SMALL AREA ESTIMATION

Isabel Molina and J. Miguel Marín

Dept. of Statistics, Univ. Carlos III de Madrid

J.N.K. Rao

School of Mathematics and Statistics, Carleton University

DESIGN VERSUS MODEL BASED INFERENCE

ELEMENT	UNDER A MODEL	UNDER THE DESIGN
Population	$y \sim P_\theta$	$U = \{1, \ldots, N\}$, $\mathcal{Y} = \{y_1, \ldots, y_N\}$
Sample	$\mathbf{y} = (y_1, \ldots, y_n)$ y_i i.i.d. as y	$s = (i_1, \ldots, i_n) \in S_\pi$, $\mathbf{y} = (y_{i_1}, \ldots, y_{i_n})$
Probab. distrib.	$P_\theta(\mathbf{y})$	$P_\pi(s)$
Parameter	θ (e.g., $\theta = E_{P_\theta}(\mathcal{Y})$)	$\theta = h(y_1, \ldots, y_N)$
Estimator	$\hat{\theta}(\mathbf{y})$	$\hat{\theta}(s)$

Sampling design: (S_π, P_π), $S_\pi \subset \mathcal{P}(U)$ set of samples, P_π probability distribution over S_π satisfying $P_\pi(s) > 0$, $\forall s \in S_\pi$, and all units $j \in U$ are contained in some sample $s \in S_\pi$.

DESIGN-BASED INFERENCE

- U **finite** population of size N.
- $y_1, \ldots, y_N$ measurements at the population units (fixed).
- **Target quantity:**

$$\delta = h(y_1, \ldots, y_N).$$

- Example: population **mean**

$$\bar{Y} = \frac{1}{N} \sum_{j=1}^{N} y_j.$$

- s random **sample** of size n drawn from the population U according to a given design.
- $r = U - s$ **non-sample** (size $N - n$).

HORVITZ-THOMPSON ESTIMATOR

- π_j probability of inclusion of unit j in the sample.
- $d_j = 1/\pi_j$ sampling weight for unit j.
- **Horvitz-Thompson (HT)** estimator of mean:

$$\hat{\bar{Y}} = \frac{1}{N}\sum_{j\in s}\frac{y_j}{\pi_j} = \frac{1}{N}\sum_{j\in s} d_j y_j.$$

- **Design variance**:

$$V_\pi(\hat{\bar{Y}}) = \frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N}(\pi_{j,k} - \pi_j\pi_k)\frac{y_j}{\pi_j}\frac{y_k}{\pi_k},$$

$\pi_{j,k}$ joint inclusion probability for units j and k.

✓ *Hansen, Hurwitz & Madow (1953)*

DESIGN VARIANCE

- **Design-unbiased variance estimator**:

$$\hat{V}_\pi(\hat{\bar{Y}}) = \frac{1}{N^2} \sum_{j \in s} \sum_{j \in s} \frac{\pi_{j,k} - \pi_j \pi_k}{\pi_{j,k}} \frac{y_j}{\pi_j} \frac{y_k}{\pi_k},$$

✓ *Särndal, Swensson and Wretman (1992), equation (5.8.5)*

- Under the approximation $\pi_{j,k} \cong \pi_j \pi_k$, $j \neq k$,

$$\hat{V}_\pi(\hat{\bar{Y}}) \cong \frac{1}{N^2} \sum_{j \in s} \left(\frac{1 - \pi_j}{\pi_j^2} \right) y_j^2 = \sum_{j \in s} d_j (d_j - 1) y_j^2.$$

EXAMPLE: FGT POVERTY INDICATORS

- E_j welfare measure for individual j: for instance, equivalized annual net income.
- z poverty line: Spanish Statistical Institute (INE) uses:

$$z = 0.6 \times \text{Median}(E_j).$$

- FGT family of poverty indicators

$$F_\alpha = \frac{1}{N} \sum_{j=1}^{N} \left(\frac{z - E_j}{z} \right)^\alpha I(E_j < z), \quad \alpha \geq 0.$$

- $\alpha = 0 \Rightarrow$ Poverty incidence
- $\alpha = 1 \Rightarrow$ Poverty gap

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

HORVITZ-THOMPSON ESTIMATOR

- Poverty indicator:

$$F_\alpha = \frac{1}{N}\sum_{j=1}^{N} F_{\alpha j}, \quad F_{\alpha j} = \left(\frac{z - E_j}{z}\right)^\alpha I(E_j < z).$$

- HT estimator of F_α:

$$\hat{F}_\alpha = \frac{1}{N}\sum_{j\in s} d_j F_{\alpha j}.$$

- Variance estimator:

$$\hat{V}_\pi(\hat{F}_\alpha) = \frac{1}{N^2}\sum_{j\in s} d_j(d_j - 1)F_{\alpha j}^2.$$

ADJUSTMENTS TO HT ESTIMATOR

- g_j adjustment factor for design weight d_j, $j \in s$.
- $w_j = d_j g_j$ adjusted weight, $j \in s$.
- Adjusted HT estimator:

$$\hat{\bar{Y}}^A = \frac{1}{N} \sum_{j \in s} w_j y_j.$$

EXAMPLE 1: RATIO HT ESTIMATOR

- HT estimator of the population size:

$$\hat{N} = \sum_{j \in s} d_j.$$

- Constant adjustment factor:

$$g_j = \frac{N}{\hat{N}}, \quad \forall j \in s.$$

- **Ratio HT estimator:**

$$\hat{\bar{Y}}^R = \frac{\hat{Y}}{\hat{N}}.$$

EXAMPLE 2: RATIO EST. WITH AUX. VARIABLE

- X known total of an auxiliary variable with population values:
$$x_1, \ldots, x_N.$$
- HT estimator of X:
$$\hat{X} = \sum_{j \in s} d_j x_j.$$
- Adjustment factor:
$$g_j = \frac{X}{\hat{X}}, \quad \forall j \in s.$$
- Ratio estimator with auxiliary variable X:
$$\hat{\hat{Y}}^{RX} = \hat{\hat{Y}} \frac{X}{\hat{X}}.$$
- The Ratio HT estimator is obtained taking $x_j = 1$, $\forall j \in U$.

EXAMPLE 3: CALIBRATION ESTIMATOR

- p auxiliary variables with known population totals $X_k,\ k = 1, \ldots, p$.
- **Idea:** Find weights w_j, $j \in s$, which minimize the $\mathcal{X}^2$ distance

$$\begin{aligned} \text{mín} \quad & \sum_{j \in s} \frac{\left(w_j - d_j\right)^2}{d_j} \\ s.t. \quad & \sum_{j \in s} w_j x_{jk} = X_k,\ k = 1, \ldots, p. \end{aligned}$$

- **Solution:** $w_j = d_j g_j$, where $g_j = 1 + \mathbf{x}_j' \hat{\mathbf{T}}^{-1}(\mathbf{X} - \hat{\mathbf{X}})$,

$$\mathbf{x}_j = (x_{j1}, \ldots, x_{jp})', \quad \mathbf{X} = (X_1, \ldots, X_p)', \quad \hat{\mathbf{T}} = \sum_{j \in s} d_j \mathbf{x}_j \mathbf{x}_j'.$$

✓ *Deville and Särndal (1992)*

EXAMPLE 3: CALIBRATION ESTIMATOR

- Linear regression model:

$$y_j = \mathbf{x}_j'\boldsymbol{\beta} + e_j, \quad E(e_j) = 0,\ E(e_j^2) = \sigma_e^2,\ j = 1, \ldots, N.$$

- Regression estimator:

$$\hat{\mathbf{B}} = \hat{\mathbf{T}}^{-1} \sum_{j \in s} d_j \mathbf{x}_j y_j$$

- Generalized regression (GREG) estimator:

$$\hat{Y}^A = \hat{Y} + (\mathbf{X} - \hat{\mathbf{X}})'\hat{\mathbf{B}}.$$

- It coincides with calibration estimator!

✓ *Deville and Särndal (1992)*

DOMAIN/AREA ESTIMATION

- U partitioned into D domains $U_1, \ldots, U_D$ of sizes $N_1, \ldots, N_D$.
- s_d sample of size n_d drawn from U_d.
- Total sample size $n = \sum_{d=1}^{D} n_d$.
- $r_d = U_d - s_d$ sample complement, of size $N_d - n_d$.

