Discussion of W.R. Bell's paper

Partha Lahiri

JPSM
University of Maryland, College Park

The Fay-Herriot Model: An Area Level Normal Model

For $i = 1, \cdots, m$, assume
Level 1: (Sampling Model) $Y_i = \theta_i + e_i$;
Level 2: (Linking Model) $\theta_i = x_i'\beta + v_i$.
Equivalently, $Y_i = x_i'\beta + v_i + e_i \ , \quad i = 1, \cdots, m \ .$

- Y_i : direct estimate for area i;
- D_i : *known* sampling variance of y_i.
- x_i : a $p \times 1$ column vector of known auxiliary variables;
- $X' = (x_1, \cdots, x_m)$, and $\Sigma(A) = diag\{A + D_j; j = 1, ..., m\}$.
- $\{e_i\}$ and $\{v_i\}$ are *independent* with $e_i \sim N[0, D_i]$ and $v_i \sim N[0, A]$,
- Lower dimensional ($p + 1$) or Hyperparameters: β and A
- Higher dimensional (m) or small area means: θ_i
- **Main Objective:** Estimation of small area means

The BP, BLUP and EBLUP

The BP (Bayes estimator) of θ_i

$$\hat{\theta}_i(Y_i; A) = B_i Y_i + (1 - B_i)x_i'\beta,$$

where $B_i = \frac{A}{D_i + A}$, $i = 1, \ldots, m$.

The BLUP of θ_i

$$\hat{\theta}_i(Y_i; A) = B_i Y_i + (1 - B_i)x_i'\hat{\beta}(A),$$

where $\hat{\beta}(A) = [X'\Sigma^{-1}(A)X]^{-1}X'\Sigma^{-1}(A)Y$ and $B_i = \frac{A}{D_i + A}$, $i = 1, \ldots, m$.

An EBLUP of θ_i

$$\hat{\theta}_i(Y_i; \hat{A}) = \hat{B}_i Y_i + (1 - \hat{B}_i)x_i'\hat{\beta}(\hat{A}) =: \hat{\theta}_i,$$

where $\hat{B}_i = \frac{\hat{A}}{D_i + \hat{A}}$, $i = 1, \ldots, m$, and $\hat{A}$ is a consistent estimator of A.

Components of the FH Model

- Level 1 Components
  - 1. Normality
  - 2. y_i are unbiased
  - 3. D_i are known
  - 4. Independence

- Level 2 Components
  - 1. Normality
  - 2. Linear mean function
  - 3. Homoscedasticity
  - 4. Independence

- Bell's paper considered impact of Level 1 misspecification of D_i
- Impact is more on MSE estimation than point estimation
- Underestimation for D_i's are more severe than underestimation

Impact of Non-normality in Both Levels: A Non-Normal Area Level Model

- $\{e_i\}$ and $\{v_i\}$ are uncorrelated with $e_i \sim [0, D_i, \kappa_{ei}]$ and $v_i \sim [0, A, \kappa_v]$, $[\mu, \sigma^2, \kappa]$ representing a probability distribution with mean μ, variance σ^2 and kurtosis κ. Let $\Phi = Diag\{\kappa_{ej}; j = 1, \cdots, m\}$.
- $\kappa = \mu_4/\sigma^4 - 3$, where μ_4 is the the fourth central moment of the distribution respectively.
- We assume that $[\beta, A, \kappa_v]$ is unknown, but $[D_i, \kappa_{ei}]$ is known.

$MSPE(\hat{\theta}_i) = E(\hat{\theta}_i - \theta_i)^2$, where the expectation is taken over the joint distribution of Y and θ under the non-normal Fay-Herriot model.

