Empirical Bayes Small Area Estimation under Multiplicative Models

Malay Ghosh,Tatsuya Kubokawa and Yuki Kawakubo
University of Florida and University of Tokyo

September 3, 2013

Outline

- INTRODUCTION
- MULTIPLICATIVE MODELS
- LOSS FUNCTIONS AND ESTIMATORS
- RISK EVALUATION AND ESTIMATION OF RISK
- DATA ANALYSIS
- SIMULATION STUDY
- SUMMARY AND CONCLUSION

INTRODUCTION

- There is a worldwide growing demand for reliable small area (small domain) estimates.
- The need is felt both in the public and private sectors.
- There are multiple uses, for instance, allocation of government funds, regional planning strategies, marketing decisions at local level, and the list can go on and on.
- Some important current day examples include estimation of K-12 children under poverty at the county and lower levels of geography, per capita income for small places, unemployment rates for local areas, proportion of people without health insurance for small domains etc.

- The ‘direct’ survey estimates, based only on domain-specific sample data may be adequate for large domains or areas, but are usually accompanied with large standard errors and coefficients of variation for small domains, due to the smallness of small sizes.
- The simple reason is that the original survey was designed to achieve a targeted level of accuracy for larger areas or domains, and one does not have the resources to carry out new surveys for these small areas or domains to achieve a desired level of accuracy.
- This makes it a necessity to borrow strength, or connect different small areas through some models.
- These models provide the necessary link by bringing in relevant auxiliary information, often collected from multiple administrative sources.

- The current literature in small area estimation is dominated by the normality of errors as well as normality of random effects.
- There are a few exceptions, especially for the analysis of binary or count small area data, where generalized mixed models are mostly used.
- These assertions apply both to the frequentist and Bayesian analysis.
- However, in many surveys, the response consists of positive outcomes, such as income, revenue, harvest yield, production and many other similar quantites of interest.
- Their disributions are quite often positively skewed, and need suitable transformations for normality to hold.
- Log transformation is one of many such transformations.
- I will be considering throughout the log-normal model in this talk.

MULTIPLICATIVE MODELS

- Positive responses y_i, $i = 1, \ldots, m$.
- Multiplicative model: $y_i = \theta_i \eta_i$, $i = 1, \ldots, m$.
- The object of estimation are the θ_i, $i = 1, \ldots, m$.
- Log transformation: $z_i = \log(y_i) = \phi_i + e_i$,
  where $\phi_i = \log\theta_i$ and $e_i = \log\eta_i$.
- Hierarchical Model:
  $z_i|\phi_i, \beta \overset{ind}{\sim} \mathrm{N}(\phi_i, d_i)$; $\phi_i|\boldsymbol{\beta} \overset{ind}{\sim} \mathrm{N}(\mathbf{x}_i^T\boldsymbol{\beta}, \tau^2)$; $\beta \sim \mathrm{uniform}(R^p)$.
- $z_1, \ldots, z_m$ are mutually independent, $\phi_1, \ldots, \phi_m$ are mutually independent, the x_i are the p-variate vectors of covariates, and β is a p-variate vector of random regression coefficients $(p < m)$.
- For the moment, τ^2 will be assumed known. This will be relaxed later.
- d_i will be assumed known throught to avoid non-identifiability.

- Notations: $\mathbf{z} = (z_1, \ldots, z_m)^T$, $\theta = (\theta_1, \ldots, \theta_m)^T$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_m)^T$, $\mathbf{X}^T = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$, $\gamma_i = d_i/(d_i + \tau^2)$, $i = 1, \ldots, m$.
- $\boldsymbol{\Sigma} = \text{diag}(d_1 + \tau^2, \ldots, d_m + \tau^2)$.
- Assume rank$(\mathbf{X}) = p$, so that $\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}$ is nonsingular.
- $\hat{\boldsymbol{\beta}} \equiv \hat{\boldsymbol{\beta}}(\tau^2) = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{z}$.
- Posterior: $\phi_i | \mathbf{z}, \tau^2 \overset{ind}{\sim} (\hat{\phi}^B, k_i(\tau^2))$.
- $\hat{\phi}_i^B = (1 - \gamma_i) z_i + \gamma_i \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$, $k_i(\tau^2) = d_i(1 - \gamma_i) + \gamma_i^2 h_i$, $h_i = \mathbf{x}_i^T (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{x}_i$.
- Recall for a lognormal varaiable W, $E(W) = \exp(\mu + \frac{1}{2}\sigma^2)$.
- $\hat{\theta}_i^B = E[\exp(\phi_i) | \mathbf{z}, \tau^2] = \exp[\hat{\phi}_i^B + (1/2) k_i(\tau^2)]$.