Example: Survey on Income and Living Conditions 2006

Total sample size: $n = 34,389$ persons.
Summary province×gender sample sizes:

(Barcelona,F)	(Córdoba,F)	(Tarragona,M)	(Soria,F)
1483	230	129	17

TRADITIONAL DIRECT ESTIMATORS

- Target quantity:

$$\delta_d = h_d(\{y_j; j \in U_d\}).$$

- Example: mean of d-th domain

$$\bar{Y}_d = \frac{1}{N_d} \sum_{j \in U_d} y_j.$$

- **Direct estimator:** Uses only area-specific sample data.
- Example: HT estimator of $\bar{Y}_d$,

$$\hat{\bar{Y}}_d^{DIR} = \frac{1}{N_d} \sum_{j \in s_d} d_j y_j.$$

- Variance estimator: Under $\pi_{j,k} \cong \pi_j \pi_k$, $j \neq k$,

$$\hat{V}_\pi(\hat{\bar{Y}}_d^{DIR}) = \frac{1}{N_d} \sum_{j \in s_d} d_j(d_j - 1) y_j^2.$$

DIRECT ESTIMATORS

TARGET INDICATORS:

- Additive in the individual observations.

DATA REQUIREMENTS:

- Final design weights d_j, $j \in s_d$ of sample units in the area.
- For the HT estimator of the domain mean and for the Hájek estimator of the total, domain population count N_d.

DIRECT ESTIMATORS

ADVANTAGES:

- **No model** assumptions (nonparametric).
- Sampling weights can be used $\Rightarrow$ Approx. **design-unbiased** and design-consistent as n_d increases.
- Additivity (**Benchmarking** property):

$$\sum_{d=1}^{D} \hat{Y}_d^{DIR} = \hat{Y}^{DIR}.$$

DISADVANTAGES:

- $V_\pi(\hat{\bar{Y}}_d^{DIR}) \uparrow$ as $n_d \downarrow$. Very **inefficient** for small domains.
- They cannot be calculated for non-sampled areas ($n_d = 0$).

LIMITS OF DISAGGREGATION OF DIRECT ESTIMATORS

RECOMMENDATIONS:

(i) Use direct estimators at the national level and for disaggregates with CV under a specified limit for all the areas.

(ii) For further disaggregations, use indirect estimators in the areas with relative absolute bias below a given limit.

(iii) For areas where indirect estimators exceed the bias limit, do not produce estimates. It is always possible to modify the survey sample size allocation so as to have a minimum number of observations in each area.

INDIRECT ESTIMATORS

- **Indirect estimator:** It **borrows strength** from other areas by making some kind of homogeneity assumption across areas (model with **common** parameters).

FIRST APPLICATION OF SYNTHETIC REGRESSION

1945 Radio Listening Survey:

- Target: to estimate the median num. of radio stations heard during the day in 500 U.S. counties.
- Mail survey: From each of 500 counties, 1000 families sampled and sent mailed questionnaire. Response rate only 20 % and incomplete coverage.
- x_d median no. of stations heard during day (mail survey) in the d-th county, for $d = 1, \ldots, 500$. Biased due to nonresponse and incomplete coverage.
- Intensive interview survey of 85 counties: Probability sample of 85 counties subsampled and subject to personal interviews.

✓ *Hansen, Hurwitz & Madow, 1953, p. 483;* ✓ *Rao, 2003*

FIRST APPLICATION OF SYNTHETIC REGRESSION

1945 Radio Listening Survey:

- y_d median no. of stations heard during day (interview) in the d-th sample county, for $d = 1, \ldots, 85$. Considered as true county medians.
- $\mathrm{corr}(y, x) = 0{,}70$
- Linear Regression:

$$y_d = \beta_0 + \beta_1 x_d + e_d, \quad d = 1, \ldots, 85.$$

- Indirect estimators for the 500-85 non-sampled counties:

$$\hat{y}_d^{SYN} = 0{,}52 + 0{,}74 x_d. \quad \text{(Regression synthetic estimators)}$$

- It does not account for between-county heterogeneity.

SYNTHETIC ESTIMATORS

Definition:

An unbiased estimator is obtained from a sample survey for a large area; when this estimate is used to derive estimates for subareas under the assumption that the small areas have the same characteristics as the large area, we identify these estimates as synthetic estimates. ✓ *González (1973)*

SIMPLE EXAMPLE:

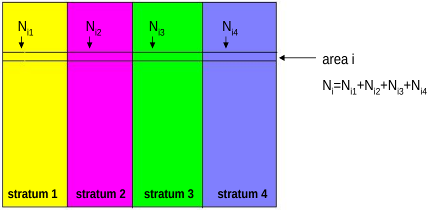
- Target: $\bar{Y}_d$ mean of domain d.
- Assumption: $\bar{Y}_d = \bar{Y}$.
- Synthetic estimator of $\bar{Y}_d$:

$$\hat{\bar{Y}}_d^{SYNT} = \hat{\bar{Y}}.$$

POST-STRATIFIED SYNTHETIC ESTIMATOR

- J post-strata ($j = 1, \ldots, J$) cut across the domains.
- N_{dj} known count in the intersection of domain d and post-stratum j.
- Total of domain d:

$$Y_d = \sum_{j=1}^{J} N_{dj} \bar{Y}_{dj}$$



- Assumption (implicit model):

$$\bar{Y}_{dj} = \bar{Y}_{+j} = Y_{+j}/N_{+j}, \ \forall d, j$$

POST-STRATIFIED SYNTHETIC ESTIMATOR

- Post-stratified synthetic estimator:

$$\hat{Y}_d^{SYN} = \sum_{j=1}^{J} N_{dj} \hat{\bar{Y}}_{+j}^{R}, \quad \hat{\bar{Y}}_{+j}^{R} = \hat{Y}_{+j}/\hat{N}_{+j}.$$

- $\hat{Y}_{+j}, \hat{N}_{+j}$ reliable direct estimators of Y_{+j}, N_{+j}.
- Need homogeneity within each post-stratum.
- Special case: When $y \in \{0, 1\}$, domain proportion P_d is Y_d/N_d, where $N_d = \sum_{j=1}^{J} N_{dj}$.
- Synthetic estimator of P_d:

$$\hat{P}_d^{SYN} = \frac{1}{N_d} \sum_{j=1}^{J} N_{dj} \hat{P}_{+j}^{R}$$

MSE OF SYNTHETIC ESTIMATOR

- Synthetic estimator $\hat{Y}_d^{SYN}$ depends on direct estimators $\hat{Y}_{+j}/\hat{N}_{+j}$ for post-stratum j. Hence, design variance of synthetic estimators small in comparison with that of the direct estimator for a small domain.
- But synthetic estimators depend on strong assumptions and hence may be biased when the assumptions are not true.
- Hence, full MSE (accounting for bias and variance) needs to be estimated.

MSE ESTIMATOR

- Approximate MSE:

$$\mathrm{MSE}_d(\hat{Y}_d^{SYN}) \approx E_d(\hat{Y}_d^{SYN} - \hat{Y}_d^{DIR})^2 - \hat{V}_d(\hat{Y}_d^{DIR})$$

- Estimated MSE:

$$\hat{\mathrm{MSE}}_d(\hat{Y}_d^{SYN}) = (\hat{Y}_d^{SYN} - \hat{Y}_d^{DIR})^2 - \hat{V}_d(\hat{Y}_d^{DIR}).$$

- $\hat{\mathrm{MSE}}_d(\hat{Y}_d^{SYN})$ is approximately unbiased but unstable.
- Average over domains: *(✓ González and Wakesberg, 1973)*

$$\hat{\mathrm{MSE}}_a(\hat{\hat{Y}}_d^{SYN}) = \frac{1}{m}\sum_{\ell=1}^{D}\frac{1}{N_\ell^2}(\hat{Y}_\ell^{SYN} - \hat{Y}_\ell^{DIR})^2 - \frac{1}{m}\sum_{\ell=1}^{D}\frac{1}{N_\ell^2}\hat{V}_d(\hat{Y}_\ell^{DIR})$$

- Limitation: $\hat{\mathrm{MSE}}_a(\hat{\hat{Y}}_d^{SYN})$ is stable but not area-specific.

SYNTHETIC ESTIMATORS

TARGET INDICATORS:

- For regression-synthetic estimator, general indicators. For post-stratified synthetic, additive parameters.

DATA REQUIREMENTS:

- For regression-synthetic estimator, aggregated values of p auxiliary variables at the domain level.
- For post-stratified synthetic estimators, post-stratum indicator in the survey and popn. sizes of crossings of poststrata and domains.

