Estimation of Complex Small Area Parameters with Application to Poverty Indicators

J.N.K. Rao

School of Mathematics and Statistics, Carleton University

(Joint work with Isabel Molina)

NOTATION

- U **finite** population of size N.
- Population partitioned into D subsets $U_1, \ldots, U_D$ of sizes $N_1, \ldots, N_D$, called **domains** or **areas**.
- Variable of interest Y.
- Y_{dj} value of Y for unit j from domain d.
- **Target:** to estimate domain parameters.

$$\delta_d = h(Y_{d1}, \ldots, Y_{dN_d}), \quad d = 1, \ldots, D.$$

- We want to use data from a sample $S \subset U$ of size n drawn from the whole population.
- $S_d = S \cap U_d$ sub-sample from domain d of size n_d.
- **Problem:** n_d **small** for some domains.

DIRECT ESTIMATORS

- **Direct estimator:** Estimator that uses only the sample data from the corresponding domain.
- **Small area/domain:** subset of the population that is target of inference and for which the direct estimator does not have enough precision.
- What does "enough precision" mean? Some National Statistical Offices (GB, Spain) allow a maximum coefficient of variation of 20 %.
- **Indirect estimator:** Borrows strength from other areas.

NESTED-ERROR REGRESSION MODEL

- **Model:** $\mathbf{x}_{dj}$ auxiliary variables at unit level,

$$Y_{dj} = \mathbf{x}_{dj}'\boldsymbol{\beta} + u_d + e_{dj}, \quad u_d \overset{iid}{\sim} N(0, \sigma_u^2), \quad e_{dj} \overset{iid}{\sim} N(0, \sigma_e^2).$$

- **Vector of variance components:**

$$\boldsymbol{\theta} = (\sigma_u^2, \sigma_e^2)'$$

- **BLUP of** $\bar{Y}_d$: Predict non-sample values $\hat{Y}_{dj} = \mathbf{x}_{dj}'\hat{\boldsymbol{\beta}}_{WLS} + \hat{u}_d$.

$$\hat{\bar{Y}}_d^{BLUP} = \frac{1}{N_d}\left(\sum_{j \in s_d} Y_{dj} + \sum_{j \in r_d} \hat{Y}_{dj}\right), \quad d = 1, \ldots, D.$$

- **Empirical BLUP (EBLUP):** $\hat{\boldsymbol{\theta}}$ estimator of $\boldsymbol{\theta}$

$$\hat{\bar{Y}}_d^{EBLUP} = \hat{\bar{Y}}_d^{BLUP}(\hat{\boldsymbol{\theta}})$$

✓ *Battese, Harter & Fuller (1988), JASA*

SOME POVERTY AND INCOME INEQUALITY MEASURES

- FGT poverty indicator
- Gini coefficient
- Sen index
- Theil index
- Generalized entropy
- Fuzzy monetary index

FGT POVERTY INDICATORS

- E_{dj} welfare measure for indiv. j in domain d: for instance, equivalised annual net income.
- $z =$ poverty line.
- **FGT family of poverty indicators for domain d:**

$$F_{\alpha d} = \frac{1}{N_d} \sum_{j=1}^{N_d} \left(\frac{z - E_{dj}}{z} \right)^{\alpha} I(E_{dj} < z), \quad \alpha = 0, 1, 2.$$

When $\alpha = 0 \Rightarrow$ **Poverty incidence**

When $\alpha = 1 \Rightarrow$ **Poverty gap**

When $\alpha = 2 \Rightarrow$ Poverty severity

✓ *Foster, Greer & Thornbecke (1984), Econometrica*

FGT POVERTY INDICATORS

- **Complex non-linear** quantities (non continuous): Even if FGT poverty indicators are also means

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

  we cannot assume normality for the $F_{\alpha dj}$.
- Not easy to obtain small area estimators with good bias and MSE properties.
- A method valid to estimate poverty measures in small areas for any α and for other poverty or inequality measures would be desirable.

SMALL AREA ESTIMATION

- Due to the relative nature of the mentioned poverty line, poverty has usually **low frequency**: Large sample size is needed.

  ✓ In Spain, poverty line for 2006: **6557 euros**, approx. **20 %** population under the line.

- Survey on Income and Living Conditions (EU-SILC) has limited sample size.

  ✓ In the Spanish SILC 2006, $n = 34{,}389$ out of $N = 43{,}162{,}384$ **(8 out 10,000)**.

