Small Area Estimation Methods, Applications and Practical Demonstration

Part 2: Model-based Methods

J.N.K. Rao

School of Mathematics and Statistics, Carleton University

Isabel Molina

Department of Statistics, Universidad Carlos III de Madrid

Fay-Herriot model

(i) Model linking area means $\bar{Y}_i = Y_i/N_i$:

$$\theta_i := g(\bar{Y}_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \upsilon_i, \quad i = 1, \ldots, m$$

$$\upsilon_i \overset{iid}{\sim} (0, \sigma_\upsilon^2), \quad \sigma_\upsilon^2 \text{ unknown}$$

(ii) Sampling model:

$$\hat{\theta}_i^{DIR} = \theta_i + e_i, \quad i = 1, \ldots, m$$

$$e_i|\theta_i \overset{ind}{\sim} (0, \psi_i), \quad \psi_i \text{ known}$$

(iii) Combined model: Linear mixed model

$$\hat{\theta}_i^{DIR} = \mathbf{x}_i^T \boldsymbol{\beta} + \upsilon_i + e_i, \quad i = 1, \ldots, m$$

Fay-Herriot model

- Usually the ψ_i's are replaced by smoothed estimators based on $v(\hat{\theta}_i^{DIR})$. These are treated as known.
- When $g(\cdot)$ is nonlinear, it is not realistic to assume $E(e_i|\theta_i) = 0$ because area sample size is small.

(ii*) More realistic sampling error model:

$$\hat{Y}_i^{DIR} = Y_i + e_i^*, \quad E(e_i^*|Y_i) = 0$$

In this case (i) and (ii*) cannot be combined to produce a linear mixed model. (i) and (ii*) are mismatched.

BLUP under the Fay-Herriot model

Best linear unbiased predictor (BLUP)

Under model (iii), the linear and unbiased estimator $\tilde{\theta}_i$ of $\theta_i = \mathbf{x}_i^T \boldsymbol{\beta} + \upsilon_i$ which minimizes $\text{MSE}(\tilde{\theta}_i) = E(\tilde{\theta}_i - \theta_i)^2$ is

$$\tilde{\theta}_i^{BLUP} = \mathbf{x}_i^T \tilde{\boldsymbol{\beta}} + \tilde{\upsilon}_i,$$

where

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{i=1}^{m} \gamma_i \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \sum_{i=1}^{m} \gamma_i \mathbf{x}_i \hat{\theta}_i^{DIR},$$

$$\tilde{\upsilon}_i = \gamma_i (\hat{\theta}_i^{DIR} - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}), \quad \gamma_i = \frac{\sigma_\upsilon^2}{\sigma_\upsilon^2 + \psi_i}$$

BLUP under the Fay-Herriot model

- BLUP can be expressed as

$$\tilde{\theta}_i^{BLUP} = \gamma_i \, \hat{\theta}_i^{DIR} + (1 - \gamma_i)\mathbf{x}_i^T \tilde{\boldsymbol{\beta}}$$

- Weighted combination of direct estimator $\hat{\theta}_i^{DIR}$ and "regression synthetic" estimator $\mathbf{x}_i^T \tilde{\boldsymbol{\beta}}$.
- It gives more weight to $\hat{\theta}_i^{DIR}$ when sampling variance ψ_i small.
- It moves towards synthetic estimator as ψ_i increases or σ_v^2 decreases.

Empirical BLUP (EBLUP)

- When random effects variance σ_v^2 is unknown, $\tilde{\theta}_i^{BLUP}$ depends on σ_v^2 through $\tilde{\boldsymbol{\beta}}$ and γ_i:

$$\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\sigma_v^2), \quad \tilde{\theta}_i^{BLUP}(\sigma_v^2)$$

- Empirical BLUP (EBLUP) of θ_i: Replace σ_v^2 in the BLUP by an estimator $\hat{\sigma}_v^2$

$$\hat{\theta}_i^{EBLUP} = \tilde{\theta}_i^{BLUP}(\hat{\sigma}_v^2), \quad i = 1, \ldots, m$$

- Non-samples areas: Use regression synthetic estimator:

$$\mathbf{x}_i^T \hat{\boldsymbol{\beta}}, \text{ where } \hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\sigma}_v^2)$$

Model fitting

- **Fay-Herriot method:** Solve iteratively for σ_v^2

$$\sum_{i=1}^{m} \frac{\left(\hat{\theta}_i^{DIR} - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}(\sigma_v^2)\right)^2}{\sigma_v^2 + \psi_i} = m - p.$$

Stop when iterations converge: $\tilde{\sigma}_v^2$

Take $\hat{\sigma}_v^2 = \max(\tilde{\sigma}_v^2, 0)$ and $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\sigma}_v^2)$.

Normality is not needed.

Model fitting

- **Maximum likelihood (ML):** Assumes normality

$$\hat{\theta}_i^{DIR} \overset{ind}{\sim} N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_\upsilon^2 + \psi_i)$$

  ML estimators remain consistent without normality.

- **Restricted maximum likelihood (REML):** Reduces the bias of ML estimators for finite sample size n.
- **Prasad-Rao method:** Based on method of moments. Good starting values for iterative fitting algorithms.

(✓ Prasad and Rao, 1990)

Best/Bayes predictor under normality

Best/Bayes estimator

Consider models (i) and (ii) with Normality assumption. The best (or Bayes) estimator of θ_i is

$$\tilde{\theta}_i^B(\boldsymbol{\beta}, \sigma_v^2) = E(\theta_i | \hat{\theta}_i^{DIR}) = \gamma_i \hat{\theta}_i^{DIR} + (1 - \gamma_i)\mathbf{x}_i^T \boldsymbol{\beta}$$

- Empirical best/Bayes (EB) estimator of θ_i:

$$\hat{\theta}_i^{EB} = \tilde{\theta}_i^B(\hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2) = \hat{\theta}_i^{EBLUP}$$

- Mean squared error of Best estimator:

$$\text{MSE}(\tilde{\theta}_i^B) = \gamma_i \psi_i := g_{1i}(\sigma_v^2)$$

Fay-Herriot model for poverty estimation

- FGT poverty indicator for area i:

$$F_{\alpha i} = \frac{1}{N_i}\sum_{j=1}^{N_i} F_{\alpha ij}, \quad F_{\alpha ij} = \left(\frac{z - E_{ij}}{z}\right)^{\alpha} I(E_{ij} < z)$$

- Direct estimator of $F_{\alpha i}$ (N_i known):

$$\hat{F}_{\alpha i}^{DIR} = \frac{1}{N_i}\sum_{j \in s_i} w_{ij} F_{\alpha ij}$$

- Fay–Herriot model can be used with

$$\theta_i = F_{\alpha i}, \quad \hat{\theta}_i^{DIR} = \hat{F}_{\alpha i}^{DIR}, \quad \psi_i = v(\hat{F}_{\alpha i}^{DIR})$$

Application 1: Estimation of mean per capita income in US small places (✓ *Fay and Herriot, 1979*)

- 20% sample in the 1970 U.S. census.
- $\bar{Y}_i$ mean per capita income (PCI) for small place i.
- $\theta_i = \log \bar{Y}_i, \;\; \hat{\theta}_i^{DIR} = \log \hat{\bar{Y}}_i^{DIR}$
- x_i associated county value of $\log(\bar{Y}_i)$ from 1970 census.
- Using Taylor approximation,

$$V(\hat{\theta}_i^{DIR}) = V(\log \hat{\bar{Y}}_i^{DIR}) \approx V(\hat{\bar{Y}}_i^{DIR})/\bar{Y}_i^2 = [CV(\hat{\bar{Y}}_i^{DIR})]^2.$$

- $CV(\hat{\bar{Y}}_i^{DIR}) \approx 3/\hat{N}_i^{\frac{1}{2}} \Rightarrow \psi_i = V(\hat{\theta}_i^{DIR}) = 9/\hat{N}_i$

  Example: If $\hat{N}_i = 100$, then $CV(\hat{\bar{Y}}_i^{DIR}) \approx 30\%$.
- Compromise EB estimator similar to compromise JS is used.