SYNTHETIC ESTIMATORS

ADVANTAGES:

- They can have pretty small variance.
- They allow us to estimate in non-sampled areas.

SYNTHETIC ESTIMATORS

DISADVANTAGES:

- They do not account for between-area heterogeneity and can thus be seriously design-biased.
- The model needs to be consciously checked (e.g. by residual plots and significance of area effect).
- If the model is known, the survey data on the target variable is not be used!
- They do not tend to the direct estimator as the domain sample size increases.
- Stable and area-specific design MSE estimators are not available.
- Benchmarking adjustment is required.

COMPOSITE ESTIMATORS

To balance the bias of a synthetic estimator and the instability of a direct estimator for a domain, take:

$$\hat{Y}_d^C = \phi_d \hat{Y}_d + (1-\phi_d)\hat{Y}_d^{SYN}, \quad 0 \leq \phi_d \leq 1.$$

- **Sample-size dependent estimator:** For a given $\delta > 0$,

$$\phi_d = \begin{cases} 1, & \text{if} \quad \hat{N}_d \geq \delta N_d; \\ \hat{N}_d/(\delta N_d), & \text{if} \quad \hat{N}_d < \delta N_d. \end{cases}$$

✓ *Drew, Singh and Choudhry (1982), SM*

SAMPLE-SIZE DEPENDENT (SSD) ESTIMATOR

- Under SRS in the population:

$$\hat{N}_d = \sum_{j \in s_d} d_j = N n_d / n$$

- $\hat{N}_d$ unbiased: $N_d = E_\pi(\hat{N}_d) = N E_\pi(n_d)/n$. Then,

$$\hat{N}_d \geq \delta N_d \Leftrightarrow N n_d / n \geq \delta N E(n_d)/n \Leftrightarrow n_d \geq \delta E(n_d).$$

- Weight of SSD estimator under SRS:

$$\phi_d = \begin{cases} 1 & \text{if} \quad n_d \geq \delta E_d(n_d); \\ n_d / \{\delta E_d(n_d)\} & \text{if} \quad n_d < \delta E_d(n_d) \end{cases}$$

SAMPLE-SIZE DEPENDENT (SSD) ESTIMATOR

- Canadian LFS: Estimates produced for Census Divisions with $\delta = 2/3$. For most areas, $1 - \phi_d = 0$; for other areas weight attached to $\hat{Y}_d^{SYN}$ is about 0.1 but never larger than 0.2.
- All variables y use the same weight ϕ_d regardless of the differences with respect to between-area homogeneity.

OPTIMAL COMPOSITE ESTIMATOR

- Find ϕ_d that minimizes $\mathrm{MSE}_d(\hat{Y}_d^C) \Rightarrow \phi_d^*$
- Optimal weight depends on true MSEs of $\hat{Y}_d^{SYN}$ and $\hat{Y}_d$.
- Estimated optimal weight:

$$\hat{\phi}_d^* = \mathrm{MSE}_d(\hat{Y}_d^{SYN})/(\hat{Y}_d^{SYN} - \hat{Y}_d)^2$$

- Limitation: $\hat{\phi}_d^*$ is unstable.
- Estimated optimal *common weight (aggregated over areas):*

$$\begin{aligned}\hat{\phi}^* &= \sum_{\ell=1}^{D} MSE_d(\hat{\hat{Y}}_\ell^{SYN}) / \sum_{\ell=1}^{D} (\hat{\hat{Y}}_\ell^{SYN} - \hat{\hat{Y}}_\ell)^2 \\ &= 1 - \left\{ \sum_{\ell=1}^{D} \hat{V}_d(\hat{\hat{Y}}_\ell) / \sum_{\ell=1}^{D} (\hat{\hat{Y}}_\ell^{SYN} - \hat{\hat{Y}}_\ell)^2 \right\}\end{aligned}$$

- $\hat{\phi}_d^*$ *is stable but it is not area-specific.*

BENCHMARKING

- Usually a reliable direct estimator for an aggregate A of areas $\hat{Y}_A^{DIR}$ is available.
- Indirect estimators of area totals Y_d do not necessarily add up to $\hat{Y}_A^{DIR}$.
- Ratio adjustment: $\tilde{Y}_d$ indirect estimator of Y_d with $\sum_{d\in A}\tilde{Y}_d \neq \hat{Y}_A^{DIR}$. Then, take the estimator

$$\tilde{Y}_d^* = \tilde{Y}_d \frac{\hat{Y}_A^{DIR}}{\sum\limits_{d\in A}\tilde{Y}_d} \Rightarrow \sum_{d\in A}\tilde{Y}_d^* = \hat{Y}_A^{DIR}$$

SSD ESTIMATORS

TARGET INDICATORS:

- Additive parameters.

DATA REQUIREMENTS:

- Same as required for the considered direct and the synthetic ones.

SSD ESTIMATORS

ADVANTAGES:

- They cannot have worse design-variance than the direct estimator or worse bias of the synthetic one.

DISADVANTAGES:

- If the domain sample size (even if small) is not smaller than the expected sample size, it does not borrow strength.
- The weight of the synthetic estimator does not depend on how well the auxiliary variables explain the variability of the target one.
- They cannot be computed for non-sampled domains.
- Stable and area-specific design MSE estimators are not available.
- Benchmarking adjustment is required.

FAY-HERRIOT MODEL

(i) Linking model:

$$\delta_d = \mathbf{x}_d'\boldsymbol{\beta} + u_d, \quad d = 1, \ldots, D$$
$$u_d \overset{iid}{\sim} (0, \sigma_u^2), \quad \sigma_u^2 \text{ unknown}$$

(ii) Sampling model:

$$\hat{\delta}_d^{DIR} = \delta_d + e_d, \quad d = 1, \ldots, D$$
$$e_d \overset{ind}{\sim} (0, \psi_d), \quad \psi_d = V_\pi(\hat{\delta}_d^{DIR} | \delta_d) \text{ known } \forall d$$
$$u_d \text{ and } e_d \text{ indep.}$$

(iii) Combined model: Linear mixed model

$$\hat{\delta}_d^{DIR} = \mathbf{x}_d'\boldsymbol{\beta} + u_d + e_d, \quad d = 1, \ldots, D$$

✓ *Fay & Herriot (1979), JASA*

BLUP UNDER FAY-HERRIOT MODEL

Best linear unbiased predictor (BLUP)

Under the combined model (iii) with $\delta_d = \mathbf{x}_d'\boldsymbol{\beta} + u_d$, the linear estimator $\tilde{\delta}_d = \alpha_1\hat{\delta}_1^{DIR} + \cdots + \alpha_D\hat{\delta}_D^{DIR}$ that solves the problem:

$$\begin{array}{ll} \min_{(\alpha_1,\ldots,\alpha_D)} & \mathrm{MSE}(\tilde{\delta}_d) = E(\tilde{\delta}_d - \delta_d)^2 \\ \text{s.t.} & E(\tilde{\delta}_d - \delta_d) = 0 \end{array}$$

is given by

$$\tilde{\delta}_d^{BLUP} = \mathbf{x}_d'\tilde{\boldsymbol{\beta}} + \tilde{u}_d,$$

where

$$\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\sigma_u^2) = \left(\sum_{d=1}^{D}\gamma_d\mathbf{x}_d\mathbf{x}_d'\right)^{-1}\sum_{d=1}^{D}\gamma_d\mathbf{x}_d\hat{\delta}_d^{DIR},$$

$$\tilde{u}_d = \tilde{u}_d(\sigma_u^2) = \gamma_d(\hat{\delta}_d^{DIR} - \mathbf{x}_d'\tilde{\boldsymbol{\beta}}), \quad \gamma_d = \frac{\sigma_u^2}{\sigma_u^2 + \psi_d}$$

BLUP UNDER FAY-HERRIOT MODEL

Proof: We prove a more general result. Let us express model (iii) in matrix notation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} + \mathbf{e},$$

where

$$\mathbf{y} = \begin{pmatrix} \hat{\delta}_1^{DIR} \\ \vdots \\ \hat{\delta}_D^{DIR} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_D' \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_D \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_D \end{pmatrix}.$$

Covariance matrices: $V(\mathbf{u}) = \sigma_u^2 \mathbf{I}_D$, $V(\mathbf{e}) = \text{diag}(\psi_d)$.
We prove that the BLUP of a mixed effect

$$\mu = \boldsymbol{\ell}'\boldsymbol{\beta} + \mathbf{m}'\mathbf{u},$$

for given $p \times 1$ and $D \times 1$ vectors $\boldsymbol{\ell}$ and $\mathbf{m}$, is

$$\tilde{\mu} = \boldsymbol{\ell}'\tilde{\boldsymbol{\beta}} + \mathbf{m}'\tilde{\mathbf{u}}, \quad \tilde{\mathbf{u}} = (\tilde{u}_1, \ldots, \tilde{u}_D)'.$$

BLUP UNDER FAY-HERRIOT MODEL

- **Linear** predictor of $\mu = \boldsymbol{\ell}'\boldsymbol{\beta} + \mathbf{m}'\mathbf{u}$:

$$\tilde{\mu} = \boldsymbol{\alpha}'\mathbf{y} + b,$$

for a given vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_D)'$ and scalar b.