SAMPLE SIZES OF PROVINCES BY GENDER

- Direct estimators for Spanish provinces are not very precise.
- Provinces $\times$ Gender $\rightarrow$ Small areas (52×2).
- CVs of direct and EB estimators of poverty incidences for 5 selected provinces:

Province	Gender	n_d	Obs. Poor	CV Dir.	CV EB
Soria	F	17	6	40.37	16.52
Tarragona	M	129	18	19.85	16.15
Córdoba	F	230	73	7.52	6.73
Badajoz	M	472	175	7.12	3.57
Barcelona	F	1483	191	6.67	5.37

EB METHOD (EMPIRICAL BEST/BAYES)

- Vector with population elements for domain d:

$$\mathbf{y}_d = (Y_{d1}, \ldots, Y_{dN_d})' = (\mathbf{y}'_{ds}, \mathbf{y}'_{dr})'$$

- **Target parameter:**

$$\delta_d = h(\mathbf{y}_d)$$

- **Best estimator:** The estimator $\hat{\delta}_d$ that minimizes the MSE is

$$\hat{\delta}_d^B = E_{\mathbf{y}_{dr}}(\delta_d | \mathbf{y}_{ds}).$$

- **Best estimator of $F_{\alpha d}$:** We need to express $\delta_d = F_{\alpha d}$ in terms of a vector $\mathbf{y}_d = (\mathbf{y}'_{ds}, \mathbf{y}'_{dr})'$,

$$F_{\alpha d} = h_\alpha(\mathbf{y}_d)$$

  for which we can derive the distribution of $\mathbf{y}_{dr} | \mathbf{y}_{ds}$.

EB METHOD FOR POVERTY ESTIMATION

- **Assumption:** there exists a transformation $Y_{dj} = T(E_{dj})$ of the welfare variables E_{dj} which follows a normal distribution (i.e., the nested error model with normal errors u_d and e_{dj}).
- FGT poverty indicator as a function of transformed variables:

$$F_{\alpha d} = \frac{1}{N_d} \sum_{j=1}^{N_d} \left\{ \frac{z - T^{-1}(Y_{dj})}{z} \right\}^{\alpha} I\left\{ T^{-1}(Y_{dj}) < z \right\}.$$

- **EB estimator of $F_{\alpha d}$:**

$$\hat{F}_{\alpha d}^{EB} = E_{\mathbf{y}_{dr}}\left[F_{\alpha d} | \mathbf{y}_{ds} \right], \quad F_{\alpha d} = h_{\alpha}(\mathbf{y}_d).$$

EB METHOD FOR POVERTY ESTIMATION

- Distribution: $\mathbf{y}_d \overset{ind}{\sim} N(\boldsymbol{\mu}_d, \mathbf{V}_d)$, $d = 1\ldots, D$, where

$$\mathbf{y}_d = \begin{pmatrix} \mathbf{y}_{ds} \\ \mathbf{y}_{dr} \end{pmatrix}, \quad \boldsymbol{\mu}_d = \begin{pmatrix} \boldsymbol{\mu}_{ds} \\ \boldsymbol{\mu}_{dr} \end{pmatrix}, \quad \mathbf{V}_d = \begin{pmatrix} \mathbf{V}_{ds} & \mathbf{V}_{dsr} \\ \mathbf{V}_{dsr} & \mathbf{V}_{dr} \end{pmatrix}.$$

- Distribution of $\mathbf{y}_{dr}$ given $\mathbf{y}_{ds}$:

$$\mathbf{y}_{dr}|\mathbf{y}_{ds} \sim N(\boldsymbol{\mu}_{dr|ds}, \mathbf{V}_{dr|ds}),$$

where

$$\begin{aligned} \boldsymbol{\mu}_{dr|ds} &= \boldsymbol{\mu}_{dr} + \mathbf{V}_{drs}\mathbf{V}_{ds}^{-1}(\mathbf{y}_{ds} - \boldsymbol{\mu}_{ds}), \\ \mathbf{V}_{dr|ds} &= \mathbf{V}_{dr} - \mathbf{V}_{drs}\mathbf{V}_{ds}^{-1}\mathbf{V}_{dsr}. \end{aligned}$$