Application 1: Choice of models

(1) $x_1 = 1$, $x_2 = \log(\text{PCI county})$; $p = 2$

(2) x_1, x_2, $x_3 = \log(\text{value of housing for the place})$
$x_4 = \log(\text{value of housing for the county})$; $p = 4$

(3) x_1, x_2, $x_5 = \log$(ave. gross income per exemption from the 1969 tax returns for the place)
$x_6 =$(ave. gross income per exemption from the 1969 tax returns for the county); $p = 4$

(4) $x_1, \ldots, x_6$; $p = 6$
$\hat{\sigma}_v^2$ average measure-of-fit of model allowing for sampling errors
$N_i = 200 \Rightarrow \psi_i = 9.0/200 = 0.045$. If also $\hat{\sigma}_v^2 = 0.045$, then $\hat{\gamma}_i = 1/2 \Rightarrow$ EB est. for 200 $\approx$ Dir est. for 400

Table 3.1: Resulting value of $\hat{\sigma}_v^2$ for states with more than 500 small places

State	Model (1)	Model (2)	Model (3)	Model (4)
Illinois	0.036	0.032	0.019	0.017
Iowa	0.029	0.011	0.017	0.000
Kansas	0.064	0.048	0.016	0.020
Minnesota	0.063	0.055	0.014	0.019
Missouri	0.061	0.033	0.034	0.017
Nebraska	0.065	0.041	0.019	0.000
N. Dakota	0.072	0.081	0.020	0.004
S. Dakota	0.138	0.138	0.014	*
Wisconsin	0.042	0.025	0.025	0.004

Application 1: Conclusions

- Models involving either tax or housing data, but especially both, provide better fit than those based on county values alone.
- Model (4) and lesser extent models (2), (3) better fit than model (1).
- $\hat{\sigma}_v^2$ for model (4) much smaller than 0.045 for North & South Dakota, Nebraska, Wisconsin, Iowa.

Evaluation study

- Complete census of a random sample of places in 1973 collecting income for 1972.
- Comparison of direct, compromise EB and synthetic (county) estimates for those places with true values
- Direct estimates and compromise EB estimates for 1972 obtained by multiplying the 1970 census estimates by updating factors f_i:

$$\hat{\bar{Y}}_i^{DIR}, \quad \exp(\hat{\theta}_i^{EB})$$

- $\exp(\hat{\theta}_i^{EB})$ not equal to EB estimator of $\bar{Y}_i$
- Percentage absolute relative error

$$\%ARE = \frac{|\text{estimate} - \text{true value}|}{\text{true value}} \times 100$$

Table 3.2 (a): Values of % ARE for places with popn. less than 500

Census area	Direct	EB	County
1	10.2	14.0	12.9
2	4.4	10.3	30.9
3	34.1	26.2	9.1
4	1.3	8.3	24.6
5	34.7	21.8	6.6
6	22.1	19.8	14.6
7	14.1	4.1	18.7
8	18.1	4.7	25.9
9	60.7	78.7	99.7
10	47.7	54.7	95.3
11	89.1	65.8	86.5
12	1.7	9.1	12.7
13	11.4	1.4	6.6
14	8.6	5.7	23.5
15	23.6	25.3	34.3
16	53.6	10.5	11.7
17	51.4	14.4	23.7
Average	28.6	22.0	31.6

Table 3.2 (b): % ARE for places with popn. between 500 and 999

Census area	Direct	EB	County
1	36.5	28.0	36.0
2	8.5	4.1	9.3
3	7.4	2.7	7.7
4	13.6	16.9	13.6
5	25.3	16.3	25.8
6	33.2	34.1	32.9
7	9.2	7.2	9.9
Average	19.1	15.6	19.3

- EB est. shows smaller average errors and lower extreme errors than either direct or county estimates.
- However, EB is consistently higher than the true value (Missing income not imputed in the special census unlike in the 1970 census: downward bias).

Application 2: Small Area Income and Poverty Estimates (SAIPE)

(✓ National Research Council, 2000)

- **Objective**: Estimate number of poor school-age children (aged 5-17) at the county level and school district level.
- These estimates used to allocate funds to counties (over $7 billion for 1997-8). States distribute these funds to school district within county.
- $\hat{Y}_i^{DIR}$ based on 3-year weighted average from the Current Population Survey (CPS).
- Sampling Model:

$$\hat{\theta}_i^{DIR} = \log(\hat{Y}_i^{DIR}) = \theta_i + e_i, \quad e_i|\theta_i \sim N(0, \sigma_e^2/n_i)$$

Application 2: Small Area Income and Poverty Estimates (SAIPE)

- x-variables: food stamps, poor from tax forms, number of exemptions, population, last census poor (all in log scale).
- Unknown sampling variances ψ_i but reliable 1990 census estimates $\hat{\psi}_{ic}$ available.
- Assume census random effects υ_{ic} follow the same distribution as υ_i and take $\sigma_\upsilon^2 = \sigma_{\upsilon c}^2$.
- From 1990 census, estimate σ_υ^2 using $\hat{\psi}_{ic}$: $\hat{\sigma}_\upsilon^2$.

  Use $\hat{\sigma}_\upsilon^2$ to get $\tilde{\sigma}_e^2$ assuming $\psi_i = \sigma_e^2/n_i$.

  Treat $\tilde{\psi}_i = \tilde{\sigma}_e^2/n_i$ as true $\psi_i \Rightarrow \hat{\theta}_i^{EB} \Rightarrow \hat{Y}_i^{EB}$
- Benchmark $\hat{Y}_i^{EB}$ to state estimate of poor based on a state model.

Application 2: Evaluation for year 1990

- Comparison of SAIPE and two more estimates (U1 and U2) based on previous census (1980) to true values obtained from 1990 census.
- Estimate U1: County shares of poor within state same as in previous census. Benchmark to state estimates for 1990.
- Estimate U2: County ratios poor/population same as in previous census: Multiply previous census ratio by current population estimate. Benchmark to current state estimate.
- SAIPE estimates based on Fay-Herriot model.
- Mean over counties of ARE (%)

U1	U2	SAIPE
26.1	26.2	16.4

- SAIPE estimates better than U1 or U2.