- Prediction error:

$$\tilde{\mu} - \mu = \boldsymbol{\alpha}'\mathbf{y} + b - \boldsymbol{\ell}'\boldsymbol{\beta} - \mathbf{m}'\mathbf{u} = \boldsymbol{\alpha}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha}'\mathbf{u} + \boldsymbol{\alpha}'\mathbf{e} + b - \boldsymbol{\ell}'\boldsymbol{\beta} - \mathbf{m}'\mathbf{u}.$$

- $\tilde{\mu}$ **model-unbiased** for μ iif $E(\tilde{\mu} - \mu) = 0$.
- Taking expected value of the prediction error,

$$E(\tilde{\mu} - \mu) = (\boldsymbol{\alpha}'\mathbf{X} - \boldsymbol{\ell}')\boldsymbol{\beta} + b = 0 \ \forall \boldsymbol{\beta} \Leftrightarrow \boldsymbol{\alpha}'\mathbf{X} = \boldsymbol{\ell}', \ \ b = 0.$$

- If $\tilde{\mu}$ is unbiased for μ, then

$$\mathrm{MSE}(\tilde{\mu}) = V(\tilde{\mu} - \mu) = V(\boldsymbol{\alpha}'\mathbf{y} - \mathbf{m}'\mathbf{u}) = \boldsymbol{\alpha}'\mathbf{V}\boldsymbol{\alpha} + \sigma_u^2\mathbf{m}'\mathbf{m} - 2\sigma_u^2\boldsymbol{\alpha}'\mathbf{m},$$

where $\mathbf{V} = V(\mathbf{y}) = \sigma_u^2\mathbf{I}_D + \mathrm{diag}(\psi_d)$.

BLUP UNDER FAY-HERRIOT MODEL

- Minimization problem:

$$\begin{array}{ll} \min_{\boldsymbol{\alpha}} & \mathrm{MSE}(\tilde{\mu}) = \boldsymbol{\alpha}'\mathbf{V}\boldsymbol{\alpha} + \sigma_u^2\mathbf{m}'\mathbf{m} - 2\sigma_u^2\boldsymbol{\alpha}'\mathbf{m} \\ \text{s.t.} & \boldsymbol{\alpha}'\mathbf{X} = \boldsymbol{\ell}' \end{array}$$

- Solve by **Lagrange multiplier** method, to obtain:

$$\boldsymbol{\alpha}' = \boldsymbol{\ell}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} + \sigma_u^2\mathbf{m}'\mathbf{V}^{-1}\left[\mathbf{I}_D - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\right].$$

- Then, the BLUP of μ is

$$\tilde{\mu}^{BLUP} = \boldsymbol{\alpha}'\mathbf{y} = \boldsymbol{\ell}'\tilde{\boldsymbol{\beta}} + \mathbf{m}'\underbrace{\sigma_u^2\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})}_{\tilde{\mathbf{u}}} = \boldsymbol{\ell}'\tilde{\boldsymbol{\beta}} + \mathbf{m}'\tilde{\mathbf{u}}.$$

- For $\boldsymbol{\ell} = \mathbf{x}_d$ and $\mathbf{m} = (\mathbf{0}'_{d-1}, 1, \mathbf{0}'_{D-d})'$, we obtain

$$\tilde{\delta}_d^{BLUP} = \mathbf{x}'_d\tilde{\boldsymbol{\beta}} + \tilde{u}_d.$$

GOOD PROPERTY OF THE BLUP

- BLUP can be expressed as

$$\tilde{\delta}_d^{BLUP} = \gamma_d \, \hat{\delta}_d^{DIR} + (1-\gamma_d)\mathbf{x}_d'\tilde{\boldsymbol{\beta}}, \quad \gamma_d = \frac{\sigma_u^2}{\sigma_u^2 + \psi_d}.$$

- Weighted combination of direct estimator $\hat{\delta}_d^{DIR}$ and "regression synthetic" estimator $\mathbf{x}_d'\tilde{\boldsymbol{\beta}}$.
- It gives **more weight** to $\hat{\delta}_d^{DIR}$ when sampling variance ψ_d small ($\hat{\delta}_d^{DIR}$ **reliable**).
- It gives **more weight** to the **synthetic** estimator $\mathbf{x}_d'\tilde{\boldsymbol{\beta}}$ when ψ_d large ($\hat{\delta}_d^{DIR}$ unreliable) or σ_u^2 small ($\mathbf{x}_d'\tilde{\boldsymbol{\beta}}$ **reliable**).

EMPIRICAL BLUP (EBLUP)

- $\tilde{\delta}_d^{BLUP}$ depends on unknown σ_u^2 through $\tilde{\boldsymbol{\beta}}$ and γ_d:

$$\tilde{\delta}_d^{BLUP} = \tilde{\delta}_d^{BLUP}(\sigma_u^2)$$

- **Empirical** BLUP (EBLUP) of δ_d: $\hat{\sigma}_u^2$ estimator of σ_u^2,

$$\hat{\delta}_d^{EBLUP} = \tilde{\delta}_d^{BLUP}(\hat{\boldsymbol{\sigma}}_{\mathbf{u}}^{\mathbf{2}}), \quad d = 1, \ldots, D$$

- The EBLUP remains **model-unbiased** provided:
  - ✓ $\hat{\sigma}_u^2$ even: $\hat{\sigma}_u^2(\mathbf{y}) = \hat{\sigma}_u^2(-\mathbf{y})$;
  - ✓ $\hat{\sigma}_u^2$ translation invariant: $\hat{\sigma}_u^2(\mathbf{y} + \mathbf{X}\boldsymbol{\gamma}) = \hat{\sigma}_u^2(\mathbf{y})$ for all $\mathbf{y}$ and $\boldsymbol{\gamma}$.

FITTING METHODS

- ✓ FH fitting method;
- ✓ Maximum Likelihood (ML);
- ✓ Restricted/Residual ML (REML);
- ✓ Prasad-Rao moments method.

FH FITTING METHOD

- It holds that

$$\hat{\theta}_d^{DIR} \stackrel{ind}{\sim} N(\mathbf{x}_d'\boldsymbol{\beta}, \sigma_u^2 + \psi_d) \Rightarrow \sum_{d=1}^{D} \frac{\left\{\hat{\theta}_d^{DIR} - \mathbf{x}_d'\tilde{\boldsymbol{\beta}}(\sigma_u^2)\right\}^2}{\sigma_u^2 + \psi_d} \sim \mathcal{X}_{D-p}^2$$

- **Fay-Herriot fitting method:** Solve iteratively for σ_u^2 the moment equation

$$h(\sigma_u^2) = \sum_{d=1}^{D} \frac{\left(\hat{\theta}_d^{DIR} - \mathbf{x}_d'\tilde{\boldsymbol{\beta}}(\sigma_u^2)\right)^2}{\sigma_u^2 + \psi_d} = D - p.$$

Stop when iterations converge to a solution $\tilde{\sigma}_u^2$

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

Normality is not needed.

OTHER FITTING METHODS

- **Maximum likelihood:** Assumes normality

$$\hat{\theta}_d^{DIR} \overset{ind}{\sim} N(\mathbf{x}_d'\boldsymbol{\beta}, \sigma_u^2 + \psi_d)$$

  ML estimators remain consistent without normality.

- **Restricted maximum likelihood (REML):** Reduces the bias of ML estimators for small sample size n compared to p.
- **Prasad-Rao method:** Based on method of moments. Provides good starting values for iterative fitting algorithms.