EB METHOD FOR POVERTY ESTIMATION

- For the nested-error model:

$$\boldsymbol{\mu}_{dr|ds} = \mathbf{X}_{dr}\boldsymbol{\beta} + \sigma_u^2 \mathbf{1}_{N_d-n_d}\mathbf{1}'_{n_d}\mathbf{V}_{ds}^{-1}(\mathbf{y}_{ds} - \mathbf{X}_{ds}\boldsymbol{\beta})$$
$$\mathbf{V}_{dr|ds} = \sigma_u^2(1-\gamma_d)\mathbf{1}_{N_d-n_d}\mathbf{1}'_{N_d-n_d} + \sigma_e^2\mathbf{I}_{N_d-n_d},$$

where

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

- Model for simulations:

$$\mathbf{y}_{dr} = \boldsymbol{\mu}_{dr|ds} + v_d\mathbf{1}_{N_d-n_d} + \boldsymbol{\epsilon}_{dr},$$

with

$$v_d \sim N\{0, \sigma_u^2(1-\gamma_d)\} \quad \text{and} \quad \boldsymbol{\epsilon}_{dr} \sim N(\mathbf{0}_{N_d-n_d}, \sigma_e^2\mathbf{I}_{N_d-n_d}).$$

- We only need to generate $N + D$ **univariate** normal random variables.

✓ *Molina and Rao (2010), CJS*

MONTE CARLO APPROXIMATION

(a) Generate L non-sample vectors $\mathbf{y}_{dr}^{(\ell)}$, $\ell = 1, \ldots, L$ from the (estimated) conditional distribution of $\mathbf{y}_{dr}|\mathbf{y}_{ds}$.

(b) Attach the sample elements to form a population vector $\mathbf{y}_d^{(\ell)} = (\mathbf{y}_{ds}, \mathbf{y}_{dr}^{(\ell)})$, $\ell = 1, \ldots, L$.

(c) Calculate the poverty measure with each population vector $F_{\alpha d}^{(\ell)} = h_\alpha(\mathbf{y}_d^{(\ell)})$, $\ell = 1, \ldots, L$. Then take the average over the L Monte Carlo generations:

$$\hat{F}_{\alpha d}^{EB} = E_{\mathbf{y}_{dr}}\left[F_{\alpha d}|\mathbf{y}_{ds}\right] \cong \frac{1}{L}\sum_{\ell=1}^{L} F_{\alpha d}^{(\ell)}.$$

NON-SAMPLED AREAS

- $Y_{dj}^{(\ell)}$ for $j = 1, \ldots, N_d$ and $\ell = 1, \ldots, L$ generated from

$$Y_{dj}^{(\ell)} = \mathbf{x}_{dj}'\hat{\boldsymbol{\beta}} + u_d^{(\ell)} + e_{dj}^{(\ell)}.$$
$$u_d^{(\ell)} \overset{iid}{\sim} N(0, \hat{\sigma}_u^2); \quad e_{dj}^{(\ell)} \overset{iid}{\sim} N(0, \hat{\sigma}_e^2).$$

- Calculate $\hat{F}_{\alpha d}^{(\ell)}$ from $\{Y_{dj}^{(\ell)}\}$ and use

$$\hat{F}_{\alpha d}^{EB} \simeq \frac{1}{L}\sum_{\ell=1}^{L} \hat{F}_{\alpha d}^{(\ell)}$$

- $\hat{F}_{\alpha d}^{EB}$ is a synthetic estimator.

MSE ESTIMATION

- Construct bootstrap populations $\{Y_{dj}^{*(b)}, b = 1, \ldots, B\}$ from

$$Y_{dj}^* = \mathbf{x}_{dj}'\hat{\beta} + u_d^* + e_{dj}^*; \quad j = 1, \ldots, N_d,\ d = 1, \ldots, D.$$
$$u_d^* \overset{iid}{\sim} N(0, \hat{\sigma}_u^2); \quad e_{dj}^* \overset{iid}{\sim} N(0, \hat{\sigma}_e^2).$$

- Calculate bootstrap population parameters $F_{\alpha d}^*(b)$

- From each bootstrap population, take the sample with the same indexes S as in the initial sample and calculate EBs $F_{\alpha d}^{EB*}(b)$ using bootstrap sample data $\mathbf{y}_s^*$ and known $\mathbf{x}_{dj}$.