LOSS FUNCTIONS AND ESTIMATORS

- Posterior mean $\hat{\boldsymbol{\theta}}^B = (\hat{\theta}_1^B, \ldots, \hat{\theta}_m^B)^T$ minimizes the posterior risk $E[\sum_{i=1}^m (\hat{\theta}_i - \theta_i)^2 | \mathbf{z}, \tau^2]$ with respect to the $\hat{\theta}_i$, $i = 1, \ldots, m$.
- This is also the BLUP of $\boldsymbol{\theta}$ for known τ^2.
- This is the most widely used estimator for small area estimation.
- However, for estimating positive quantities, there are other options.
- One such option is the Kullback-Leibler(KL) or the entropy loss.
- Stein (1964) used this loss for variance estimation.
- KL loss: $L_{KL}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \sum_{i=1}^m [\hat{\theta}_i/\theta_i - \log(\hat{\theta}_i/\theta_i) - 1]$.
- Minimization of $E[L_{KL}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) | \mathbf{z}, \tau^2]$ with respect to $\hat{\boldsymbol{\theta}}$ leads to the estimator $[E(\theta_i^{-1} | \mathbf{z}, \tau^2)]^{-1} = [E\{\exp(-\phi_i) | \mathbf{z}, \tau^2\}]^{-1}] = \exp[\hat{\phi}_i^B - (1/2) k_i(\tau^2)]$.

- An alternative is the weighted KL loss: $L_{WKL}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \sum_{i=1}^m \theta_i[\hat{\theta}_i/\theta_i - \log(\hat{\theta}_i/\theta_i) - 1] = \sum_{i=1}^m [\hat{\theta}_i - \theta_i - \theta_i \log(\hat{\theta}_i/\theta_i)]$.
- Minimization of $E[L_{WKL}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})|\mathbf{z}, \tau^2]$ with respect to $\hat{\boldsymbol{\theta}}$ leads to the estimator $\hat{\theta}_i^B$ of the θ_i.
- We prefer L_{WKL} over L_{KL} for a couple of reasons.
- First, it leads to the BLUP of $\boldsymbol{\theta}$ for known τ^2.
- Second, it is now possible to compare the risk of the same estimator $\hat{\theta}_i^B$ under the two losses-squared error and L_{WKL}.

- The loss L_{WKL} has a certain edge over the squared error loss under one particular benchmarking formulation.
- Suppose we want to compare the weighted squared eror loss $L_W(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \sum_{i=1}^m w_i(\hat{\theta}_i - \theta_i)^2$ with $L_{WWKL}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \sum_{i=1}^m w_i[\hat{\theta}_i - \theta_i - \log(\hat{\theta}_i/\theta_i)]$, where the w_i are known weights, not depending on the θ_i.
- Then $\hat{\theta}_i^B$ continues to be the Bayes estimator of θ_i under both losses.
- Suppose now we impose the benchmarking constraint $\sum_{i=1}^m w_i\hat{\theta}_i = M(\mathbf{y})$, where $M(\mathbf{y}) = \sum_{j=1}^m w_j y_j$ or even some other function of $\mathbf{y}$.
- Then the constrained Bayes estimator of θ_i under L_W is $\hat{\theta}_i^B - \sum_{j=1}^m w_j\hat{\theta}_j^B + M(\mathbf{y})$, which can even take negative values.
- But the constrained Bayes estimator of θ_i under L_{WWKL} is $\hat{\theta}_i^B[M(\mathbf{y})/\sum_{j=1}^m w_j\hat{\theta}_j^B]$, which is guaranteed to be positive as long as $M(\mathbf{y})$ is positive.

