Estimation of Poverty and Social Exclusion Indicators at the Small Area Level

Hukum Chandra
Indian Agricultural Statistics Research Institute, New Delhi
Email: hchandra@iasri.res.in

Acknowledgments

Ray Chambers, University of Wollongong, Australia
Nicola Salvati, University of Pisa, Italy

Outline

- Background
- Small Area Estimation (SAE) methods
- Poverty Indicators
- SAE methods of poverty estimation
- SAE methods for Unit-level data
  - World Bank Method **(Elbers, Lanjouw and Lanjouw, 2003)**
  - Empirical Best Prediction Method **(Molina and Rao, 2010)**
  - M-Quantile Method **(Tzavidis *et al.,* 2008)**
  - GLMM
- SAE methods for Aggregated (or Area) level data
  - GLMM
  - Spatial models
- Application to real data –poverty mapping

Background

- Most national surveys are planned to give **reliable** estimates at national and large domain levels and are **not** appropriate to produce small domain level estimates due to **small sample sizes**

- **Direct estimation:** estimates derived using only the domain (or area) specific sample data are known as direct estimates

- An area is regarded as **small** if the area specific sample size is not large enough to support direct estimates of adequate precision

- Small sample size $\Rightarrow$ large standard errors for estimates

- If **no** sample data at all $\Rightarrow$ impossible to construct direct estimator

"Direct estimation is typically not an option"

Background

- A need for special techniques to produce estimates for such small domains or small areas referred as small area estimation (SAE)

"SAE answers the problem of small sample sizes"

- **The SAE methods** look at producing estimates with **adequate** precision through an estimation procedure that ***'borrows strength'*** and thus increase the overall (**effective**) sample size and precision

- They **borrow information** (data) from other small areas or times (or both) by use of statistical models either based on **implicit** or **explicit** models that link related small areas through auxiliary information

- **Unit** level **linear mixed models** with area specific random effects are quite popular in small area estimation **(Rao, 2003)**

Background

- In many SAE problems, it is not always possible to use the unit level small area model simply because of the unavailability of the unit level data

- In such circumstances, SAE is carried out under **area level random effect models**

- The Fay–Herriot model is widely used area level model in SAE **(Fay and Herriot, 1979)** - one of the most popular methods of SAE

Background

Based on the level of auxiliary information available and utilised, two types of mixed models are used for SAE:

- **Unit level model -** when the unit level auxiliary information is available. They relate the unit values of a study variable to unit-specific covariates **(Battese, Harter and Fuller, 1988)**

- **Area level model -** when auxiliary information is available only at area level. They relate small area direct estimates to area-specific covariates **(Fay and Herriot, 1979)**

- These are special cases of the mixed model, usually referred as area level and unit level small area models

- The empirical best linear unbiased predictor **(EBLUP)** is widely used technique of SAE under these models and proven to be efficient, see **Rao (2003)**

Background

- **United Nation Millennium Development Goal-1 (UNMDG-1):** To eradicate extreme poverty and hunger

- To monitor the progress and target aiming at the reduction of poverty requires micro/regional (or small area) level estimates on various parameters related to the living conditions of people

- From existing large scale surveys, sample sizes in small areas tend to be too **small** to provide small area specific reliable direct estimates for poverty parameters

- Therefore, we need to apply small area estimation techniques

Problem

- Commonly used small area estimation methods are typically developed for the estimation of linear characteristics, such as means/totals

- Many measures of poverty and inequality are nonlinear functions of a quantitative welfare variable for the population units. As a results, the standard methods are not suitable estimation of poverty parameters

- Potential approaches for SAE of poverty parameters
  - World Bank Method **(Elbers, Lanjouw and Lanjouw, 2003)**
  - Empirical Best Prediction Method **(Molina and Rao, 2010)**
  - M-Quantile Method **(Tzavidis *et al.*, 2008)**
  - GLMM - Empirical Prediction **(Saei and Chambers, 2003)**