$$mse^*(\hat{F}_{\alpha d}^{EB}) = \frac{1}{B}\sum_{b=1}^{B}\{\hat{F}_{\alpha d}^{EB*}(b) - F_{\alpha d}^*(b)\}^2$$

WORLD BANK (WB) / ELL METHOD

- Elbers et al. (2003) also used nested error model on transformed variables Y_{dj}, using clusters as d.
- For comparability we take cluster as small area.
- Generate A bootstrap populations $\{Y_{dj}^*(a), a = 1, \ldots, A\}$
- Calculate $F_{\alpha d}^*(a), a = 1, \ldots, A$. Then ELL estimator is:

$$\hat{F}_{\alpha d}^{(ELL)} = \frac{1}{A}\sum_{a=1}^{A} F_{\alpha d}^*(a) = F_{\alpha d}^*(\cdot)$$

WORLD BANK (WB) / ELL METHOD

- MSE estimator:

$$mse(\hat{F}_{\alpha d}^{ELL}) = \frac{1}{A}\sum_{a=1}^{A}\{F_{\alpha d}^{*}(a) - F_{\alpha d}^{*}(\cdot)\}^2$$

- If the mean $\bar{Y}_d$ is the parameter of interest, then

$$\hat{\bar{Y}}_d^{(ELL)} \simeq \bar{X}_d\hat{\boldsymbol{\beta}}$$

- $\hat{\bar{Y}}_d^{(ELL)}$ is a regression synthetic estimator.
- For non-sampled areas, $\hat{F}_{\alpha d}^{ELL}$ is essentially equivalent to $\hat{F}_{\alpha d}^{EB}$.

MODEL-BASED EXPERIMENT

- We simulated $I = 1000$ populations from the nested error model;
- For each population, we computed the true domain poverty measures.
- We computed the MSE of the EB estimators as

$$\mathrm{MSE}(\hat{F}_{\alpha d}^{EB}) = \frac{1}{I} \sum_{i=1}^{I} \left(\hat{F}_{\alpha d}^{EB(i)} - F_{\alpha d}^{(i)} \right)^2, \quad d = 1, \ldots, D.$$

- Similarly for direct and ELL estimators.

MODEL-BASED EXPERIMENT

Sizes:

$N = 20000$
$D = 80$
$N_d = 250,\ d = 1, \dots, D$
$n_d = 50,\ d = 1, \dots, D$

Variance components:

$\sigma_e^2 = (0{,}5)^2$
$\sigma_u^2 = (0{,}15)^2$

MODEL-BASED EXPERIMENT

Explanatory variables:

$$X_1 \in \{0,1\}, \quad p_{1d} = 0.3 + 0.5d/80, \quad d = 1, \dots, D.$$
$$X_2 \in \{0,1\}, \quad p_{2d} = 0.2, \quad d = 1, \dots, D.$$

Coefficients:

$$\beta = (3, 0.03, -0.04)'.$$

- The response increases when moving from $X_1 = 0$ to $X_1 = 1$, and decreases when moving from $X_2 = 0$ to $X_2 = 1$.
- The "richest" people are those with $X_1 = 1$ and $X_2 = 0$.
- The last areas have "richer" individuals than the first areas, i.e., poverty decreases with the area index.

POVERTY INCIDENCE

- Bias negligible for all three estimators (EB, direct and ELL).
- EB much more efficient than ELL and direct estimators.
- ELL even less efficient than direct estimators!

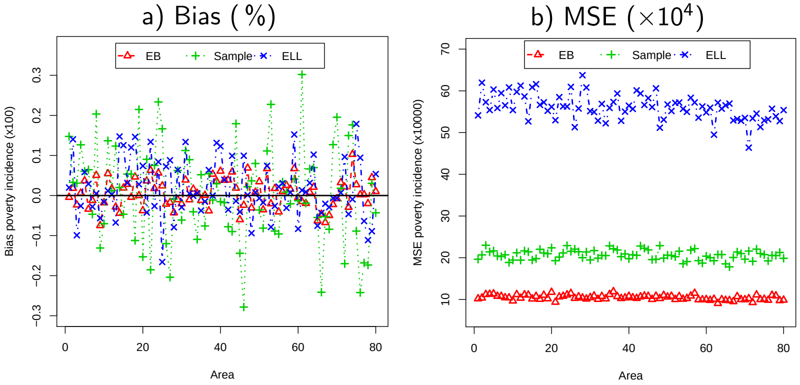


Figure 1. a) Bias and b) MSE of EB, direct and ELL estimators of poverty incidences F_{0d} for each area d.