We decompose the MSPE of EBLUP $\hat{\theta}_i$ as

$$
\begin{aligned}
MSPE[\hat{\theta}_i(Y_i, \hat{A})] &= MSPE[\hat{\theta}_i(Y_i, A)] + E[\hat{\theta}_i(Y_i, \hat{A}) - \hat{\theta}_i(Y_i, A)]^2 \\
&\quad + 2E[\hat{\theta}_i(Y_i, \hat{A}) - \hat{\theta}_i(Y_i, A)][\hat{\theta}_i(Y_i, A) - \theta_i].
\end{aligned}
$$

where $\hat{\theta}_i(Y_i, A)$ is the BLUP of θ_i.

A second-order expansion to MSPE of EBLUP $\hat{\theta}_i$ is given by

$$
\begin{aligned}
& AMSPE_i \\
= \; & g_{1i}(A) + g_{2i}(A) + g_{3i}(A, \kappa_v) + 2g_{4i}(A, \kappa_v) \\
= \; & \frac{AD_i}{A + D_i} + \frac{D_i^2}{(A + D_i)^2} var[\hat{\beta}(A)] \\
& + \frac{D_i^2}{(A + D_i)^3} var(\hat{A}) + \frac{2AD_i^2}{m(A + D_i)^3} \left[D_i \kappa_{ei} - A\kappa_v \right] c(\hat{A}; A) \\
= \; & AMSPE_{i,N} + \frac{D_i^2}{(A + D_i)^3} \eta(\hat{A}; A, \kappa_v) + 2g_{4i}(A, \kappa_v), .
\end{aligned}
$$

where $AMSPE_{i,N}$ is the normality-based MSPE approximation as given in Prasad and Rao (1990) and Datta, Rao and Smith (2005).

Comments

- The term $g_{3i}(A, \kappa_v)$ is the additional uncertainty due to the estimation of the variance component A and the term $2g_{4i}(A, \kappa_v)$ is needed to adjust for the non-normality.
- Under the regularity conditions, $g_{1i}(A)$ is the leading term [of order $O(1)$] and the remaining terms are all of order $O(m^{-1})$.
- Note that non-normality affects both $var(\hat{A})$ and the cross-product term $2E[\hat{\theta}_i(\hat{A}, Y) - \hat{\theta}_i(A, Y)][\hat{\theta}_i(A, Y) - \theta_i]$.
- When both $\{e_i\}$ and $\{v_i\}$ are normal, the above approximation reduces to the Prasad-Rao (1990) approximation when $\hat{A} = \hat{A}_{PR}$ and the Datta-Rao-Smith (2005) approximation when $\hat{A} = \hat{A}_{FH}$.
- When the $\{e_i\}$ are normal and $\hat{A} = \hat{A}_{PR}$, the MSPE approximation reduces to the Lahiri-Rao (1995) approximation.

Estimation of small area proportion

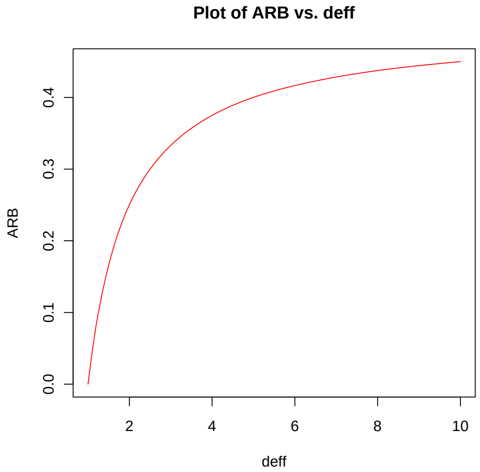
For $i = 1, \cdots, m$, assume
Level 1: (Sampling Model) $Y_i|\theta_i \overset{ind}{\sim} Bin(n_i, \theta_i)$:
Level 2: (Linking Model) $\theta_i \overset{iid}{\sim}$ Beta(α, β).