Notation

- U finite population of size N
- Population U partitioned into D non-overlapping subsets $U_1,....,U_D$ of *sizes* $N_1,....., N_D$ called **domains**, small areas, or simply **areas**
- $U = \bigcup_{d=1}^{D} U_d$; $N = \sum_{d=1}^{D} N_d$ and N_d population size of domain d
- Variable of interest Y
- Y_{dj} value of Y for unit *j* from domain d
- Auxiliary variable X
- x_{dj} - $p \times 1$ vector of values of X for unit *j* from domain d

Notation

- We want to use data from a sample $s \subset U$ of size n drawn from the whole population

- Non-sample component of U, containing $N - n$ units, denoted by r

- $s_d = s \cap U_d$ sub-sample from domain d of size $n_d = | s_d |$ such that $n = \sum_{d=1}^{D} n_d$ and $s = \bigcup_{d=1}^{D} s_d$

- r_d non-sample units of domain d of $N_d - n_d = | r_d |$
- $n_d = 0$ if area is not sampled

- If $\pi_j = \sum_{j \in s} p(s)$ are the first order inclusion probabilities then $w_j = \pi_j^{-1}$ defines the design weight of element j

- **Problem**: Domain specific sample size n_d too small for some domains

Direct Estimation

- Direct estimator: Estimator that uses only the sample data from the corresponding domain

- **The estimate of population mean** of Y in the small area d, $\bar{Y}_d = N_d^{-1} \sum_{j \in U_d} y_d$ is

$$\hat{\bar{Y}}_d^{Hajek} = \left(\sum_{j \in s_d} w_j \right)^{-1} \left(\sum_{j \in s_d} w_j y_j \right)$$

or, if the population size N_i of the small area i is known

$$\hat{\bar{Y}}_d^{HT} = N_d^{-1} \left(\sum_{j \in s_d} w_j y_j \right)$$

- These are the **direct estimators** of small area d mean $\bar{Y}_d$

For example, under simple random sampling

- **With no auxiliary information**: $\pi_j = nN^{-1}$ and $w_j = Nn^{-1} \Rightarrow$
- **A direct estimator** of the mean of Y for area d is

$$\hat{\bar{Y}}_d = \begin{cases} \bar{y}_d = \sum_{s_d} w_j y_j \Big/ \sum_{s_d} w_j = n_d^{-1} \sum_{j \in s_d} y_j; & if \quad n_d \geq 1 \\ 0 & otherwise \end{cases}$$

- An unbiased estimator for **variance** is

$$v(\hat{\bar{Y}}_d \mid n_d) = (1 - f_d) s_d^2 / n_d \qquad \text{when } N_d \text{ is known.}$$

$$v(\hat{\bar{Y}}_d \mid n_d) = (1 - f) s_d^2 / n_d \qquad \text{when } N_d \text{ unknown}$$

where $s_d^2 = (n_d - 1)^{-1} \sum_{j \in s_d} (y_j - \bar{y}_d)^2$

- It is obvious that for **small sample size** n_d, the variance will be **larger** unless the variability of the Y values is sufficiently small

Nested Error Unit Level Regression model
(Battese, Harter and Fuller, 1988)

$$y_{dj} = \mathrm{x}_{dj}^T \boldsymbol{\beta} + u_d + e_{dj}; (d = 1, \ldots, D;\ j = 1, \ldots, N_d)$$

- u_d and e_{dj} are **independent** with zero means and variances σ_u^2 and σ_e^2 respectively

- u_d's represent the combined effect of area characteristics, not accounted for by the auxiliary variables

- It is assumed that the area means $\bar{\mathrm{X}}_d = N_d^{-1} \sum_{i=1}^{N_d} \mathrm{x}_{di}$ are **known**

- The population mean of the ***y***-values in area *d*

$$\bar{Y}_d = N_d^{-1} \sum_{i=1}^{N_d} y_{dj} = \bar{\mathrm{X}}_d^T \boldsymbol{\beta} + u_d + \bar{e}_d$$