POVERTY GAP

- Same conclusions as for poverty incidence.

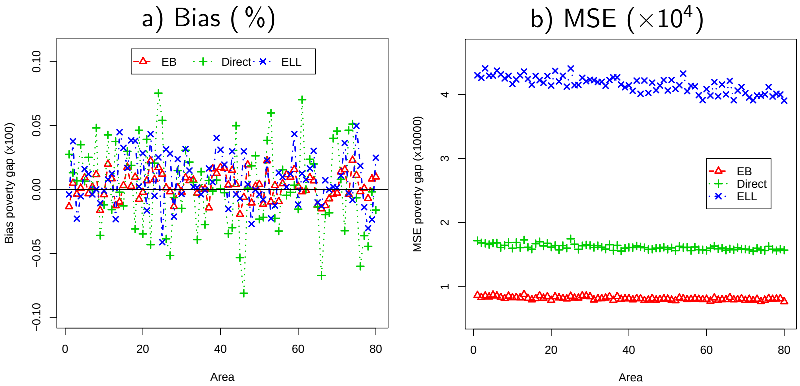


Figure 2. a) Bias and b) MSE of EB, direct and ELL estimators of poverty gaps F_{1d} for each area d.

BOOTSTRAP MSE

- The bootstrap MSE tracks true MSE.

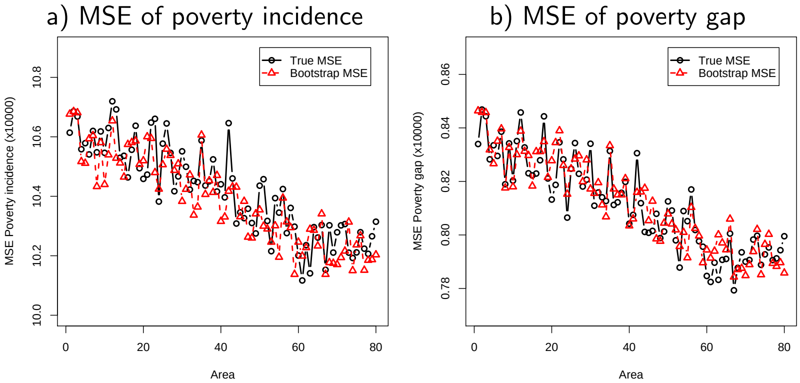


Figure 3. True MSEs and bootstrap estimators ($\times 10^4$) of EB estimators with $B = 500$ for each area d.

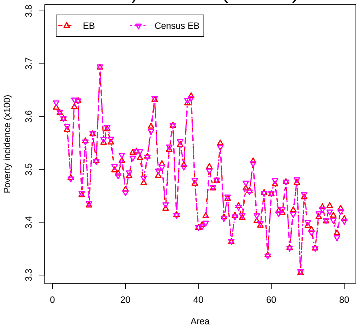
CENSUS EB METHOD

- When sample data cannot be linked with census auxiliary data, in steps (a) and (b) of EB method generate a full census from

$$\mathbf{y}_d = \hat{\mu}_{d|ds} + v_d \mathbf{1}_{N_d} + \epsilon_d, \quad \hat{\mu}_{d|ds} = \mathbf{X}_d \hat{\boldsymbol{\beta}} + \hat{\sigma}_u^2 \mathbf{1}_{N_d} \mathbf{1}'_{n_d} \hat{\mathbf{V}}_{ds}^{-1} (\mathbf{y}_{ds} - \mathbf{X}_{ds} \hat{\boldsymbol{\beta}}).$$

- Practically the same as original EB method.

a) Mean ($\times 100$)



b) MSE ($\times 10^4$)

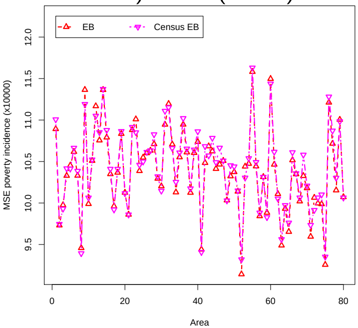


Figure 4. a) Mean and b) MSE of EB and Census EB estimators of poverty gaps F_{1d} for each area d.

