQB, C-EQB, ELL-MOM, and MR-N to the MSE of the simple direct estimator (poverty incidence, α=0).

Results on the poverty gap ($\alpha = 1$), from Figure 6, show that only the two quasi-best estimators (marginal and conditional) do better than the direct estimator with improvement about 25% for the lowest area indicators to about 33% for highest area indicators. The improved ELL method using method of moments has a performance similar to the direct estimator. The MR normality-based estimator is much worse than the other three predictors and the director estimator. The performance of the MR normality-based estimator is worse for the poverty gap ($\alpha = 1$) than for the previous less complex poverty incidence ($\alpha = 0$). Relative to the direct estimator, the other three predictors are less affected by the extra complexity of the parameter of interest than the MR normality-based estimator. Results on poverty severity ($\alpha = 2$) were similar to those for poverty gap. A value of λ_e smaller than 3 would produce better result for the MR normality-based. The choice of $\lambda_e = 3$ may be extreme under the log transformation of the welfare variable.

The traditional ELL predictor had much larger MSEs then any of the four estimators in the last two figures. It was not included in Figures 5 and 6 to avoid squeezing the graphs and make them unreadable. For poverty incidence, small area MSEs of the traditional ELL estimator were on average about 5.2 times larger than the MSEs of the MR normal-based estimator. That ratio reduces to about 3 for poverty gap and 1.4 for poverty severity. For all three poverty measures, the traditional ELL method performs worse than the direct estimator.

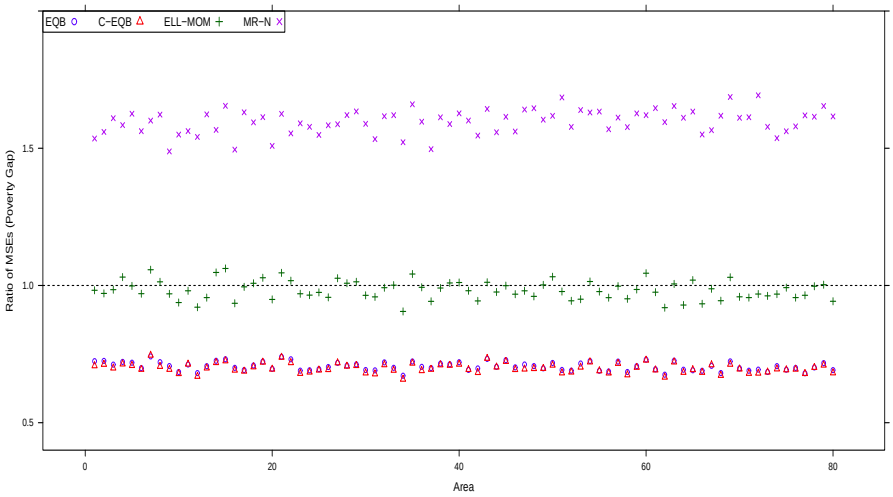

Figure 6: Ratio of the MSEs of EQB, C-EQB, ELL-MOM, and MR-N to the MSE of the simple direct estimator (poverty gap, α=1).

References

Arellano-Valle, R. B. and Azzalini, A. (2006). On the unification of the families of skew-normal distributions. *Scandinavian Journal of Statistics*, 33:516–574.

Battese, G. E., Harter, R. M., and 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*, 83:28–36.

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

Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52:761–766.

Fuller, W. A. and Battese, G. (1973). Transformations for estimation of linear models with nested-error structure. *Journal of the American Statistical Association*, 68:626–632.

Genton, M. G. (2004). *Skew-Elliptical Distributions And Their Applications, A Journey Beyond Normality*. Chapman and Hall/CRC, New York.

Molina, I. and Rao, J. N. K. (2010). Small area estimation of poverty indicators. *The Canadian Journal of Statistics*, 38:369–385.

Rao, J. N. K. (2003). *Small Area Estimation*. John Wiley and Sons, Hoboken, New Jersey.