- Empirical Bayes estimation of the θ_i needs estimation of the unknown τ^2.
- Possible approaches: (a) the iterative method of moments estimator by Fay and Herriot (1979) and Morris (1983), (b) the Prasad-Rao (1990) method of moment estimator and (c) the ML and REML estimators proposed by Datta and Lahiri (2000).
- (a) and (c) do not provide closed form estimators, but (b) does.
- The Prasad-Rao estimator of τ^2 is $\hat{\tau}^2 = \max[0, (m-p)^{-1} \sum_{j=1}^{m} \{(z_j - \mathbf{x}_j^T \tilde{\boldsymbol{\beta}})^2 - d_j(1-h_j)\}]$.
- $\tilde{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{z}$.
- The EB estimator of θ_i is $\hat{\theta}_i^{EB} = \exp[\hat{\phi}^B(\hat{\tau}^2) + (1/2)k_i(\hat{\tau}^2)]$.

- An alternate approach is to originally assume both $\boldsymbol{\beta}$ and τ^2 to be known and find the Bayes estimator $\tilde{\theta}_i = E(\theta_i|\boldsymbol{\beta}, \tau^2, \mathbf{z}) = \exp[(1-\gamma_i)z_i + \gamma_i \mathbf{x}_i^T \boldsymbol{\beta} + (1/2)d_i(1-\gamma_i)]$.
- Then find the EB estimator of θ_i by substituting estimators of $\boldsymbol{\beta}$ and τ^2.
- We prefer the present approach since $\hat{\theta}_i^B = E(\theta_i|\mathbf{z}, \tau^2) = E[E(\theta_i|\boldsymbol{\beta}, \tau^2, \mathbf{z})|\tau^2, \mathbf{z}] = E[\tilde{\theta}_i|\tau^2, \mathbf{z}]$.
- By the Rao-Blackwell theorem, $\hat{\theta}_i^B$ has smaller risk than that of $\tilde{\theta}_i$ for any convex loss, while maintaining the same marginal exectation as that of the latter.
- Both L_{WKL} and squared error loss are convex in $\hat{\theta}$.

RISK EVALUATION AND ESTIMATION OF RISK

- Our objective is to find a second order correct (namely, correct up to $O(m^{-1})$ expression for the risk of $\hat{\theta}_i^{EB}$ under the loss $L_i = \hat{\theta}_i^{EB} - \theta_i - \theta_i \log(\hat{\theta}_i^{EB}/\theta_i)$.
- Assumptions: (A1) $\mathbf{X}^T\mathbf{X}/m$ converges to a positive definite matrix;
  (A2) $\max_{1 \le i \le m} \mathbf{x}_i^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i = O(m^{-1})$;
  (A3) $0 < d_L \le \min_{1 \le i \le m} d_i \le \max_{1 \le i \le m} d_i \le d_U < \infty$.
  (A4) $\hat{\tau}^2(\mathbf{z}) = \hat{\tau}^2(\mathbf{z} - \mathbf{X}\boldsymbol{\alpha})$ for all $\mathbf{z}$ and α ;$\hat{\tau}^2(\mathbf{z}) = \hat{\tau}^2(-\mathbf{z})$ for all $\mathbf{z}$.
  (A5) $\hat{\tau}^2 - \tau^2 = O_p(m^{-1/2})$.
- Notation: $s_i = \exp[\mathbf{x}_i^T\boldsymbol{\beta} + (1/2)\tau^2]$.
- Recall also the notations $\gamma_i = d_i/(d_i + \tau^2)$, $h_i = \mathbf{x}_i^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i$.
- $\hat{\phi}_i^B = (1-\gamma_i)z_i + \gamma_i \mathbf{x}_i^T \hat{\boldsymbol{\beta}}(\tau^2)$; $k_i(\tau^2) = d_i(1-\gamma_i) + \gamma_i^2 h_i$.

Theorem. Under the assumptions (A1)-(A5), and the loss L_i, the risk of the estimator $\hat{\theta}_i^B$ of θ_i is

$$
\begin{aligned}
R_{\boldsymbol{\beta},\tau^2}(\hat{\theta}_i^{EB}) &= E_{\boldsymbol{\beta},\tau^2} L_i \\
&= \frac{1}{2} d_i(1-\gamma_i)s_i + (1/2)s_i h_i[1+\tau^2\gamma_i(1/2-\gamma_i)] \\
&+ (\gamma_i^4/2)\exp\{2\mathbf{x}_i^T\boldsymbol{\beta} + d_i(1-\gamma_i)\} \\
&\times E(\hat{\tau}^2-\tau^2)^2\exp\{2(z_i-\mathbf{x}_i^T\hat{\boldsymbol{\beta}})\}(\frac{z_i-\mathbf{x}_i^T\hat{\boldsymbol{\beta}}}{d_i}+1)^2] \\
&+ O(m^{-3/2}).
\end{aligned}
$$