FAST EB METHOD

- For large populations or computationally complex indicators.
- Instead of generating a full census in the EB method, generate only samples from the conditional distribution and compute direct estimators instead of true values.
- Fast EB method quite close to original EB.

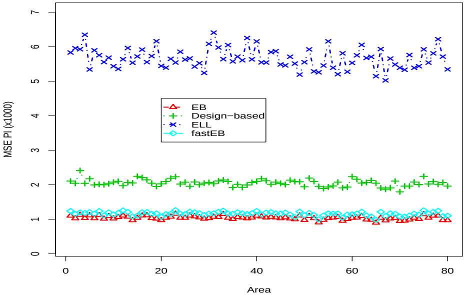


Figure 5. MSE ($\times 10^4$) of EB, direct, ELL and fast EB estimators of PI.

✓ *Ferretti, Molina & Lemmi, Submitted to JISAS*

SKEW-NORMAL EB

- Nester error model with e_{dj} skew normal

$$u_d \overset{iid}{\sim} N(0, \sigma_u^2), \quad e_{dj} \overset{iid}{\sim} SN(0, \sigma_e^2, \lambda_e)$$

$$\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma_u^2, \sigma_e^2, \lambda_e)'$$

$\lambda_e = 0$ corresponds to Normal

- As in the Normal case, EB estimator can be computed by generating only **univariate** normal variables, conditionally given a half-normal variable $T = t$.
- SN-EB was computed assuming $\boldsymbol{\theta}$ is known.

SKEW-NORMAL EB SIMULATION

- EB biased under significant skewness ($\lambda > 1$) unlike SN EB.

a) Bias of SN-EB estimator b) Bias of EB estimator

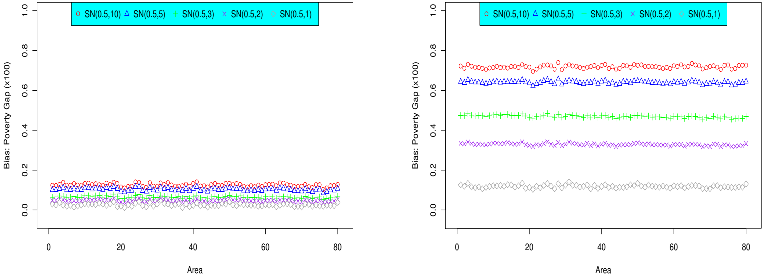


Figure 6. Bias of a) SN-EB estimator and b) EB estimator under skew normal distributions for error term for $\lambda = 1, 2, 3, 5, 10$.

✓ *Diallo & Rao, Work in progress*

SKEW-NORMAL EB SIMULATION

- RMSE = MSE(EB)/MSE(SN-EB)
- SN-EB significantly more efficient than EB when $\lambda > 1$.

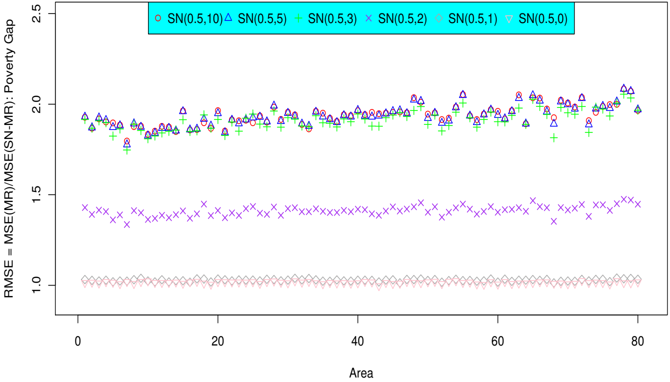


Figure 7. RMSE for skewness parameter $\lambda = 1, 2, 3, 5, 10$.

SMALL AREA DISTRIBUTION FUNCTION

- EB good for estimating other non-linear characteristics such as distrib. function.