(✓ Prasad and Rao, 1990)

MEAN SQUARED ERROR

- Under normality of u_d and e_d, as $D \to \infty$,

$$\mathrm{MSE}(\hat{\delta}_d^{EBLUP}) = g_{1d}(\sigma_u^2) + g_{2d}(\sigma_u^2) + g_{3d}(\sigma_u^2) + o(D^{-1}),$$

where

$$\begin{aligned} g_{1d}(\sigma_u^2) &= \gamma_d \psi_d = O(1), \\ g_{2d}(\sigma_u^2) &= (1-\gamma_d)^2 \mathbf{x}_d' \left(\sum_{d=1}^{D} \gamma_d \mathbf{x}_d \mathbf{x}_d' \right)^{-1} \mathbf{x}_d = O(D^{-1}), \\ g_{3d}(\sigma_u^2) &= (1-\gamma_d)^2 \gamma_d \sigma_u^{-2} \bar{V}(\hat{\sigma}_u^2) = O(D^{-1}), \end{aligned}$$

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

✓ *Prasad & Rao (1990), JASA*

MEAN SQUARED ERROR

Sketch of proof (ML estimation): We have obtained $\boldsymbol{\mu} = \boldsymbol{\alpha}'\mathbf{u}$, where

$$\boldsymbol{\alpha}' = \boldsymbol{\ell}'\mathbf{Q}\mathbf{X}'\mathbf{V}^{-1} + \sigma_u^2\mathbf{m}'\mathbf{P}, \quad \mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}\mathbf{Q}\mathbf{X}'\mathbf{V}^{-1},$$

for $\mathbf{Q} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = (\sum_d \gamma_d \mathbf{x}_d\mathbf{x}_d')^{-1}$.

Replacing $\boldsymbol{\alpha}'$ and $\mathbf{m}' = (\mathbf{0}'_{d-1}, 1, \mathbf{0}'_{D-d})$ in $\mathrm{MSE}(\tilde{\mu})$, and noting that

$$\mathbf{PVP} = \mathbf{P},\ \mathbf{PX} = \mathbf{0}_D,$$

we obtain

$$\mathrm{MSE}(\tilde{\delta}_d^{BLUP}) = g_{1d}(\sigma_u^2) + g_{2d}(\sigma_u^2),$$

where

$$\begin{aligned} g_{1d}(\sigma_u^2) &= \gamma_d\psi_d, \\ g_{2d}(\sigma_u^2) &= (1-\gamma_d)^2\mathbf{x}_d'(\sum_d \gamma_d\mathbf{x}_d\mathbf{x}_d')^{-1}\mathbf{x}_d. \end{aligned}$$

MEAN SQUARED ERROR

- MSE decomposition:

$$\begin{aligned}
\text{MSE}(\hat{\delta}_d^{EBLUP}) &= E(\hat{\delta}_d^{EBLUP} - \delta_d)^2 \\
&= E(\hat{\delta}_d^{EBLUP} - \tilde{\delta}_d^{BLUP} + \tilde{\delta}_d^{BLUP} - \delta_d)^2 \\
&= \text{MSE}(\tilde{\delta}_d^{BLUP}) + E(\hat{\delta}_d^{EBLUP} - \tilde{\delta}_d^{BLUP})^2 \\
&+ 2E(\hat{\delta}_d^{EBLUP} - \tilde{\delta}_d^{BLUP})(\tilde{\delta}_d^{BLUP} - \delta_d).
\end{aligned}$$

- If $\hat{\sigma}_u^2$ is even and translation invariant, then under normality

$$E(\hat{\delta}_d^{EBLUP} - \tilde{\delta}_d^{BLUP})(\tilde{\delta}_d^{BLUP} - \delta_d) = 0.$$

- Then,

$$\text{MSE}(\hat{\delta}_d^{EBLUP}) = \text{MSE}(\tilde{\delta}_d^{BLUP}) + E(\hat{\delta}_d^{EBLUP} - \tilde{\delta}_d^{BLUP})^2.$$

✓ *Kackar & Harville (1984), JASA*

MEAN SQUARED ERROR

- First-order Taylor expansion of $\tilde{\delta}_d^{BLUP}(\hat{\sigma}_u^2)$ around σ_u^2:

$$\hat{\delta}_d^{EBLUP} \approx \tilde{\delta}_d^{BLUP} + \frac{\partial \tilde{\delta}_d^{BLUP}}{\partial \sigma_u^2}(\hat{\sigma}_u^2 - \sigma_u^2).$$

- Then,

$$E(\hat{\delta}_d^{EBLUP} - \tilde{\delta}_d^{BLUP})^2 \approx E\left[\left(\frac{\partial \tilde{\delta}_d^{BLUP}}{\partial \sigma_u^2}\right)^2 (\hat{\sigma}_u^2 - \sigma_u^2)^2\right].$$

- Replace $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v}$ in $\tilde{\delta}_d^{BLUP} = \boldsymbol{\alpha}'\mathbf{y}$:

$$\tilde{\delta}_d^{BLUP} = \boldsymbol{\ell}'\boldsymbol{\beta} + \mathbf{b}'\mathbf{v}, \quad \mathbf{v} = \mathbf{u} + \mathbf{e} \sim N(\mathbf{0}_D, \mathbf{V}).$$

  Then,

$$\frac{\partial \tilde{\delta}_d^{BLUP}}{\partial \sigma_u^2} = \frac{\partial \mathbf{b}'}{\partial \sigma_u^2}\mathbf{v}.$$

MEAN SQUARED ERROR

- Then

$$E(\hat{\delta}_d^{EBLUP} - \tilde{\delta}_d^{BLUP})^2 \approx E\left[\frac{\partial \mathbf{b}'}{\partial \sigma_u^2}\mathbf{v}\mathbf{v}'\left(\frac{\partial \mathbf{b}'}{\partial \sigma_u^2}\right)'(\hat{\sigma}_u^2 - \sigma_u^2)^2\right].$$

- $\hat{\sigma}_u^2$ ML estimator of σ_u^2.
- By first-order Taylor expansion of $s(\hat{\sigma}_u^2) = \partial \log L(\hat{\sigma}_u^2)/\partial \hat{\sigma}_u^2$ around σ_u^2, and noting that $\partial s(\sigma_u^2)/\partial \sigma_u^2 \xrightarrow{P} -\mathcal{I}(\sigma_u^2)$, where $\mathcal{I}(\sigma_u^2)$ is the Fisher information,

$$\hat{\sigma}_u^2 \approx \sigma_u^2 + \mathcal{I}(\sigma_u^2)s(\sigma_u^2)$$

MEAN SQUARED ERROR

- log-likelihood function:

$$\log L(\sigma_u^2) = -\frac{D}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}(\mathbf{y}-\mathbf{X}\beta)'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\beta).$$

- Score function:

$$s(\sigma_u^2) = -\frac{1}{2}\mathrm{tr}(\mathbf{V}^{-1}) - \underbrace{(\mathbf{y}-\mathbf{X}\beta)'}_{\mathbf{v}'}\mathbf{V}^{-3}\underbrace{(\mathbf{y}-\mathbf{X}\beta)}_{\mathbf{v}}.$$

- Fisher information:

$$\mathcal{I}(\sigma_u^2) = -\frac{1}{2}\mathrm{tr}(\mathbf{V}^{-2}).$$

- Then calculate the expected value

$$E(\hat{\delta}_d^{EBLUP} - \tilde{\delta}_d^{BLUP})^2 \approx E\left[\frac{\partial \mathbf{b}'}{\partial \sigma_u^2}\mathbf{v}\mathbf{v}'\left(\frac{\partial \mathbf{b}'}{\partial \sigma_u^2}\right)'\mathcal{I}^2(\sigma_u^2)s^2(\sigma_u^2)\right].$$

MSE ESTIMATOR

- It holds

$$E[g_{1d}(\hat{\sigma}_u^2)] \approx g_{1d}(\sigma_u^2) - g_{3d}(\sigma_u^2),$$
$$E[g_{2d}(\hat{\sigma}_u^2)] \approx g_{2d}(\sigma_u^2), \quad E[g_{3d}(\hat{\sigma}_u^2)] \approx g_{3d}(\sigma_u^2).$$

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

$$\mathrm{mse}(\hat{\delta}_d^{EBLUP}) = g_{1d}(\hat{\sigma}_u^2) + g_{2d}(\hat{\sigma}_u^2) + 2g_{3d}(\hat{\sigma}_u^2)$$

- Nearly unbiased:

$$E\left[\mathrm{mse}(\hat{\delta}_d^{EBLUP})\right] = \mathrm{MSE}(\hat{\delta}_d^{EBLUP}) + o(D^{-1})$$

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

✓ *Prasad & Rao (1990), JASA*

EBLUP BASED ON FAY-HERRIOT MODEL

TARGET INDICATORS:

- General indicators.