- For sufficiently large N_d , $\bar{e}_d = N_d^{-1} \sum_{i=1}^{N_d} e_i \approx 0 \left[V(\bar{e}_d) \approx \sigma_e^2 / N_d \right]$,

- Then the mean of Y in small area d is approximated by

$$\mu_d = \bar{\mathrm{X}}_d^T \boldsymbol{\beta} + u_d = E(\bar{Y}_d \mid \bar{\mathrm{X}}_d, u_d)$$

- For **known variances,** $\theta = (\sigma_u^2, \sigma_e^2)$, following the **Henderson (1975)**, the Best Linear Unbiased Predictor (**BLUP**) of μ_d is,

$$\hat{\mu}_d = \bar{\mathrm{X}}_d^T \tilde{\boldsymbol{\beta}}_{GLS} + \gamma_d (\bar{y}_d - \bar{\mathrm{x}}_d^T \tilde{\boldsymbol{\beta}}_{GLS})$$

$$= \gamma_d \left\{ \bar{y}_d + (\bar{\mathrm{X}}_d - \bar{\mathrm{x}}_d)^T \tilde{\boldsymbol{\beta}}_{GLS} \right\} + (1 - \gamma_d) \bar{\mathrm{X}}_d^T \tilde{\boldsymbol{\beta}}_{GLS}$$

- $\bar{y}_d$ and $\bar{\mathrm{x}}_d$ are the sample means of y and x for area d, and $\gamma_d = \sigma_u^2 \left(\sigma_u^2 + n_d^{-1} \sigma_e^2 \right)^{-1}$

- The weight, $0 \le \gamma_d \le 1$ called '**shrinkage factor**', provides a trade off between the approximately design-unbiased regression estimator and the synthetic estimator and measures the model variance σ_u^2 relative to total variance $\left(\sigma_u^2 + n_d^{-1} \sigma_e^2 \right)$

- For a small value of σ_u^2, weight γ_d will be small and consequently the synthetic part get more weight and vice versa

- For $n_d = 0$, i.e. areas with **no** samples, $\gamma_d \to 0$ and $\hat{\mu}_i = \bar{\mathrm{X}}_d^T \tilde{\boldsymbol{\beta}}_{GLS}$

- When **sampling fraction**, $f_d = n_d / N_d$ is **non-negligible**, then the estimated parameter is, $\bar{Y}_d = \bar{\mathrm{X}}_d^T \boldsymbol{\beta} + u_d + \bar{e}_d$, the **BLUP** for $\bar{Y}_d$ is

$$\hat{\bar{Y}}_d^{BLUP} = N_d^{-1}\left[\sum_{i \in s_d} y_i + \sum_{i \in r_d} (\bar{\mathrm{X}}_d^T \tilde{\boldsymbol{\beta}}_{GLS} + \hat{u}_d)\right]$$

$$= f_d \bar{y}_d + (1 - f_d)\left\{\bar{\mathrm{X}}_{dr}^T \tilde{\boldsymbol{\beta}}_{GLS} + \gamma_d (\bar{y}_d - \bar{\mathrm{x}}_d^T \tilde{\boldsymbol{\beta}}_{GLS})\right\}$$

where $\bar{\mathrm{X}}_{dr} = (N_d \bar{\mathrm{X}}_d - n_d \bar{\mathrm{x}}_d)/(N_d - n_d)$, is the mean of x for $(N_d - n_d)$ **non-sampled** units for small area d

Variance of BLUP (known variances)

$$Var(\hat{\mu}_d) = E(\hat{\mu}_d - \mu_d)^2 = g_{1d}(\theta) + g_{2d}(\theta)$$

$$g_{1d}(\theta) = (1-\gamma_d)\sigma_u^2 = \gamma_d(\sigma_e^2 / n_d)$$

$$g_{2d}(\theta) = c_d^T \left(\sum_d \mathrm{X}_d^T \mathrm{V}_d^{-1} \mathrm{X}_d \right)^{-1} c_d ; c_d^T = (\bar{X}_d^T - \gamma_d \bar{x}_d^T)$$