Application 3: 1991 Census undercount in Canada

(✓ *Dick, 1995*)

- i province × age × sex ($i = 1, \ldots, 96$)
- T_i true (unknown) count
- C_i census count
- $\theta_i = T_i / C_i$ adjustment factor
- M_i = number missed $= T_i - C_i = C_i(\theta_i - 1)$
- $\hat{M}_i^{DIR}$ direct estimate of M_i (post-enumeration survey)
- Sampling variances ψ_i: linear regression of log-direct variance est. $\hat{\psi}_i = v(\hat{M}_i^{DIR})$ on $\log(C_i)$:

  Fitted model: $\log \hat{\psi}_i = -6.13 - 0.28 \log(C_i)$

Application 3: 1991 Census undercount in Canada

- Selection of x-variables: 42 variables subjected to backward stepwise regression.
- Model diagnostics: Analysis of residuals

$$r_i = (\hat{\theta}_i^{EB} - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})/(\hat{\sigma}_v^2 + \psi_i)^{\frac{1}{2}}$$

- EB estimator for each province p and age-sex group a:

$$\hat{\theta}_{pa}^{EB} \Rightarrow \hat{M}_{pa}^{EB} = C_{pa}(\hat{\theta}_{pa}^{EB} - 1)$$

- Raking to make them confirm to margins $\hat{M}_{p+}^{DIR}$ and $\hat{M}_{+a}^{DIR}$ (reliable direct estimates).
- Raking tends to shrink EB towards direct estimates.

Mean squared error

- Mean squared error of EB/EBLUP with respect to model:

$$\mathrm{MSE}(\hat{\theta}_i^{EB}) = E(\hat{\theta}_i^{EB} - \theta_i)^2$$

- Approximation of MSE by Taylor linearization method: Under normality,

$$\mathrm{MSE}(\hat{\theta}_i^{EB}) \approx g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) + g_{3i}(\sigma_v^2),$$

$g_{1i}(\sigma_v^2) = O(1)$ due to prediction of random effects
$g_{2i}(\sigma_v^2) = O(1/m)$ due to estimation of $\boldsymbol{\beta}$
$g_{3i}(\sigma_v^2) = O(1/m)$ due to estimation of σ_v^2

Mean squared error

- Explicit expressions for g_{1i}, g_{2i} and g_{3i}:

$$
\begin{aligned}
g_{1i}(\sigma_v^2) &= \gamma_i \psi_i, \\
g_{2i}(\sigma_v^2) &= \sigma_v^2 (1-\gamma_i)^2 \mathbf{x}_i^T \left(\sum_{i=1}^{m} \gamma_i \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \mathbf{x}_i, \\
g_{3i}(\sigma_v^2) &= (1-\gamma_i)^2 \gamma_i \sigma_v^{-2} \bar{V}(\hat{\sigma}_v^2),
\end{aligned}
$$

- $\bar{V}(\hat{\sigma}_v^2)$ asymptotic variance of $\hat{\sigma}_v^2$: It depends on the estimation method used for σ_v^2.

Mean squared error

- Nearly unbiased MSE estimator when $\hat{\sigma}_v^2$ is obtained by REML:

$$\text{mse}(\hat{\theta}_i^{EB}) = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{3i}(\hat{\sigma}_v^2)$$

  Nearly unbiasedness property:

$$E\left\{\text{mse}(\hat{\theta}_i^{EB})\right\} = \text{MSE}(\hat{\theta}_i^{EB}) + o(1/m)$$

- When $\hat{\sigma}_v^2$ is obtained by FH or ML methods, an extra term due to bias in $\hat{\sigma}_v^2$ must be added.

✓ *Rao (2003, p. 129)*

Jackknife estimation of MSE

$$\text{MSE}(\hat{\theta}_i^{EB}) = \text{MSE}(\tilde{\theta}_i^{B}) + E(\hat{\theta}_i^{EB} - \tilde{\theta}_i^{B})^2 =: M_{1i} + M_{2i}$$

(i) Delete $l-$th area and calculate estimators of $\boldsymbol{\beta}$ and σ_v^2: $\hat{\boldsymbol{\beta}}(l)$ and $\hat{\sigma}_v^2(l)$. Calculate EB estimator: $\hat{\theta}_i^{EB}(l) = \tilde{\theta}_i^{B}(\hat{\boldsymbol{\beta}}(l), \hat{\sigma}_v^2(l))$

(ii) Calculate a Jackknife estimator of M_{2i}:

$$\hat{M}_{2i} = \frac{m-1}{m} \sum_{l=1}^{m} [\hat{\theta}_i^{EB}(l) - \hat{\theta}_i^{EB}]^2$$

(iii) Calculate a bias-corrected est. of M_{1i}:

$$\hat{M}_{1i} = g_{1i}(\hat{\sigma}_v^2) - \frac{m-1}{m} \sum_{l=1}^{m} [g_{1i}(\hat{\sigma}_v^2(l)) - g_{1i}(\hat{\sigma}_v^2)]$$

(iv) Nearly unbiased Jackknife estimator: $\text{mse}_J(\hat{\theta}_i^{EB}) = \hat{M}_{1i} + \hat{M}_{2i}$

✓ *Jiang, Lahiri and Wan (2002)*

Parametric bootstrap estimation of MSE

(1) Model fitting: $\hat{\sigma}_v^2$, $\hat{\boldsymbol{\beta}}$ by FH, ML or REML.

(2) Generate bootstrap area effects:

$$v_i^* \stackrel{iid}{\sim} N(0, \hat{\sigma}_v^2), \quad i = 1, \ldots, m$$

(3) Generate, independently of $v_1^*, \ldots, v_m^*$, sampling errors:

$$e_i^* \stackrel{iid}{\sim} N(0, \psi_i), \quad i = 1, \ldots, m$$

(4) Generate bootstrap true values from the linking model:

$$\theta_i^* = \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + v_i^*, \quad i = 1, \ldots, m$$

and bootstrap direct estimators from the sampling model:

$$\hat{\theta}_i^{DIR*} = \theta_i^* + e_i^*, \quad i = 1, \ldots, m$$

Parametric bootstrap estimation of MSE

(5) Fit model to new bootstrap data $\{\hat{\theta}_i^{DIR*}, \mathbf{x}_i\}$ and calculate the bootstrap EB estimator: $\hat{\theta}_i^{EB*}$

(6) Repeat (2)–(5) a large number of times B:

$\theta_i^*(b)$ true value, $\hat{\theta}_i^{EB*}(b)$ EB estimator, $b = 1, \ldots, B$

(7) Bootstrap MSE estimator:

$$\text{mse}_B(\hat{\theta}_i^{EB}) = \frac{1}{B}\sum_{b=1}^{B}\{\hat{\theta}_i^{EB*}(b) - \theta_i^*(b)\}^2$$

- **Note 1:** Bias of order $O(1/m)$:

$$\text{mse}_B(\hat{\theta}_i^{EB}) \approx g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + g_{3i}(\hat{\sigma}_v^2)$$

- **Note 2:** Bootstrap applicable to other small area parameters: for example, EB estimator of small area total Y_i.

EB estimation of total Y_i

- Possible estimator of total $Y_i = g^{-1}(\theta_i)$:

$$\hat{Y}_i = g^{-1}(\hat{\theta}_i^{EB}) \rightarrow \text{It is not the EB estimator of } Y_i.$$

- Best estimator of $Y_i = g^{-1}(\theta_i)$:

$$\tilde{Y}_i^B = E_{\theta_i}\left[g^{-1}(\theta_i)|\hat{\theta}_i^{DIR}\right] \rightarrow \text{No closed form expression.}$$

- EB estimator of Y_i by Monte Carlo approximation: simulate a large number L of values $\theta_i(\ell)$, $\ell = 1, \ldots, L$, from the conditional distribution $\theta_i|\hat{\theta}_i^{DIR} \sim N(\tilde{\theta}_i^B, \gamma_i \psi_i)$, evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ and $\sigma_v^2 = \hat{\sigma}_v^2$. Then average over the L simulations:

$$\hat{Y}_i^{EB} \approx \frac{1}{L}\sum_{\ell=1}^{L} g^{-1}\{\theta_i(\ell)\},$$

(✓ Rao, 2003; p. 182).