DATA REQUIREMENTS:

- Aggregated values of p auxiliary variables at the domain level.
- Domain population sizes.

EBLUP BASED ON FAY-HERRIOT MODEL

ADVANTAGES:

- Requires only **area level** auxiliary information, which is **easily** available and avoids confidentiality issues.
- Makes use of the **sampling weights** when $\gamma_d \neq 0$. Design-consistent as $n_d \to \infty$. Hence, less affected by informative sampling.
- Automatically gives **more weight** to the regression estimator when sample size is **too small** in a given area.
- It often has better efficiency than the direct estimator.
- It accounts for unexplained between-area heterogeneity if $\gamma_d \neq 0$.

EBLUP BASED ON FAY-HERRIOT MODEL

ADVANTAGES:

- It tends to the direct estimator as the domain sample size increases (ψ_d decreases).
- For **linear** direct estimators, CLT applies for areas with not so small sample sizes, so goodness-of-fit **minimally ensured**. Isolated outliers have small effect because of averaging.
- Prasad-Rao MSE estimator stable estimator of design MSE and **design-unbiased** when **averaging** over a large number of areas. *For non-sampled domains, the synthetic component can be used ($\gamma_d = 0$).*

EBLUP BASED ON FAY-HERRIOT MODEL

DISADVANTAGES:

- Information **loss** in the aggregation process of auxiliary variables.
- Only D (typically $<< n$) observations to fit the model. In our examples, very **mild** gains over direct estimators.
- Model checking is required. Potential **linearity** problems for non-linear parameters.
- It requires **preliminary** estimation of sampling variances ψ_d. Same small area problem!

EBLUP BASED ON FAY-HERRIOT MODEL

DISADVANTAGES:

- If we wish to estimate several indicators defined in terms of the same target variable, different modelling is required.
- Estimators cannot be disaggregated for subdomains.
- MSE estimator by Prasad-Rao formula correct **under the model** with normality (not design-unbiased for design MSE for a given area).
- Benchmarking adjustment required.

NESTED-ERROR MODEL

- y_{dj} value of target variable for unit j within area d
- u_d random effect of area d
- **Nested error** linear regression model:

$$y_{dj} = \mathbf{x}'_{dj}\boldsymbol{\beta} + u_d + e_{dj}, \quad j = 1, \dots, N_d, \; d = 1, \dots, D$$
$$u_d \overset{iid}{\sim} N(0, \sigma_u^2), \quad e_{dj} \overset{iid}{\sim} N(0, \sigma_e^2)$$

- Model in matrix notation:

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

- Marginal expectation and variance:

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

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

BLUP: GENERAL LINEAR MODEL

More general linear model:

- $\mathbf{y} = (y_1, \ldots, y_N)'$ population vector (**random**)
- Linear model:

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

- 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:

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

BLUP: GENERAL LINEAR MODEL

Best linear unbiased predictor (BLUP): V known

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

$$\begin{array}{ll} \text{mín}_{\boldsymbol{\alpha} \in \boldsymbol{R}^n} & \text{MSE}(\tilde{\delta}) = E(\tilde{\delta} - \delta)^2 \\ \text{s.t.} & E(\tilde{\delta} - \delta) = 0 \end{array}$$

is given by

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

where

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

✓ *Royall (1970), Biometrika*

BLUP UNDER NESTED ERROR MODEL

- Under the **nested-error model**, the BLUP of $\delta = \bar{Y}_d$ is:

$$\tilde{\bar{Y}}_d^{BLUP} = \frac{1}{N_d}\left(\sum_{j\in s_d} y_{dj} + \sum_{j\in r_d} \tilde{y}_{dj}^{BLUP}\right),$$

where

$$\tilde{y}_{dj}^{BLUP} = \mathbf{x}_{dj}'\tilde{\boldsymbol{\beta}} + \tilde{u}_d,\ \tilde{\boldsymbol{\beta}} \text{ WLS estimator of } \boldsymbol{\beta},$$
$$\tilde{u}_d = \gamma_d(\bar{y}_d - \bar{\mathbf{x}}_d'\tilde{\boldsymbol{\beta}}),\ \gamma_d = \sigma_u^2/(\sigma_u^2 + \sigma_e^2/n_d).$$

- When $n_d/N_d \approx 0$,

$$\tilde{\bar{Y}}_d^{BLUP} \approx \gamma_d\left\{\bar{y}_d + (\bar{\mathbf{X}}_d - \bar{\mathbf{x}}_d)'\tilde{\boldsymbol{\beta}}\right\} + (1-\gamma_d)\bar{\mathbf{X}}_d'\tilde{\boldsymbol{\beta}}$$

- Weighted average of **"survey regression"** estimator $\bar{y}_d + (\bar{\mathbf{X}}_d - \bar{\mathbf{x}}_d)'\tilde{\boldsymbol{\beta}}$ and **regression synthetic** estimator $\bar{\mathbf{X}}_d'\tilde{\boldsymbol{\beta}}$.

EMPIRICAL BLUP (EBLUP)

- BLUP depends on unknown $\boldsymbol{\theta} = (\sigma_u^2, \sigma_e^2)'$:

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

- EBLUP of δ: $\hat{\boldsymbol{\theta}} = (\hat{\sigma}_u^2, \hat{\sigma}_e^2)'$ estimator of $\boldsymbol{\theta}$

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

- Estimators of σ_u^2 and σ_e^2:
  - ✓ Henderson method III (moments method);
  - ✓ ML;
  - ✓ REML.

EBLUP UNDER A UNIT LEVEL MODEL

TARGET INDICATORS:

- Means or totals of the variable of interest.

DATA REQUIREMENTS:

- Microdata for the p auxiliary variables in the survey.
- Domain indicator in the survey.
- Population means of the p auxiliary variables for the domains.

EBLUP UNDER A UNIT LEVEL MODEL

ADVANTAGES:

- It uses unit level auxiliary information, which is typically much **richer** than area level information.
- Total sample size is typically very **large** ($n >> D$), so borrowing a lot of strength.
- It accounts for unexplained between-area heterogeneity.
- Model checking is required.
- It does not require sampling variances of direct estimators.
- It automatically borrows strength when domain sample size is small and tends to the "survey regression" estimator as the domain sample size grows.

EBLUP UNDER A UNIT LEVEL MODEL

ADVANTAGES:

- Estimates can be disaggregated for subareas (without sub-area effect) or even for individuals.
- **Unbiased** estimators under the model (normality not really needed, only symmetry).
- **Nearly unbiased MSE** estimators under the model with **normality**.
- Model MSE estimator stable for design-based MSE and design-unbiased when averaging for many domains.
- The synthetic part can be used for non-sampled areas.

EBLUP UNDER A UNIT LEVEL MODEL

DISADVANTAGES:

- Unit level auxiliary information **not easily available**.
- Only applicable to **linear** parameters.
- Does **not use sampling weights**, so not good design properties for a given area. Problems under **informative** sampling.
- It can be affected by outliers and/or lack of normality.

EBLUP UNDER A UNIT LEVEL MODEL

DISADVANTAGES:

- **Sensitive** to model departures. **Model checking** very important.
- MSE estimator by Prasad-Rao formula correct **under the model** with normality (not design-unbiased for design MSE for a given area).
- Benchmarking adjustment required.

BEST PREDICTOR

Best Predictor (BP)

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

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

Proof: Define $\delta^0 = E_{\mathbf{y}_r}(\delta|\mathbf{y}_s)$. Note that

$$\mathrm{MSE}(\tilde{\delta}) = E_{\mathbf{y}}\{(\tilde{\delta} - \delta^0)^2\} + 2\, E_{\mathbf{y}}\{(\tilde{\delta} - \delta^0)(\delta^0 - \delta)\} + E_{\mathbf{y}}\{(\delta^0 - \delta)^2\}.$$

The last term does not depend on $\tilde{\delta}$. For the second term,

$$\begin{aligned} E_{\mathbf{y}}\{(\tilde{\delta} - \delta^0)(\delta^0 - \delta)\} &= E_{\mathbf{y}_s}\left[E_{\mathbf{y}_r}\left\{(\tilde{\delta} - \delta^0)(\delta^0 - \delta)|\mathbf{y}_s\right\}\right] \\ &= E_{\mathbf{y}_s}\left[(\tilde{\delta} - \delta^0)\left\{\delta^0 - E_{\mathbf{y}_r}(\delta|\mathbf{y}_s)\right\}\right] = 0. \end{aligned}$$

The minimizer of $E_{\mathbf{y}}\{(\tilde{\delta} - \delta^0)^2\}$ is exactly $\tilde{\delta}^{BP} = \delta^0 = E_{\mathbf{y}_r}(\delta|\mathbf{y}_s)$.