- Outline of Proof of Theorem 1:
  Since $E(\theta_i) = E[E(\theta_i|\tau^2, \mathbf{z})] = E(\hat{\theta}_i^B)$,
  $E[\hat{\theta}_i^{EB} - \theta_i - \theta_i \log(\hat{\theta}_i^{EB}/\theta_i)] =$
  $E[\hat{\theta}_i^{EB} - \hat{\theta}_i^B + \theta_i \log(\theta_i) - \hat{\theta}_i^B \log(\hat{\theta}_i^B) - \hat{\theta}_i^B \log(\hat{\theta}_i^{EB}/\hat{\theta}_i^B)]$.
- We need the fact that if $W \sim \mathrm{N}(\mu, \sigma^2)$, then $E[W\exp(W)] = (\mu + \sigma^2)\exp(\mu + \frac{1}{2}\sigma^2)$.
- Then

$$\begin{aligned} E(\theta_i \log(\theta_i)] &= E[\phi\exp(\phi_i)] = E[E\{\phi\exp(\phi_i)|\mathbf{z}\}] \\ &= E[\{\hat{\phi}_i^B + k_i(\tau^2)\}\exp(\hat{\phi}_i^B + (1/2)k_i(\tau^2))] \end{aligned}$$

- $E[\hat{\theta}_i^B \log(\hat{\theta}_i^B)] = E[(\hat{\phi}_i^B + (1/2)k_i(\tau^2))\exp(\hat{\phi}_i^B + (1/2)k_i(\tau^2))]$.
- $E(\theta_i \log\theta_i - \hat{\theta}_i^B \log\hat{\theta}_i^B) =$
  $(1/2)k_i(\tau^2)\exp\{(1/2)k_i(\tau^2)\}E[\exp(\hat{\phi}_i^B)] =$
  $(1/2)d_i(1-\gamma_i)s_i + (1/2)s_i h_i[1 + d_i(1-\gamma_i)(1/2-\gamma_i)]$.
- $E[\hat{\theta}_i^B \log(\hat{\theta}_i^{EB}/\hat{\theta}_i^B)]$ leads to the third term.

- The first term in the right hand side of the expression for $R_{\boldsymbol{\beta},\tau^2}(\hat{\theta}_i^{EB})$ is $O(1)$.
- The remaining terms are of $O(m^{-1})$.
- For the terms which are $O(m^{-1})$, one needs to plug in estimators for $\boldsymbol{\beta}$ and τ^2.
- The bias of these plugged in estimators is $O(m^{-3/2})$.
- Thus the problem reduces to find an estimator of $u_i(\tau^2) = (1/2)d_i(1-\gamma_i)\exp(\mathbf{x}_i^T\boldsymbol{\beta} + (1/2)\tau^2)$ whose bias is of the order $O(m^{-3/2})$.
- First step is the Taylor expansion
  $u_i(\hat{\tau}^2) =$
  $u_i(\tau^2) + (\hat{\tau}^2 - \tau^2)u_i'(\tau^2) + \frac{1}{2}(\hat{\tau}^2 - \tau^2)^2 u_i''(\tau^2) + O(m^{-3/2})$.
- After much simplification, the required bias corrected estimator of $u_i(\tau^2)$ is $u_i(\hat{\tau}^2)+$ A VERY LONG EXPRESSION of $O(m^{-1})$.

DATA ANALYSIS

- Apply the empirical Bayes (EB) estimator to the data in the Survey of Family Income and Expenditure (SFIE) in Japan.
- Also we investigate the performance of the second-order unbiased estimators of the risks of EB.
- In this study, we use the data of the disbursement item "Education" in the survey in November, 2011.
- The average disbursement (scaled by 1,000 Yen) at each capital city of 47 prefectures in Japan is denoted by y_i for $i = 1, \ldots, 47$.
- The variance is based on data of the disbursement "Education" at the same city every November in the past ten years.
- Although the average disbursements in SFIE are reported every month, the sample sizes are around 100 for most prefectures, and data of the item "Education" have high variability.