EB estimation of area total Y_i

- Mean squared error of $\hat{Y}_i^{EB}$: Apply parametric bootstrap similarly as for $\hat{\theta}_i^{EB}$.
- For each bootstrap sample b, calculate $\hat{Y}_i^{EB*}(b)$ as described before and then take

$$\mathrm{mse}_B(\hat{Y}_i^{EB}) = \frac{1}{B}\sum_{b=1}^{B}\{\hat{Y}_i^{EB*}(b) - Y_i^*(b)\}^2,$$

where $Y_i^*(b) = g^{-1}\{\theta_i^*(b)\}$.

Bootstrap confidence intervals

- Pivot for a confidence interval for θ_i:

$$T_i = \frac{\hat{\theta}_i^{EB} - \theta_i}{\sqrt{g_{1i}(\hat{\sigma}_v^2)}}$$

- $t_1 = T_i(\alpha/2)$ and $t_2 = T_i(1-\alpha/2)$ quantiles of the distribution of T_i.
- $1-\alpha$ confidence interval for θ_i:

$$CI_{1-\alpha}(\theta_i) = \left[\hat{\theta}_i^{EB} - t_2\sqrt{g_{1i}(\hat{\sigma}_v^2)}, \hat{\theta}_i^{EB} - t_1\sqrt{g_{1i}(\hat{\sigma}_v^2)}\right]$$

- Distribution of T_i not known: Use bootstrap. ✓ *Chatterjee, Lahiri and Li (2008)*

Bootstrap confidence intervals

- Repeat steps (1)–(5) of the parametric bootstrap procedure a large number B of times. Calculate:

$$T_i^*(b) = \frac{\hat{\theta}_i^{EB*}(b) - \theta_i^*(b)}{\sqrt{g_{1i}(\hat{\sigma}_v^{2*}(b))}}, \quad b = 1, \ldots, B$$

- Order $T_i^*(b)$: $T_i^*(1) \leq \ldots \leq T_i^*(B)$

  Take sample quantiles $t_1^* = T_i^*(\alpha/2)$ and $t_2^* = T_i^*(1-\alpha/2)$

- Bootstrap interval for θ_i:

$$CI^*(\theta_i) = \left[\hat{\theta}_i^{EB} - t_2^* \sqrt{g_{1i}(\hat{\sigma}_v^2)}, \hat{\theta}_i^{EB} - t_1^* \sqrt{g_{1i}(\hat{\sigma}_v^2)}\right]$$

- Second-order accurate:

$$P(\theta_i \in CI^*(\theta_i)) = 1 - \alpha + o(1/m)$$

Estimation in non-sampled areas

- For an area k without sample data, we take the synthetic regression estimator of θ_k:

$$\hat{\theta}_k^{SYN} = \mathbf{x}_k^T \hat{\boldsymbol{\beta}}$$

- Mean squared error:

$$\begin{aligned} \text{MSE}(\hat{\theta}_k^{SYN}) &= E(\mathbf{x}_k^T \hat{\boldsymbol{\beta}} - \theta_k)^2 \\ &\approx \sigma_\upsilon^2 \mathbf{x}_k^T \left(\sum_{i=1}^{m} \gamma_i \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \mathbf{x}_k + \sigma_\upsilon^2 =: \tilde{g}_k(\sigma_\upsilon^2) + \sigma_\upsilon^2 \end{aligned}$$

- Nearly unbiased MSE estimator:

$$\text{mse}(\hat{\theta}_k^{SYN}) = \tilde{g}_k(\hat{\sigma}_\upsilon^2) + \hat{\sigma}_\upsilon^2$$

Estimation in non-sampled areas

- Estimator of total Y_k: $\tilde{Y}_k^{SYN} = g^{-1}(\hat{\theta}_k^{SYN})$
- Better estimator of Y_k: Using Monte Carlo approximation, generate $\theta_k(\ell)$, $\ell = 1, \ldots, L$, from $N(\mathbf{x}_k^T \hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2)$, and then take

$$\hat{Y}_k^{SYN} \approx \frac{1}{L} \sum_{\ell=1}^{L} g^{-1}\{\theta_k(\ell)\}$$

- MSE estimator obtained by parametric bootstrap.

Nested error model

- y_{ij} value of target variable for unit j within area i
- υ_i random effect of area i
- Nested error linear regression model:

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \upsilon_i + e_{ij}, \quad j = 1, \ldots, N_i, \; i = 1, \ldots, m$$
$$\upsilon_i \overset{iid}{\sim} N(0, \sigma_\upsilon^2), \quad e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$$

- Model in matrix notation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\upsilon} + \mathbf{e}$$

- Marginal expectation and variance:

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}, \quad V(\mathbf{y}) = \sigma_\upsilon^2 \mathbf{Z}\mathbf{Z}^T + \sigma_e^2 \mathbf{I}_N$$

BLUP under a finite population

More general linear model:

- $\mathbf{y} = (y_1, \ldots, y_N)^T$ population vector
- Linear model:

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}, \quad V(\mathbf{y}) = \mathbf{V}$$

- Decomposition into sample and non-sample parts:

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_s \\ \mathbf{y}_r \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_s \\ \mathbf{X}_r \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_{ss} & \mathbf{V}_{sr} \\ \mathbf{V}_{rs} & \mathbf{V}_{rr} \end{pmatrix}$$

- Linear target parameter:

$$\delta = \mathbf{a}^T\mathbf{y} = \mathbf{a}_s^T\mathbf{y}_s + \mathbf{a}_r^T\mathbf{y}_r$$

BLUP under a finite population

Best linear unbiased predictor (BLUP): V known

The linear predictor $\tilde{\delta} = \boldsymbol{\alpha}^T \mathbf{y}_s$ that is solution to the problem:

$$\begin{array}{ll} \min_{\boldsymbol{\alpha}} & \text{MSE}(\tilde{\delta}) = E(\tilde{\delta} - \delta)^2 \\ \text{s.t.} & E(\tilde{\delta} - \delta) = 0 \text{ (model unbiased)} \end{array}$$

is given by

$$\tilde{\delta}^{BLUP} = \mathbf{a}_s^T \mathbf{y}_s + \mathbf{a}_r^T \tilde{\mathbf{y}}_r^{BLUP},$$

where

$$\tilde{\mathbf{y}}_r^{BLUP} = \mathbf{X}_r \tilde{\boldsymbol{\beta}} + \mathbf{V}_{rs} \mathbf{V}_{ss}^{-1} (\mathbf{y}_s - \mathbf{X}_s \tilde{\boldsymbol{\beta}}),$$
$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}_s^T \mathbf{V}_{ss}^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s^T \mathbf{V}_{ss}^{-1} \mathbf{y}_s$$

✓ Scott & Smith (1969); ✓ Royall (1970)

BLUP under a finite population

- For a small area mean $\delta = \bar{Y}_i$, the BLUP takes the form:

$$\tilde{\bar{Y}}_i^{BLUP} = \frac{1}{N_i}\left(\sum_{j \in s_i} y_{ij} + \sum_{j \in r_i} \tilde{y}_{ij}^{BLUP}\right),$$

where $\tilde{y}_{ij}^{BLUP} = \mathbf{x}_{ij}^T\tilde{\boldsymbol{\beta}} + \tilde{\upsilon}_i$ and $\tilde{\upsilon}_i = \gamma_i(\bar{y}_i - \bar{\mathbf{x}}_i^T\tilde{\boldsymbol{\beta}})$ with $\gamma_i = \sigma_\upsilon^2/(\sigma_\upsilon^2 + \sigma_e^2 n_i^{-1})$.