EMPIRICAL BEST PREDICTOR

- The best predictor is unbiased:

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

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

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

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

$$\hat{\delta}^{EBP} = \tilde{\delta}^{BP}(\tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}).$$

BEST PREDICTOR: LINEAR PARAMETER

- Particular case: Consider a linear target parameter

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

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

$$\tilde{\delta}^{BP} = \mathbf{a}_s'\mathbf{y}_s + \mathbf{a}_r'\tilde{\mathbf{y}}_r^{BP},$$

where

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

- In this case, EBP equals EBLUP.

EB METHOD: POVERTY ESTIMATION

- Domain poverty indicators:

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

- The distribution of incomes E_{dj} is highly right skewed.
- Select a transformation $T()$ such that the distribution of $y_{dj} = T(E_{dj})$ is approximately Normal.
- **Assumption:** $y_{dj} = T(E_{dj})$ satisfies the nested error model

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

EB METHOD: POVERTY ESTIMATION

- Area vector: $\mathbf{y}_d = (y_{d1}, \ldots, y_{dN_d})'$.
- Poverty indicators in terms of $\mathbf{y}_d$:

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

- Partition $\mathbf{y}_d$ into sample and non-sample: $\mathbf{y}_d = (\mathbf{y}_{ds}', \mathbf{y}_{dr}')'$
- **Best estimator:**

$$\tilde{F}_{\alpha d}^{BP} = E_{\mathbf{y}_{dr}}\left[F_{\alpha d} | \mathbf{y}_{ds} \right].$$

✓ *Molina and Rao (2010), CJS*

EB METHOD

- Distribution of $\mathbf{y}_{dr}$ given $\mathbf{y}_{ds}$ under nested-error model:

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

where

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

and

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

- The conditional distrib. depends on $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma_u^2, \sigma_e^2)'$.
- **Empirical best (EB) estimator:** Replace a consistent estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$

$$\hat{F}_{\alpha d}^{EBP} = \tilde{F}_{\alpha d}^{BP}(\hat{\boldsymbol{\theta}}).$$

✓ *Molina and Rao (2010), CJS*

MONTE CARLO APPROXIMATION

(a) Generate L out-of-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 target parameter 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:

$$\hat{F}_{\alpha d}^{EBP} = \frac{1}{L} \sum_{\ell=1}^{L} F_{\alpha d}^{(\ell)}.$$

(d) MSE estimated by parametric bootstrap.

✓ *Molina and Rao (2010), CJS*

PARAMETRIC BOOTSTRAP MSE

(i) From the fitted model, generate B bootstrap populations

$$\mathbf{y}^{*(b)} = (\mathbf{y}_1^{*(b)}, \ldots, \mathbf{y}_D^{*(b)}), \quad b = 1, \ldots, B.$$

(ii) Calculate true bootstrap parameters

$$\delta_d^{*(b)} = h(\mathbf{y}_d^{*(b)}), \quad b = 1, \ldots, B.$$

(iii) With the sample part $\mathbf{y}_s^{*(b)} = (\mathbf{y}_{1s}^{*(b)}, \ldots, \mathbf{y}_{Ds}^{*(b)})'$ of the population vector $\mathbf{y}^{*(b)}$, compute EB estimators

$$\hat{\delta}_d^{EBP*(b)}, \quad b = 1, \ldots, B.$$

(iv) **Naive parametric bootstrap MSE estimator:**

$$mse_*(\hat{\delta}_d^{EBP}) = \frac{1}{B}\sum_{b=1}^{BP}\left(\hat{\delta}_d^{EBP*(b)} - \delta_d^{*(b)}\right)^2$$

✓ *González-Manteiga et al. (2008), JSCS*

EB UNDER A UNIT LEVEL MODEL

TARGET INDICATORS:

- General indicators defined in terms of one continuous variable (e.g. income), which will be modeled.

DATA REQUIREMENTS:

- Microdata for the p auxiliary variables in the survey.
- Domain indicator in the survey.
- Microdata for the p auxiliary variables for all the population units (census or admin. register).

EB UNDER A UNIT LEVEL MODEL

ADVANTAGES:

- It uses unit level auxiliary information, which is typically much **richer** than area level information.
- Total sample size is typically very **large** ($n >> D$), so borrowing a lot of strength.
- It accounts for unexplained between-area heterogeneity.
- Applicable to estimate **general** non-linear parameters $h(\mathbf{y})$, where $\mathbf{y}$ is normally distributed.

EB UNDER A UNIT LEVEL MODEL

ADVANTAGES:

- Full censuses are generated. Then, **several indicators** can be obtained at the same time without new modelling and generation.
- Approximately **unbiased and optimal** estimators under the model with normality.
- The same fitted model can be used to estimate several indicators.
- Estimates can be disaggregated to whatever subdomains (without subdomain effect), even at the unit level.
- **Nearly unbiased MSE** estimators under the model with **normality**.
- Model MSE estimator stable for design MSE and design-unbiased when averaging for many domains.

EB UNDER A UNIT LEVEL MODEL

DISADVANTAGES:

- Unit level auxiliary information for each population unit (census/register) **not easily available**.
- Computationally **intensive**.
- Does **not use sampling weights**, so not good design properties for a given area. Problems under **informative** sampling.
- **Sensitive** to model departures. Finding the correct transformation of variables and **model checking** very important. Model checking is crucial.
- MSE estimators obtained by bootstrap are computationally intensive.

GENERALIZED LINEAR MIXED MODELS

- $y_{dj} \in \{0, 1\}$, where 1=presence of the characteristic of interest, 0=absence.
- Target parameters: Proportions of individuals with the characteristic,

$$P_d = \frac{1}{N_d} \sum_{j=1}^{N_d} y_{dj}, \quad d = 1, \ldots, D.$$

- Logistic mixed model:

$$y_{dj}|u_d \stackrel{ind.}{\sim} \text{Bern}(p_{dj}), \quad j = 1, \ldots, N_d,\ d = 1, \ldots, D,$$
$$p_{dj} = \frac{\exp(\mathbf{x}_{dj}'\boldsymbol{\beta} + u_d)}{1 + \exp(\mathbf{x}_{dj}'\boldsymbol{\beta} + u_d)}, \quad u_d \stackrel{iid}{\sim} N(0, \sigma_u^2).$$

✓ *González-Manteiga et al. (2007), CSDA*

SMALL AREA ESTIMATORS

- Best predictor:

$$\hat{P}_d^{BP} = \frac{1}{N_d}\left\{\sum_{j\in s_d} y_{dj} + \sum_{j\in r_d} E(y_{dj}|\mathbf{y}_{ds})\right\}, \quad d = 1,\dots,D.$$

- The expectation $E(y_{dj}|\mathbf{y}_{ds})$ cannot be calculated analytically: approximations (e.g. Laplace) or Monte Carlo simulation methods are required.
- Simple plug-in estimator:

$$\hat{P}_d^{Plug} = \frac{1}{N_d}\left(\sum_{j\in s_d} y_{dj} + \sum_{j\in r_d} \hat{p}_{dj}^{Plug}\right), \quad d = 1,\dots,D.$$

- $\hat{p}_{dj}^{Plug} = \exp(\mathbf{x}'_{dj}\hat{\boldsymbol{\beta}} + \hat{u}_d)/\{1 + \exp(\mathbf{x}'_{dj}\hat{\boldsymbol{\beta}} + \hat{u}_d)\}$ predicted probabilities through the GLMM fit.

FITTING METHODS

- Sample likelihood:

$$f(\mathbf{y}_s) = \int_{\boldsymbol{R}^D} f(\mathbf{y}_s, \mathbf{u}) d\mathbf{u} = \int_{\boldsymbol{R}^D} f_1(\mathbf{y}_s|\mathbf{u}) f_2(\mathbf{u}) d\mathbf{u}$$

- No analytical expression for the likelihood.
- ML: Approximations (e.g. Laplace) or numerical methods are required to maximize the likelihood.