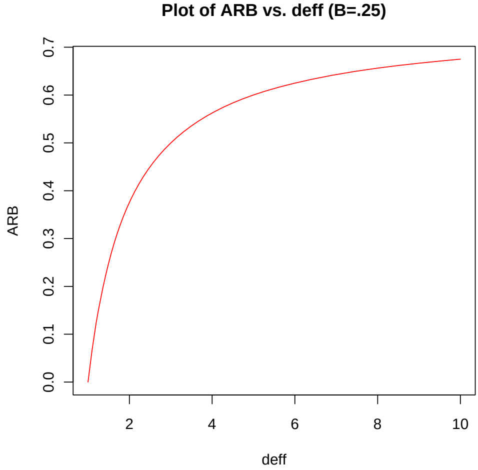
$\mu = \frac{\alpha}{\alpha+\beta}, \; A = \gamma\mu(1-\mu)$, with $gamma = \frac{1}{\alpha+\beta+1}$

- Y_i is the number of units favoring an event out of a sample size of n_i.
- Beta$[\mu, A]$ denotes a Beta distribution with mean μ and variance A
- Let us assume μ and A are known. So we drop the small area index i in the subsequent discussion.
- Under complex sampling, a more reasonable Level 1 might be Bin$(\tilde{n}, \theta)$, where $\tilde{n} = n/\delta$ and δ is the design effect.

Estimation of small area proportion

- The posterior distribution of θ, under misspecified model, is given by Beta$[\theta^B = (1-B)p + B\mu,\ v^B = \gamma\theta^B(1-\theta^B)]$, where $B = \frac{\alpha+\beta}{\alpha+\beta+n}$
- Under the complex sampling model, the posterior distribution of θ is given by Beta$[\tilde{\theta}^B = (1-\tilde{B})\tilde{p} + \tilde{B}\mu,\ v^B = \gamma\tilde{\theta}^B(1-\tilde{\theta}^B)]$, where $\tilde{B} = \frac{\alpha+\beta}{\alpha+\beta+\tilde{n}}$
- ARB $= \frac{E(\theta^B - \tilde{\theta}^B)}{\theta} = -(1-B)(1-\frac{1}{\delta})$
- $\frac{1}{\alpha+\beta+1}(1-\frac{1}{\delta}) < |ARB| < 1 - \frac{1}{\delta}$


Plot of ARB vs. deff
ARB
0.0
0.1
0.2
0.3
0.4
2
4
6
8
10
deff


Plot of ARB vs. deff (B=.25)
ARB
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
2
4
6
8
10
deff

Ref: Liu, Lahiri, Kalton (2007)

Model 1

For $i = 1, \cdots, m$, assume

Level 1: (Sampling Model) $p_i \mid \theta_i \overset{ind}{\sim} N[\theta_i, \theta_i(1-\theta_i)\delta_i];$

Level 2: (Linking Model) $h(\theta_i) \overset{ind}{\sim} N[x_i'\beta, A].$

Model 2

For $i = 1, \cdots, m$, assume

Level 1: (Sampling Model) $p_i \mid \theta_i \overset{ind}{\sim} Beta[\theta_i, \theta_i(1-\theta_i)\delta_i];$

Level 2: (Linking Model) $h(\theta_i) \overset{ind}{\sim} N[x_i'\beta, A].$

Some Comments

- Level 1 modeling could be problematic in the presence of sizable number of zeroes for small area.
- $\delta_i = \frac{Deff_i}{n_i} = \frac{\sum_h W_{ih}^2 \theta_{ih}(1-\theta_{ih})/n_{ih}}{\theta_i(1-\theta_i)}$,
  where $W_{ih} = N_{ih}/N_i,\ N_i = \sum_h N_{ih},\ n_i = \sum_h n_{ih}$.
- θ_{ih} is the population proportion for stratum h in area i .
- The design effect $Deff_i$ is a function of θ_{ih}, which are unknown.
- If $\theta_{ih} \approx \theta_i$, then $\delta_i \approx \sum_h W_{ih}^2/n_{ih}$.
- For complex designs, certain approximations of design effects are given in Kish (1987), Gabler, Häder and Lahiri (1999), Gabler, Ganninger, Lahiri (2011), Hawala and Lahiri (2010).