- When $n_i/N_i \approx 0$, then:

$$\tilde{\bar{Y}}_i^{BLUP} \approx \gamma_i\left\{\bar{y}_i + (\bar{\mathbf{X}}_i - \bar{\mathbf{x}}_i)^T\tilde{\boldsymbol{\beta}}\right\} + (1-\gamma_i)\bar{\mathbf{X}}_i^T\tilde{\boldsymbol{\beta}}$$

- Weighted average of “survey regression” estimator $\bar{y}_i + (\bar{\mathbf{X}}_i - \bar{\mathbf{x}}_i)^T\tilde{\boldsymbol{\beta}}$ and regression synthetic estimator $\bar{\mathbf{X}}_i^T\tilde{\boldsymbol{\beta}}$.

BLUP under a finite population

- Usually $\mathbf{V}$ depends on unknown parameters:

$$\mathbf{V} = \mathbf{V}(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \text{ unknown}$$

- $\hat{\boldsymbol{\theta}}$ estimator of $\boldsymbol{\theta}$ obtained by:

  (a) Henderson method III (fitting constants method),
  (b) ML,
  (c) REML.

- Empirical BLUP (EBLUP) of δ:

$$\hat{\delta}^{EBLUP} = \tilde{\delta}^{BLUP}(\hat{\boldsymbol{\theta}})$$

Best predictor under a finite population

Best Predictor (BP)

Consider the target quantity $\delta = h(\mathbf{y})$, not necessarily linear. The predictor $\tilde{\delta}$ which minimizes $\text{MSE}(\tilde{\delta}) = E(\tilde{\delta} - \delta)^2$ is

$$\tilde{\delta}^B = E_{\mathbf{y}_r}(\delta | \mathbf{y}_s).$$

- For a linear model with $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $V(\mathbf{y}) = \mathbf{V}(\boldsymbol{\theta})$ with $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ unknown, the BP depends on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$:

$$\tilde{\delta}^B = \tilde{\delta}^B(\boldsymbol{\beta}, \boldsymbol{\theta}).$$

- Empirical Best Predictor (EBP): $\hat{\boldsymbol{\theta}}$ estimator of $\boldsymbol{\theta}$. Then

$$\hat{\delta}^{EB} = \tilde{\delta}^B(\tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}})$$

Best predictor under a finite population

- Particular case: Consider a linear target parameter

$$\delta = \mathbf{a}^T \mathbf{y} = \mathbf{a}_s^T \mathbf{y}_s + \mathbf{a}_r^T \mathbf{y}_r$$

If $\mathbf{y}$ is normally distributed, then BP is

$$\tilde{\delta}^B = \mathbf{a}_s^T \mathbf{y}_s + \mathbf{a}_r^T \tilde{\mathbf{y}}_r^B,$$

where

$$\tilde{\mathbf{y}}_r^B = \mathbf{X}_r \boldsymbol{\beta} + \mathbf{V}_{rs} \mathbf{V}_{ss}^{-1} (\mathbf{y}_s - \mathbf{X}_s \boldsymbol{\beta}).$$

- In this case EBP is equal to EBLUP.

EB method for poverty estimation

- **Assumption:** There exists a transformation $y_{ij} = T(E_{ij})$ of the welfare variables E_{ij} with Normal distribution

$$\mathbf{y} \sim N(\boldsymbol{\mu}, \mathbf{V})$$

- If $\boldsymbol{\mu}$ and $\mathbf{V}$ contain unknown parameters, we estimate them.
- The distribution of $\mathbf{y}_r$ given $\mathbf{y}_s$ is

$$\mathbf{y}_r | \mathbf{y}_s \sim N(\boldsymbol{\mu}_{r|s}, \mathbf{V}_{r|s}),$$

where

$$\boldsymbol{\mu}_{r|s} = \boldsymbol{\mu}_r + \mathbf{V}_{rs}\mathbf{V}_s^{-1}(\mathbf{y}_s - \boldsymbol{\mu}_s), \quad \mathbf{V}_{r|s} = \mathbf{V}_r - \mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{V}_{sr}$$

EB method for poverty estimation

- FGT poverty indicator for area i:

$$F_{\alpha i} = \frac{1}{N_i}\sum_{j=1}^{N_i} F_{\alpha ij}, \quad F_{\alpha ij} = \left(\frac{z - E_{ij}}{z}\right)^{\alpha} I(E_{ij} < z) =: h_\alpha(y_{ij})$$

- EB estimator of $F_{\alpha i}$:

$$\begin{aligned}\hat{F}_{\alpha i}^{EB} &= E_{\mathbf{y}_r}\left[\frac{1}{N_i}\sum_{j=1}^{N_i} F_{\alpha ij} \middle| \, \mathbf{y}_s\right] = \frac{1}{N_i}\sum_{j=1}^{N_i} E_{\mathbf{y}_r}\left[F_{\alpha ij} \middle| \, \mathbf{y}_s\right] \\ &= \frac{1}{N_i}\left(\sum_{j\in s_i} F_{\alpha ij} + \sum_{j\in r_i} E_{\mathbf{y}_r}\left[F_{\alpha ij}|\mathbf{y}_s\right]\right)\end{aligned}$$

Monte Carlo approximation of EBP

- The conditional expectations $E_{\mathbf{y}_r}[F_{\alpha ij}|\mathbf{y}_s]$ can be approximated by Monte Carlo.
- Generate L non-sample vectors $\mathbf{y}_r^{(\ell)}$, $\ell = 1, \ldots, L$ from the conditional distribution of $\mathbf{y}_r|\mathbf{y}_s$.
- For each element $y_{ij}^{(\ell)}$ of $\mathbf{y}_r^{(\ell)}$, calculate $F_{\alpha ij}^{(\ell)} = h_\alpha(y_{ij}^{(\ell)})$, $\ell = 1, \ldots, L$ and average over the L Monte Carlo replicates,

$$E_{\mathbf{y}_r}[F_{\alpha ij}|\mathbf{y}_s] \cong \frac{1}{L}\sum_{\ell=1}^{L} F_{\alpha ij}^{(\ell)}.$$

Parametric bootstrap

(1) Model fitting by ML, REML or Henderson method III: $\hat{\sigma}_u^2$, $\hat{\sigma}_e^2$, $\hat{\boldsymbol{\beta}}$

(2) Generate bootstrap domain effects:

$$\upsilon_i^* \overset{iid}{\sim} N(0, \hat{\sigma}_\upsilon^2), \quad i = 1, \dots, m$$

(3) Generate, independently of $\upsilon_1^*, \dots, \upsilon_m^*$, disturbances:

$$e_{ij}^* \overset{iid}{\sim} N(0, \hat{\sigma}_e^2), \quad j = 1, \dots, N_i, \; i = 1, \dots, m$$

(4) Generate a bootstrap population from the model:

$$y_{ij}^* = \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} + \upsilon_i^* + e_{ij}^*, \quad j = 1, \dots, N_i, \; i = 1, \dots, m$$

✓ *González-Manteiga et al. (2008)*

Parametric bootstrap

(5) Calculate target quantities for the bootstrap population

$$F^*_{\alpha i} = \frac{1}{N_i}\sum_{j=1}^{N_i} F^*_{\alpha ij}, \quad F^*_{\alpha ij} = h_\alpha(y^*_{ij}), \quad i = 1, \ldots, m$$

(6) Take the elements y^*_{ij} with indices contained in the sample s: $\mathbf{y}^*_s$. Fit the model to bootstrap sample $\mathbf{y}^*_s$: $\hat{\sigma}^{2*}_{\upsilon}$, $\hat{\sigma}^{2*}_{e}$, $\hat{\boldsymbol{\beta}}^*$

(7) Obtain the bootstrap EBP (through Monte Carlo): $\hat{F}^{EB*}_{\alpha i}$

(8) Repeat (2)–(7) B times: $F^*_{\alpha i}(b)$ true value, $\hat{F}^{EB*}_{\alpha i}(b)$ EBP for bootstrap sample b, $b = 1, \ldots, B$.