- Here $g_{1d}(\theta)$ is the **leading** term whereas in MSE of the simple regression estimator leading term is (σ_e^2 / n_d)

- This shows that the **BLUP** is **superior** to the simple regression estimator in terms of MSE if the shrinkage factor γ_d is small

The case of unknown variances

In practice $\theta = (\sigma_u^2, \sigma_e^2)$ are **unknown** and estimated from sample data using standard method of estimation such as **ANOVA**, **ML/REML** methods of estimation **(Rao, 2003)**

The **EBLUP** estimators are obtained by replacing the unknown variances θ by their sample estimates $\hat{\theta} = (\hat{\sigma}_u^2, \hat{\sigma}_e^2)$

$$\hat{\mu}_d = \bar{\mathrm{X}}_d^T \hat{\boldsymbol{\beta}}_{GLS} + \hat{\gamma}_d (\bar{y}_d - \bar{\mathrm{x}}_d^T \hat{\boldsymbol{\beta}}_{GLS})$$

or

$$\hat{\bar{Y}}_d^{EBLUP} = f_d \bar{y}_d + (1 - f_d) \left\{ \bar{\mathrm{X}}_{dr}^T \hat{\boldsymbol{\beta}} + \hat{\gamma}_d (\bar{y}_d - \bar{\mathrm{x}}_d^T \hat{\boldsymbol{\beta}}) \right\}$$

Estimation of MSE of EBLUP

- The **EBLUP** is obtained by replacing the **unknown** variances by their estimates, everywhere they appear in the expressions for the **BLUP**

- A **naive MSE** estimator is obtained by applying similar substitutions in the expression for the variance of the **BLUP**
  - **ignores** the error resulting from variance estimation
  - underestimates the true MSE

Prasad and Rao (1990) approximate $MSE[\hat{\mu}_d(\hat{\theta})]$ (**true** MSE) to order **o(1/*D*)**, assuming **normality** of the error terms as,

$$MSE[\hat{\mu}_d(\hat{\theta})] = g_{1d}(\theta) + g_{2d}(\theta) + g_{3d}(\theta)$$

where

$$g_{3d}(\theta) = n_d^{-1}(\sigma_u^2 + \sigma_e^2 n_d^{-1})^{-3} Var(\sigma_e^2 \hat{\sigma}_u^2 - \sigma_u^2 \hat{\sigma}_e^2)$$

For the expression of $Var(\hat{\sigma}_e^2)$; $Var(\hat{\sigma}_u^2)$ and $Cov(\hat{\sigma}_u^2, \hat{\sigma}_e^2)$ under normality of the error terms, see **Prasad and Rao (1990)**

An estimator of $MSE[\hat{\mu}_d(\hat{\theta})]$with bias **o(1/D)** is,

$$mse[\hat{\mu}_d(\hat{\theta})] = g_{1d}(\hat{\theta}) + g_{2d}(\hat{\theta}) + 2g_{3d}(\hat{\theta})$$

with $g_{3d}(\hat{\theta}) = n_d^{-1}(\hat{\sigma}_u^2 + \hat{\sigma}_e^2 n_d^{-1})^{-3} \hat{Var}(\sigma_e^2 \hat{\sigma}_u^2 - \sigma_u^2 \hat{\sigma}_e^2)$

An estimator of $MSE[\hat{\bar{Y}}_d(\hat{\theta})]$ is

$$mse[\hat{\bar{Y}}_d(\hat{\theta})] = (1 - f_d)^2 mse(\hat{\mu}_d^*) + N_d^{-1}(1 - f_d)\hat{\sigma}_e^2$$

$mse(\hat{\mu}_d^*) = mse[\hat{\mu}_d^*(\hat{\theta})]$ is obtained from $MSE[\hat{\mu}_d(\hat{\theta})]$ replacing $g_{2d}(\hat{\theta})$ by $g_{2d}^*(\hat{\theta})$. Here $g_{2d}^*(\hat{\theta})$ is obtained from $g_{2d}(\hat{\theta})$ by replacing $\bar{X}_d$ by $\bar{X}_{rd}$