- On the other hand, we have data in the National Survey of Family Income and Expenditure (NSFIE) for 47 prefectures.
- Since NSFIE is based on much larger sample than SFIE, the average disbursements in NSFIE are more reliable, but this survey is conducted only every 5 years.
- In this study, we use the log-transformed data of the item "Education" of NSFIE in 2009, which is denoted by x_i for $i = 1, \ldots, 47$.
- We use the multiplicative model for the y_i, and work with the $z_i = \log y_i$ as mentioned earlier.
- For τ^2, we used the iterative Fay-Herriot estimator which yields $\hat{\tau}^{2FH} = 0.052$ in this example.
- The resulting EB estimator of θ_i is denoted by EB.
- Second order asymptotic unbiased MSE estimatotrs are denoted by R_{EB} while parametric bootstrap estimators are denoted by R^*_{EB}.

- Among 47 prefectures in Japan, we select the seven prefectures in the Kanto region around Tokyo, and their values of n_i, y_i, EB, R_{EB} and R^*_{EB} are reported in the following table.
- EB shrinks y_i more for relatively smaller n_i.
- The risk estimates R_{EB} and R^*_{EB} are close to each other for all perfectures.

Table1. Values of y_i, EB and the risk estimators for EB

Prefecture	n_i	y_i	EB	R_{EB}	R^*_{EB}
Ibaraki	95	8.10	8.89	19.46	19.48
Tochigi	95	10.03	9.48	23.59	23.92
Gunma	94	5.21	7.71	19.66	19.61
Saitama	95	12.33	12.73	19.90	19.78
Chiba	94	30.71	13.10	32.12	32.11
Tokyo	386	15.45	14.45	12.39	12.54
Kanagawa	142	15.99	23.25	14.06	29.58

SIMULATION STUDY

- Framework of Datta, Rao and Smiith (Biometrika, 2005).
- There are five groups $G_1, \ldots, G_5$ and three small areas in each group.
- The sampling variances d_i are the same for areas within the same group.
- Transformed Fay-Herriot Model with $m = 15$, $\tau^2 = 1$ and two d_i patterns.
- (a): $d_i = .7, .6, .5, .4, .3$; (b) $d_i = 2.0, 0.6, 0.5, 0.4, 0.2$.
- Intercept model: the model only has a constant term.
- Simulated MSE of y_i and $\hat{\theta}_i^{EB}$ are based on 100,000 simulations.
- The relative bias (Rbias) and the relative root mean squared error (RMSE) of the $\hat{R}_i$, the estimator of the MSE are also calculated.

Table 2. Simulated MSE's, MSE Estimates, Relative Biases and Relative MSE's for Pattern (i) using Fay-Herriot Estimates of τ^2.

i	d_i	y_i	$\hat{\theta}_i^{EB}$	$\hat{R}_i$	Rbias	RMSE
1	0.7	1.87	1.11	1.08	-.03	.30
2	0.6	1.56	0.99	0.98	-.01	.28
3	0.5	1.26	0.88	0.86	-.02	.26
4	0.4	0.99	0.74	0.73	-.01	.26
5	0.3	0.73	0.59	0.58	-.02	.23

Table 3. Simulated MSE's, MSE Estimates, Relative Biases and Relative MSE's for Pattern (ii) using Fay-Herriot Estimates of τ^2.

i	d_i	y_i	$\hat{\theta}_i^{EB}$	$\hat{R}_i$	Rbias	RMSE
1	2.0	7.81	2.04	1.83	-.10	.47
2	0.6	1.56	1.01	0.98	-.03	.30
3	0.5	1.27	0.88	0.87	-.01	.27
4	0.4	1.00	0.75	0.74	-.01	.25
5	0.2	0.47	0.41	0.42	- 02	.22

SUMMARY AND CONCLUSION

- We have considered the problem of estimating positive quantities with the transformed Fay-Herriot multiplicative model.
- EB estimators using weighted Kullback-Leibler loss function are found.
- Second-order approximation of the risk of the EB estimators under the weighted Kullback-Leibler loss is calculated.
- Second-order unbiased estimator of the risk via two approaches, namely, the analytical method based on the Taylor series expansion and the parametric bootstrap are found.
- The performance is compared both through simulation and a numerical example.
- As future projects, one can extend the results to the unit-level multiplicative models and to the benchmark issues.
- It is also of interest to consider the problem of constructing confidence intervals of the θ_i with second-order accuracy.