(9) Bootstrap estimator:

$$\text{mse}_B(\hat{F}^{EB}_{\alpha i}) = B^{-1}\sum_{b=1}^{B}\left\{\hat{F}^{EB*}_{\alpha i}(b) - F^*_{\alpha i}(b)\right\}^2$$

Application: Estimation of county crop areas

(✓ Battese, Harter & Fuller, 1988)

- $m = 12$ counties in North-central Iowa.
- i county, j area segment.
- county sample sizes n_i =1 to 5.
- y_{ij} number of hectares of corn in j-th segment of i-th county (from farm interview data).
- Auxiliary variables (from LANDSAT satellite data):

  x_{1ij} number of pixels (picture elements of about 0.45 hectares) classified as corn in (ij)-th segment.
  x_{2ij} number of pixels classified as soybeans in (ij)-th segment.
- EB estimates adjusted to agree with survey regression estimate for the entire area covering the 12 counties.

Application: Model validation

- Model with quadratic terms x_{1ij}^2, x_{2ij}^2 included: associated regression coefficients not significant.
- Transformed residuals:

$$r_{ij} = (y_{ij} - \hat{\tau}_i \bar{y}_i) - (\mathbf{x}_{ij} - \hat{\tau}_i \bar{\mathbf{x}}_i)^T \hat{\boldsymbol{\beta}},$$

  where $\hat{\tau}_i = 1 - (1 - \hat{\gamma}_i)^{1/2}$.

- Result: $r_{ij} \overset{iid}{\cong} N(0, \sigma_e^2)$
- Normality assumptions on υ_i and e_{ij}:

  Shapiro-Wilk statistic W applied to r_{ij}:
  $W = 0.985$, p-value: $0.921 \Rightarrow$ No evidence against Normality

Table 3.3: EB estimates with standard errors

County	n_i	$\hat{\bar{y}}_i^{EB}$	s.e.(EB)	s.e.(surv.reg.)	s.e.(EB)/s.e.(surv.reg.)
1	1	122.2	9.6	13.7	0.70
2	1	126.3	9.5	12.9	0.74
3	1	106.2	9.3	12.4	0.75
4	2	108.0	8.1	9.7	0.83
5	3	145.0	6.5	7.1	0.91
6	3	112.6	6.6	7.2	0.92
7	3	112.4	6.6	7.2	0.92
8	3	122.1	6.7	7.3	0.92
9	4	115.8	5.8	6.1	0.95
10	5	124.3	5.3	5.7	0.93
11	5	106.3	5.2	5.5	0.94
12	5	143.6	5.7	6.1	0.93

- As n_i decreases from 5 to 1, s.e.(EB)/s.e.(survey reg.) decreases from 0.95 to 0.7.
- Significant reduction of s.e. when $n_i \leq 3$.

Pseudo-EBLUP of an area mean

- Consider design weights d_{ij} such that

$$\sum_{j \in s_i} d_{ij} = N_i$$

- Normalized weights: $w_{ij} = d_{ij}/N_i$
- Direct estimators of area means and population totals:

$$\bar{y}_{iw} = \sum_{j \in s_i} w_{ij} y_{ij}, \quad \hat{Y}_w = \sum_{i=1}^{m} N_i \bar{y}_{iw}$$

$$\bar{\mathbf{x}}_{iw} = \sum_{j \in s_i} w_{ij} \mathbf{x}_{ij}, \quad \hat{\mathbf{X}}_w = \sum_{i=1}^{m} N_i \bar{\mathbf{x}}_{iw}$$

Pseudo-EBLUP of an area mean

- Pseudo-EBLUP:

$$\bar{y}_{iw}^{PEB} = \hat{\gamma}_{iw} \left\{ \bar{y}_{iw} + (\bar{\mathbf{X}}_i - \bar{\mathbf{x}}_{iw})^T \hat{\boldsymbol{\beta}}_w \right\} + (1 - \hat{\gamma}_{iw}) \bar{\mathbf{X}}_i^T \hat{\boldsymbol{\beta}}_w,$$

where $\hat{\gamma}_{iw} = \hat{\sigma}_v^2 / (\hat{\sigma}_v^2 + \hat{\sigma}_e^2 \sum_{j \in s_i} w_{ij}^2)$.

- σ_v^2 and σ_e^2 estimated from the unit-level data.
- $\hat{\boldsymbol{\beta}}_w$ chosen to ensure automatic benchmarking to the "sample regression estimator" of total Y:

$$\sum_{i=1}^{m} N_i \bar{y}_{iw}^{PEB} = \hat{Y}_w + (\mathbf{X} - \hat{\mathbf{X}}_w)^T \hat{\boldsymbol{\beta}}_w$$

(✓ You and Rao, 2002)

Hierarchical Bayes (HB) approach

- $\boldsymbol{\mu}$ vector of small area parameters of interest, with density, given a vector of unknown model parameters $\boldsymbol{\lambda}_1$: $f(\boldsymbol{\mu}|\boldsymbol{\lambda}_1)$.
- $\mathbf{y}$ observed data, with density, given a vector of unknown parameters $\boldsymbol{\lambda}_2$: $f(\mathbf{y}|\boldsymbol{\lambda}_2)$.
- $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^T, \boldsymbol{\lambda}_2^T)^T$ vector of unknown model parameters, with "prior" density: $f(\boldsymbol{\lambda})$
- Joint posterior density of unknown quantities $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ given observed data $\mathbf{y}$: By Bayes theorem,

$$f(\boldsymbol{\mu}, \boldsymbol{\lambda}|\mathbf{y}) = \frac{f(\mathbf{y}, \boldsymbol{\mu}|\boldsymbol{\lambda}) f(\boldsymbol{\lambda})}{f(\mathbf{y})},$$

where $f(\mathbf{y})$ is the marginal density of $\mathbf{y}$:

$$f(\mathbf{y}) = \int f(\mathbf{y}, \boldsymbol{\mu}|\boldsymbol{\lambda}) f(\boldsymbol{\lambda}) d\boldsymbol{\mu} d\boldsymbol{\lambda}$$

Hierarchical Bayes (HB) approach

- Posterior density of parameters of interest $\boldsymbol{\mu}$ given observed data $\mathbf{y}$:

$$f(\boldsymbol{\mu}|\mathbf{y}) = \int f(\boldsymbol{\mu}, \boldsymbol{\lambda}|\mathbf{y}) d\boldsymbol{\lambda} = \int f(\boldsymbol{\mu}|\boldsymbol{\lambda}, \mathbf{y}) f(\boldsymbol{\lambda}|\mathbf{y}) d\boldsymbol{\lambda}$$

- Hierarchical Bayes (HB) estimator of $\delta = h(\boldsymbol{\mu})$:

$$\hat{\delta}^{HB} = E_{\boldsymbol{\mu}}(\delta|\mathbf{y})$$

- Measure of uncertainty associated with the HB estimator: Posterior variance

$$V_{\boldsymbol{\mu}}(\delta|\mathbf{y})$$

- In practice, $f(\boldsymbol{\mu}|\mathbf{y})$ does not have a closed analytical form. Then HB estimator and posterior variance are obtained numerically, using Markov Chain Monte Carlo (MCMC) methods.