Area level model for small area estimation
(Fay and Herriot, 1979)

- Used when individual measurements for auxiliary variables are **not** available or when the auxiliary information is only at the **area** level

- The simple area specific two stage model suggested by **Fay and Herriot (1979)** is described as

$$y_d = Y_d + e_d, \text{ and } Y_d = \mathrm{x}_d^T \boldsymbol{\beta} + u_d (d = 1, \ldots, D)$$

- The **first stage** accounts for the **sampling variability** of the survey estimates y_d of true area means Y_d and the **second stage** links the true area means Y_d to a vector of known auxiliary variables x_d

- ***D*** - number of small areas with sample data

- We can express as an area level **linear mixed model**

$$y_d = \mathrm{x}_d^T \boldsymbol{\beta} + u_d + e_d$$

- $\boldsymbol{\beta}$ is a p-vector of unknown fixed effect parameters

- u_d's are iid normal random errors with $E(u_d) = 0$ and $Var(u_d) = \sigma_u^2$

- e_d's are independent **sampling errors** of the **direct estimator** normally distributed with $E(e_d \mid Y_d) = 0,\ Var(e_d \mid Y_{di}) = \sigma_{ed}^2,\ u_d \perp e_d$

- Assuming **known** variances σ_{ed}^2 and σ_u^2, the **BLUP** estimate for Y_i is **(Henderson, 1975; Fay & Herriot, 1979)**

$$\hat{Y}_d^{BLUP} = \mathrm{x}_d^T \tilde{\boldsymbol{\beta}}_{GLS} + \hat{u}_d = \mathrm{x}_d^T \tilde{\boldsymbol{\beta}}_{GLS} + \gamma_d (y_d - \mathrm{x}_d^T \tilde{\boldsymbol{\beta}}_{GLS})$$

$$= \gamma_d y_d + (1 - \gamma_d) \mathrm{x}_d^T \hat{\boldsymbol{\beta}}_{GLS}$$

$$\tilde{\boldsymbol{\beta}}_{GLS} = \left(\sum_{d=1}^{D} V_d^{-1} \mathrm{x}_d \mathrm{x}_d^T \right)^{-1} \left(\sum_{d=1}^{D} V_d^{-1} \mathrm{x}_d y_d \right); \quad V_d = \sigma_{ed}^2 + \sigma_u^2$$

$$\gamma_d = \sigma_u^2 (\sigma_{ed}^2 + \sigma_u^2)^{-1}$$

- When $n_d = 0$, $\hat{Y}_d^{SYN-EBLUP} = \mathrm{x}_d^T \hat{\boldsymbol{\beta}}_{GLS}$, **synthetic estimator**

- **No design-based solution!!**

- $\hat{Y}_d^{BLUP}$ has form of a **composite** estimator with the weight γ_d

- γ_d is called "**shrinkage factor**" since it 'shrinks' the direct estimator towards the **synthetic estimator,** $\mathrm{x}_d^T\tilde{\boldsymbol{\beta}}_{GLS}$

Variance of (Prediction error) of BLUP

$$Var(\hat{Y}_d^{BLUP} - Y_d) = E(\hat{Y}_d^{BLUP} - Y_d)^2 \left[\text{since } E(\hat{Y}_d^{BLUP} - Y_d) = 0\right]$$

$$= \gamma_d \sigma_{ed}^2 + (1-\gamma_d)^2 \mathrm{x}_d^T Var(\tilde{\boldsymbol{\beta}}_{GLS}) \mathrm{x}_d = g_{1d} + g_{2d}$$

- For large D, $Var(\tilde{\boldsymbol{\beta}}_{GLS})$ is small and $Var(\hat{Y}_d^{BLUP}) \approx g_{1d} = \gamma_d \sigma_{ed}^2$
- $g_{1d} \approx O(1)$ and $g_{2d} \approx O(1/D)$