PENALIZED QUASI-LIKELIHOOD (PQL) +APPROXIMATE ML

(A) σ_u^2 known: PQL algorithm (✓ Breslow and Clayton, 1993):

$$(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}) = \text{argmax}_{(\boldsymbol{\beta},\mathbf{u})} f(\mathbf{y}_s, \mathbf{u})$$

(B) $\boldsymbol{\beta}$ and $\mathbf{u}$ known: approximate ML

$$\hat{\sigma}_u^2 = \text{argmax}_{\sigma_u^2} f_L(\mathbf{y}_s)$$

f_L multivariate normal likelihood of a linear mixed model approximating the GLMM.

✓ *Schall (1991)* ✓ *Saei and Chambers (2003)*

PENALIZED QUASI-LIKELIHOOD (PQL) +APPROXIMATE ML

- It delivers possibly inconsistent estimators.
- MSE can be estimated by parametric bootstrap.
- GLMM fitting+EBP+Bootstrap MSE: highly computationally intensive. Unfeasible for large populations.
- GLMM fitting+Plug-in estimator+Bootstrap MSE: more feasible but not optimal.

EXTENSION: SEVERAL CATEGORIES

- Y_{d1} total unemployed in area d;
- Y_{d2} total employed in area d;
- R_d unemployment rate in area d;

$$R_d = \frac{Y_{d1}}{Y_{d1} + Y_{d2}} \times 100.$$

MULTINOMIAL LOGISTIC MIXED MODEL

- Three exclusive categories:
  y_{dj1} 1=unemployed, 0=otherwise
  y_{dj2} 1=employed, 0=otherwise
  y_{dj3} 1=inactive, 0=otherwise

- **Multivariate model:**

$$(y_{dj1}, y_{dj2}, y_{dj3}) \sim \text{Multin}(m_{dj}; p_{dj1}, p_{dj2}, p_{dj3})$$

$$\begin{array}{ll} \text{Unemployed :} & \log(p_{dj1}/p_{dj3}) = \mathbf{x}'_{dj1}\boldsymbol{\beta}_1 + u_{d1} \\ \text{Employed :} & \log(p_{dj2}/p_{dj3}) = \mathbf{x}'_{dj2}\boldsymbol{\beta}_2 + u_{d2} \end{array}$$

- Category-specific random effects: $\mathbf{u} = (u_{d1}, u_{d2})' \sim N_2(0, \Sigma_u)$.

MULTINOMIAL LOGISTIC MIXED MODEL

- Plug-in estimates of unemployed/employed totals:

$$\hat{Y}_{dk}^{Plug} = \sum_{j \in s_d} y_{djk} + \sum_{j \in r_d} \hat{p}_{djk}^{Plug}, \quad k = 1, 2.$$

- Plug-in estimates of unemployment rates:

$$R_d^{Plug} = \frac{\hat{Y}_{d1}^{Plug}}{\hat{Y}_{d1}^{Plug} + \hat{Y}_{d2}^{Plug}} \times 100.$$

✓ *Molina, Saei and Lombardía (2007), JRSSA*

MODELS FOR BINARY DATA

TARGET INDICATORS:

- Proportions o totals of a binary variable (e.g. absence of certain commodity).

DATA REQUIREMENTS:

- Microdata for the p auxiliary variables in the survey.
- Domain indicator in the survey.
- Microdata for the p auxiliary variables for all the population units (census or admin. register).

MODELS FOR BINARY DATA

ADVANTAGES:

- It uses unit level auxiliary information, which is typically much **richer** than area level information.
- Total sample size is typically very **large** ($n >> D$), so borrowing a lot of strength.
- It accounts for unexplained between-area heterogeneity.
- EB approximately **unbiased and optimal** under the model.
- Estimates can be disaggregated to whatever subdomains (without subdomain effect), even at the unit level.

MODELS FOR BINARY DATA

ADVANTAGES:

- The synthetic part can be used to estimate in non-sampled areas.
- El estimador del ECM bajo el modelo obtenido e.g. por procedimientos bootstrap es un estimador estable del ECM bajo el diseño y es insesgado bajo el diseño cuando se promedia a lo largo de muchas áreas.
- Bootstrap MSE estiamtor stable for design MSE and design-unbiased when averaging for many domains.

MODELS FOR BINARY DATA

DISADVANTAGES:

- Unit level auxiliary information for each population unit (census/register) **not easily available**.
- Computationally **intensive**.
- Does **not use sampling weights**, so not good design properties for a given area. Problems under **informative** sampling.
- **Sensitive** to model departures. Finding the correct transformation of variables and **model checking** very important. Model checking is crucial.
- EB estimator (unlike plug-in) is computationally intensive.
- MSE estimators obtained by bootstrap are computationally intensive (even more for EB estimator).
- Benchmarking adjustment is required.

SOFTWARE

The R package `sae` contains functions:

- Direct estimators: `direct`.
- Traditional indirect estimators: `pssynt`, `ssd`.
- FH model: `eblupFH`, `mseFH`.
- Spatial FH model: `eblupSFH`, `mseSFH`, `pbmseSFH`, `npbmseSFH`.
- Spatio-temporal FH model: `eblupSTFH`, `pbmseSTFH`.
- Nested error model: `eblupBHF`, `pbmseBHF`.
- EB method: `ebBHF`, `pbmseebBHF`.
- Data sets and examples.

✓ *Molina and Marhuenda (2015), The R Journal*

DISCUSSION

(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 different institutions needed.

(c) Model validation crucial. External evaluation studies are also needed.

(d) Area-level models have wider scope than unit level models 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. Unit level models can gain much more efficiency if unit level information is available.

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, *J. Amer. Statist. Assoc.*, **83**, 28–36.
- Breslow, N.E., Clayton, D.G. (1993), Approximate inference in generalized linear mixed models. J. Amer. Statist. Assoc. **88**, 9–25.
- Deville, J.C. and Särndal, C.E. (1992). Calibration estimation in Survey Sampling, *J. Amer. Statist. Assoc.*, **87**, 376–382.
- Drew, D., Singh, M.P. and Choudhry, G.H. (1982). Evaluation of Small Area Estimation Techniques for the Canadian Labour Force Survey, *Survey Methodology*, **8**, 17–47.
- Foster, J., Greer, J. and Thorbecke, E. (1984). A class of decomposable poverty measures, *Econometrica*, **52**, 761–766.

REFERENCES

- Fay, R.E. and Herriot, R.A. (1979). Estimation of Income for Small Places: An Application of James-Stein Procedures to Census Data, *J. Amer. Statist. Assoc.*, **74**, 269–277.
- González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D. and Santamaría, L. (2007), Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model, *Computational Statistics and Data Analysis*, **51**, 2720–2733.
- Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). Sample Survey Methods and Theory I, New York: Wiley.
- Kackar, R.N. and Harville, D.A. (1984). Approximations for Standard Errors of Estimators of Fixed Random Effects in Mixed Linear Models, *J. Amer. Statist. Assoc.*, **79**, 853–862.
- Prasad, N.G.N. and Rao, J.N.K. (1990). The Estimation of the Mean Squared Error of Small-Area Estimators, *J. Amer. Statist. Assoc.*, **85**, 163–171.

REFERENCES

- Molina, I. and Marhuenda, Y. (2015), sae: An R Package for Small Area Estimation, *The R Journal*, **7**.
- Molina, I., Saei, A. and Lombardía, M.J. (2007), Small area estimates of labour force participation under a multinomial logit mixed model, *Journal of the Royal Statistical Society, Series A*, **170**, 975–1000.
- Molina, I. and Rao (2010). Small Area Estimation of Poverty Indicators. *Canadian Journal of Statistics*, **38**, 369–385.
- Rao, J.N.K. and Molina, I. (2015). Small Area Estimation, Second Edition. Wiley: Hoboken, NJ.
- Royall, R.M. (1970). On Finite Population Sampling Theory Under Certain Linear Regression, *Biometrika*, **57**, 377–387.

REFERENCES

- Saei, A., Chambers, R., 2003. Small area estimation under linear and generalized linear mixed models with time and area effects. S3RI Methodology Working Paper M03/15. Southampton Statistical Sciences Research Institute, University of Southampton.
- Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika* **78**, 719—727.
- Särndal, C.E., Swenson, B. and Wretman, J.H. (1992). Model Assisted Survey Sampling, New York: Springer-Verlag.