HB Fay–Herriot model: σ_v^2 known

- Observed data: $\mathbf{y} = \hat{\boldsymbol{\theta}} = (\hat{\theta}_1^{DIR}, \ldots, \hat{\theta}_m^{DIR})^T$ with distribution:

$$\hat{\theta}_i^{DIR} | \theta_i, \boldsymbol{\beta} \overset{ind}{\sim} N(\theta_i, \psi_i), \quad i = 1, \ldots, m$$

- Parameters of interest: $\boldsymbol{\mu} = \boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^T$ with distribution:

$$\theta_i | \boldsymbol{\beta} \overset{iid}{\sim} N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_v^2), \quad i = 1, \ldots, m$$

- Model parameters: $\boldsymbol{\lambda} = \boldsymbol{\beta}$ with prior distribution:

$$f(\boldsymbol{\beta}) \propto 1$$

HB estimator: σ_v^2 known

- HB estimator of $\delta = \theta_i$:

$$\tilde{\theta}_i^{HB} = E_{\theta_i}(\theta_i|\hat{\boldsymbol{\theta}}) = \gamma_i \hat{\theta}_i^{DIR} + (1-\gamma_i)\mathbf{x}_i^T \tilde{\boldsymbol{\beta}}(\sigma_v^2) = \tilde{\theta}_i^{BLUP}$$

- Posterior variance:

$$V_{\theta_i}(\theta_i|\hat{\boldsymbol{\theta}}) = g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) = \text{MSE}(\tilde{\theta}_i^{BLUP})$$

- In the EB method, no prior on $\boldsymbol{\beta}$ is assumed. An estimator $\tilde{\boldsymbol{\beta}}$ is obtained from the marginal distribution:

$$\hat{\theta}_i^{DIR} \stackrel{ind}{\sim} N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_v^2 + \psi_i)$$

HB Fay–Herriot model: σ_v^2 unknown

- Observed data: $\mathbf{y} = \hat{\boldsymbol{\theta}} = (\hat{\theta}_1^{DIR}, \ldots, \hat{\theta}_m^{DIR})^T$ with distribution:

$$\hat{\theta}_i^{DIR} | \theta_i, \boldsymbol{\beta} \overset{ind}{\sim} N(\theta_i, \psi_i), \quad i = 1, \ldots, m$$

- Parameters of interest: $\boldsymbol{\mu} = \boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^T$ with distribution:

$$\theta_i | \boldsymbol{\beta}, \sigma_v^2 \overset{iid}{\sim} N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_v^2), \quad i = 1, \ldots, m$$

- Model parameters: $\boldsymbol{\lambda} = (\boldsymbol{\beta}^T, \sigma_v^2)^T$ with prior distribution:

$$f(\boldsymbol{\lambda}) = f(\boldsymbol{\beta}) f(\sigma_v^2) \propto f(\sigma_v^2)$$

HB estimator: σ_v^2 unknown

- HB estimator of θ_i for σ_v^2 unknown:

$$\hat{\boldsymbol{\theta}}_i^{HB} = E_{\theta_i}(\theta_i|\hat{\boldsymbol{\theta}}) = \int \tilde{\theta}_i^{HB}(\sigma_v^2) f(\sigma_v^2|\hat{\boldsymbol{\theta}})\, d\sigma_v^2 = E_{\sigma_v^2}[\tilde{\theta}_i^{HB}(\sigma_v^2)|\hat{\boldsymbol{\theta}}]$$

- Posterior variance of σ_v^2:

$$V_{\theta_i}(\theta_i|\hat{\boldsymbol{\theta}}) = E_{\sigma_v^2}[g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2)|\hat{\boldsymbol{\theta}}] + V_{\sigma_v^2}[\tilde{\theta}_i^{HB}(\sigma_v^2)|\hat{\boldsymbol{\theta}}]$$

- $V_{\theta_i}(\theta_i|\hat{\boldsymbol{\theta}})$ is used as a measure of variability associated with the HB estimator.

Choice of prior density for σ_v^2

- Flat prior:

$$f(\sigma_v^2) \propto 1$$

- Inverse Gamma:

$$f(1/\sigma_v^2) = G(a, b), \quad a > 0, \; b > 0$$

- Prior ensuring that posterior variance $V_{\theta_i}(\theta_i|\hat{\boldsymbol{\theta}})$ is nearly unbiased for the frequentist MSE($\hat{\boldsymbol{\theta}}_i^{HB}$):

$$f_i(\sigma_v^2) \propto (\sigma_v^2 + \psi_i)^2 \sum_{l=1}^{m} (\sigma_v^2 + \psi_l)^{-2}$$

If $\psi_i = \psi$, then $f_i(\sigma_v^2) \propto 1$ (flat prior)

✓ *Datta, Rao and Smith (2002)*

Implementation of HB approach

- When $E_{\theta_i}(\theta_i|\hat{\boldsymbol{\theta}})$ and $V_{\theta_i}(\theta_i|\hat{\boldsymbol{\theta}})$ involve only one dimensional integral, numerical integration can be applied.
- Numerical integration not feasible in complex problems involving high dimensional integration: Use MCMC methods.
- Generate MCMC samples $\{\theta_i^{(\ell)},\ \ell = 1, \ldots, L\}$ from $f(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$.
- Monte Carlo approximation of $E_{\theta_i}(\theta_i|\hat{\boldsymbol{\theta}})$:

$$\hat{\theta}_i^{HB} = E_{\theta_i}(\theta_i|\hat{\boldsymbol{\theta}}) \approx \frac{1}{L}\sum_{l=1}^{L}\theta_i^{(l)}$$

- Monte Carlo approximation of $V_{\theta_i}(\theta_i|\hat{\boldsymbol{\theta}})$:

$$V(\theta_i|\hat{\boldsymbol{\theta}}) \approx \frac{1}{L}\sum_{r=1}^{L}\left\{\theta_i^{(r)} - \frac{1}{L}\sum_{l=1}^{L}\theta_i^{(l)}\right\}^2$$

Time series models

- T time periods observed
- θ_{it} parameter of interest for small area i at time t
- $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{iT})^T$ vector of parameters
- $\hat{\boldsymbol{\theta}}_i^{DIR} = (\hat{\theta}_{i1}^{DIR}, \ldots, \hat{\theta}_{iT}^{DIR})^T$ vector of direct estimators
- $\boldsymbol{\Psi}_i = V(\hat{\boldsymbol{\theta}}_i^{DIR})$ known covariance matrix, $i = 1, \ldots, m$
- Sampling model:

$$\hat{\theta}_{it}^{DIR} = \theta_{it} + e_{it}, \quad (e_{i1}, \ldots, e_{iT})^T \stackrel{ind}{\sim} (\mathbf{0}, \boldsymbol{\Psi}_i), \quad \boldsymbol{\Psi}_i \text{ known}$$