- $g_{1d} \leq \sigma_{ed}^2$ and is much smaller if σ_u^2 is small ('good' covariate information). This variance relationship shows the **superiority** of the use of the model over the use of the direct estimator

The case of unknown variance σ_u^2

- The variance σ_u^2 can be estimated from the sample data by **MLE/REML**, or by use of **Methods of Moments**

- Substituting $\hat{\sigma}_u^2$ for σ_u^2 yields the **Empirical-BLUP(EBLUP)**

$$\hat{Y}_d^{EBLUP} = \mathrm{x}_d^T \hat{\boldsymbol{\beta}}_{GLS} + \hat{\gamma}_d (y_d - \mathrm{x}_d^T \hat{\boldsymbol{\beta}}_{GLS})$$

- $\hat{\boldsymbol{\beta}}_{GLS} = \left(\sum_{d=1}^{D} \hat{V}_d^{-1} \mathrm{x}_d \mathrm{x}_d^T \right)^{-1} \left(\sum_{d=1}^{D} \hat{V}_d^{-1} \mathrm{x}_d y_d \right); \ \hat{V}_d = \sigma_{ed}^2 + \hat{\sigma}_u^2$
- $\hat{\gamma}_d = \hat{\sigma}_u^2 (\sigma_{ed}^2 + \hat{\sigma}_u^2)^{-1}$
- When $\sigma_u^2 = 0 \Rightarrow$ EBLUP=$\mathrm{x}_d^T \hat{\boldsymbol{\beta}}_{GLS} \rightarrow$ **synthetic estimator**

Estimation of MSE of EBLUP

- The **EBLUP** is obtained by replacing the **unknown** variances by their estimates, everywhere they appear in the expressions for the **BLUP**

- A **naive MSE** estimator is obtained by applying similar substitutions in the expression for the variance of the **BLUP**
  - **ignores** the error resulting from variance estimation
  - underestimates the true MSE

- **Prasad and Rao (1990)** approximate $MSE\left[\hat{Y}_d^{EBLUP}\right]$ to order **o(1/*D*)**, assuming **normality** of the error terms as,

$$MSE\left[\hat{Y}_d^{EBLUP}\right] = g_{1d}(\sigma_u^2) + g_{2d}(\sigma_u^2) + g_{3d}(\sigma_u^2)$$

- $g_{2d}(\sigma_u^2)$ is the excess in MSE due to estimation of β and

- $g_{3d}(\sigma_u^2) = [\sigma_{ed}^4 / (\sigma_{ed}^2 + \sigma_u^2)^3] \times Var(\hat{\sigma}_u^2)$ is the **excess in MSE** due to estimation of σ_u^2

- An estimator of $MSE\left[\hat{Y}_d^{EBLUP}\right]$ with bias of order **o(1/*D*)** is then

$$mse\left[\hat{Y}_d^{EBLUP}\right] = g_{1d}(\hat{\sigma}_u^2) + g_{2d}(\hat{\sigma}_u^2) + 2\,g_{3d}(\hat{\sigma}_u^2)$$

- When estimating $\hat{\sigma}_u^2 = \hat{\sigma}_{FC}^2$= fitting of constants

$$Var(\hat{\sigma}_{FC}^2) \approx 2\,D^{-2} \sum_{d=1}^{D} \left(\sigma_{ed}^2 + \sigma_u^2\right)^2$$

- **Datta and Lahiri (2000)** noticed that when ML estimation is used for parameter estimation, there is an extra bias correction term in MSE estimate of $\hat{Y}_d^{EBLUP}$

Software for Small Area Estimation

- **SAS** – provide the facility of generating EBLUP and its MSE estimates under both area and unit level small area model, see **Mukhopadhyay. and McDowell (2011)**

- **R** – R codes are easily available for both area and unit level model, and associated MSE estimates **(SAE2 package in R),** http://www.bias-project.org.uk/SAE_tutorial/

- For fitting area level model using R. See **Chandra et al. (2013)** http://sample.iasri.res.in/ssrs/SAE_Using_R.pdf