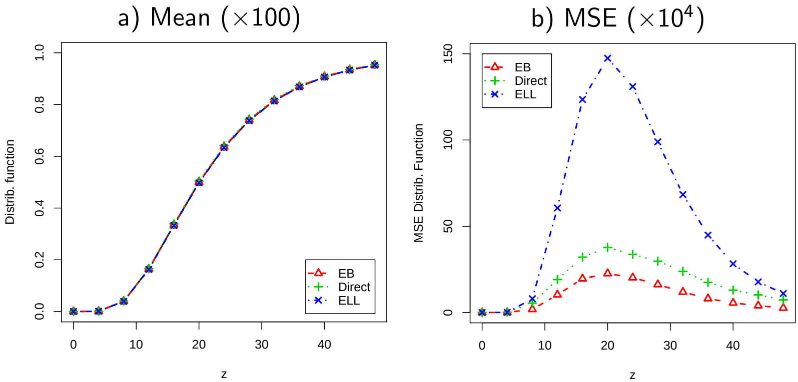


Figure 8. a) Mean of true, EB, direct and ELL estimators of the distribution function and b) MSE of estimators for area $d = 1$.

HIERARCHICAL BAYES METHOD

- **Reparameterized nested-error model:**

$$y_{di}|u_d,\boldsymbol{\beta},\sigma^2 \stackrel{ind}{\sim} N(\mathbf{x}'_{di}\boldsymbol{\beta}+u_d,\sigma^2)$$
$$u_d|\rho,\sigma^2 \stackrel{ind}{\sim} N\left(0,\frac{\rho}{1-\rho}\,\sigma^2\right)$$

- Noninformative prior: $\pi(\boldsymbol{\beta},\sigma^2,\rho)\propto 1/\sigma^2$.
- Proper posterior density (provided $\mathbf{X}$ full column rank):

$$\pi(\mathbf{u},\boldsymbol{\beta},\sigma^2,\rho|\mathbf{y}_s)=\pi_1(\mathbf{u}|\boldsymbol{\beta},\sigma^2,\rho,\mathbf{y}_s)\,\pi_2(\boldsymbol{\beta}|\sigma^2,\rho,\mathbf{y}_s)\,\pi_3(\sigma^2|\rho,\mathbf{y}_s)\,\pi_4(\rho|\mathbf{y}_s)$$

- $u_i|\boldsymbol{\beta},\sigma^2,\rho,\mathbf{y}_s\sim_{\text{ind}}$ *Normal*, $\boldsymbol{\beta}|\sigma^2,\rho,\mathbf{y}_s\sim$ *Normal*, $\sigma^{-2}|\rho,\mathbf{y}_s\sim$ *Gamma*.
- $\pi_4(\rho|\mathbf{y}_s)$ is not simple but ρ-values from it can be generated using a grid method.

✓ *Rao, Nandram & Molina, Work in progress*

HIERARCHICAL BAYES METHOD

- Very similar to original EB method (frequencial validity).

a) Mean ($\times 100$) b) MSE ($\times 10^4$)

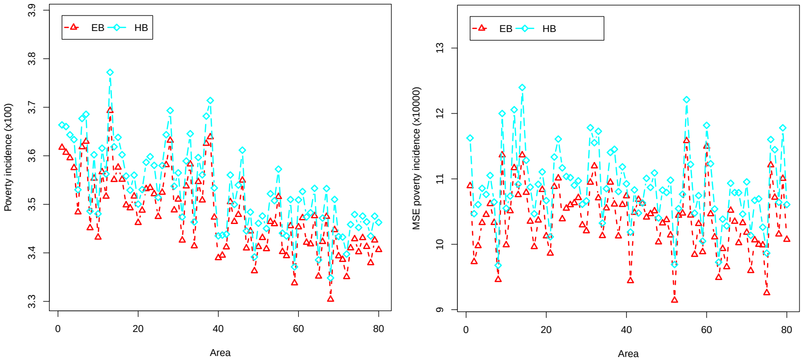


Figure 9. a) Mean and b) MSE of EB and HB estimators of poverty gaps F_{1d} for each area d.

✓ *Rao, Nandram & Molina, Work in progress*

CONCLUSIONS

- We studied **EB and HB** estimation of **complex** small area parameters.
- Method applicable to **unit level** data.
- EB method assumes **normality** for some transformation of the variable of interest. EB work extended to **skew normal** distributions.
- It requires the knowledge of **all population values** of the auxiliary variables.
- It requires **computational effort** because large number of populations are generated. **Fast EB method** available.

CONCLUSIONS

- Original EB method, unlike ELL method, requires **linking** sample with census data for the auxiliary variables. **Census EB** method avoids the linking and is practically the same as original EB.
- Both EB and ELL methods assume that the sample is **non-informative**, that is, the model for the population holds good for the sample. Under informative sampling, probably both methods are biased. Currently an extension of EB method accounting for **informative** sampling is being studied.