Times series models

- Linking model:

$$\theta_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \upsilon_i + u_{it}, \quad \upsilon_i \stackrel{iid}{\sim} (0, \sigma_\upsilon^2)$$

(a) AR(1) model:

$$u_{it} = \rho u_{i,t-1} + \varepsilon_{it}, \;\; |\rho| < 1, \quad \varepsilon_{it} \stackrel{iid}{\sim} (0, \sigma_\epsilon^2)$$

✓ *Rao and Yu (1992, 1994)*

(b) Random walk model:

$$u_{it} = u_{i,t-1} + \varepsilon_{it}, \quad \varepsilon_{it} \stackrel{iid}{\sim} (0, \sigma_\epsilon^2)$$

✓ *Datta, Lahiri and Maiti (2002)*

Application

(✓ Datta, Lahiri and Maiti, 2002)

- Target: Estimate median income of four-person families in year 1989 for the 50 US states and the District of Columbia:
- Random walk model.
- Data: Direct estimates for years 1981–1989 obtained from CPS ($T = 9$).
- $\mathbf{x}_{it} = (1, x_{it})^T$ where x_{it} is the 1979 census estimate, adjusted by the proportional growth in per capita income.
- Evaluation: 1989 estimates obtained from 1990 census taken as true values.

Table 4.2: Distribution of CV (%)

	CV		
Estimator	$2-4\%$	$4-6\%$	$\geq 6\%$
CPS	6	7	38
HB	10	37	4
EB	49	2	0

- Both EB, HB better than CPS estimate.
- EB performs better than HB in terms of CV.

Disease Mapping

- y_i small area count
- n_i number exposed in area i
- λ_i true incidence rate
- Counts distribution:

$$y_i|\lambda_i \overset{ind}{\sim} \text{Pois}(n_i\lambda_i), \quad i = 1, \ldots, m$$

- **Model 1:** $\lambda_i \overset{ind}{\sim} \text{Gamma}(a, b)$, $a > 0$, $b > 0$
- **Model 2:** $\beta_i = \log \lambda_i \overset{ind}{\sim} N(\mu, \sigma^2)$
- **Model 3:** Spatial dependence for β_i's through a Conditional Autoregression (CAR) model: Relates each β_i to a set of neighbourhood areas of area i

Disease Mapping

Application: HB approach to lip cancer incidence in Scotland

- Estimation of lip cancer incidence for each of 56 counties in Scotland.
- Data from registered cases in years 1975–1980.
- HB approach with Models 1–3.
- HB estimates similar under Models 2 and 3 but standard errors smaller under Model 3.

Logistic linear mixed models

- Binary target variable: $y_{ij} \in \{0, 1\}$
- $\theta_{ij} := P(y_{ij} = 1)$ true probability for unit j in area i
- Target parameters: small area proportions

$$P_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}, \quad i = 1, \ldots, m.$$

- Logistic model with random area effects:

$$y_{ij}|\theta_{ij} \overset{ind}{\sim} \text{Bernoulli}(\theta_{ij})$$
$$\log\{\theta_{ij}/(1-\theta_{ij})\} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \upsilon_i; \quad \upsilon_i \overset{iid}{\sim} N(0, \sigma_\upsilon^2)$$

- EB estimators of P_i obtained by Monte Carlo approximation, and associated standard errors obtained by jackknife.

✓ *Jiang, Lahiri & Wan (1999)*

Recommendations:

(a) Preventive measures (design issues) may reduce the need for indirect estimates significantly.

(b) Good auxiliary information related to variables of interest plays vital role in model-based estimation. Expanded access to auxiliary information through coordination and cooperation among federal agencies needed.

(c) Internal evaluation: Model validation plays important role. More work on model diagnostics needed. External evaluation studies are also needed.

(d) Area-level models have wider scope because area-level auxiliary information more readily available. But assumption of known sampling variances is restrictive. More work on getting good approximations to sampling variances is needed.

Recommendations:

(e) HB approach is powerful, but caution should be exercised in the choice of improper priors on model parameters. Practical issues in implementing MCMC need to be addressed.
(✓ Rao, 2003, Section 10.2.4)

(f) Model-based estimates of totals and means not suitable if objective is to identify areas with extreme population values or to rank areas or to identify areas that fall below or above some pre-specified level. *(✓ Rao, 2003, Section 9.6)*

(g) Model-based estimates should be distinguished clearly from traditional area-specific direct estimates. Errors in small area estimates may be more transparent to users than errors in large area estimates.

Recommendations:

(h) Proper criterion for assessing quality of model-based estimates is whether they are sufficiently accurate for the intended uses. Even if they are better than direct estimates, they may not be sufficiently accurate to be acceptable.

(i) Overall program should be developed that covers issues related to sample design and data development, organization and dissemination, in addition to those pertaining to methods of estimation for small areas.

References

- 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.
- Chatterjee, S., Lahiri, P. and Li, H. (2008). Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models, *Annals of Statistics*, **36**, 1221–1245.
- Datta, G.S., Rao, J.N.K. and Smith, D.D. (2005). On measuring the variability of small area estimators under a basic area level model, *Biometrika*, **92**, 183–196.
- Datta, G.S., Lahiri, P. and Maiti, T. (2002). Empirical Bayes Estimation of Median Income of Four-Person Families by State Using Time Series and Cross-Sectional Data, *Journal of Statistical Planning and Inference*, **102**, 83–97.

References

- Dick, P. (1995). Modelling Net Undercoverage in the 1991 Canadian Census, *Survey Methodology*, **21**, 45–54.
- Fay, R.E. and Herriot, R.A. (1979). Estimation of Income for Small Places: An Application of James-Stein Procedures to Census Data, *Journal of the American Statistical Association*, **74**, 269–277.
- González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D. and Santamaría, L. (2008). Bootstrap Mean Squared Error of a Small-Area EBLUP, *Journal of Statistical Computation and Simulation*, **75**, 443–462.
- Jiang, J., Lahiri, P. and Wan, S.-M. (2002). A Unified Jackknife Theory, *Annals of Statistics*, **30**, 1782–1810.
- National Research Council (2000). Small Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology, C.F. Citro and G. Kalton (eds.), Committee on National Statistics, Washington DC: National Academy Press.

References

- Prasad, N.G.N. and Rao, J.N.K. (1990). The Estimation of the Mean Squared Error of Small-Area Estimators, *Journal of the American Statistical Association*, **85**, 163–171.
- Rao, J.N.K. (2003). Small Area Estimation, Hoboken, New Jersey: Wiley.
- Rao, J.N.K. and Yu, M. (1992). Small Area Estimation by Combining Time Series and Cross-sectional Data, *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 1–9.
- Rao, J.N.K. and Yu, M. (1994). Small Area Estimation by Combining Time Series and Cross-sectional Data, *Canadian Journal of Statistics*, **22**, 511–528.
- Royall, R.M. (1970). On Finite Population Sampling Theory Under Certain Linear Regression, *Biometrika*, **57**, 377–387.

References

- Scott, A.J. and Smith, T.M.F. (1969). Estimation in Multi-stage Surveys, *Journal of the American Statistical Association*, **71**, 657–664.
- You, Y. and Rao, J.N.K. (2002). A Pseudo-Empirical Best Linear Unbiased Prediction Approach to Small Area Estimation Using Survey Weights, *Canadian Journal of Statistics*, **30**, 431–439.