Prediction in Heteroscedastic Nested Error Regression Models with Random Dispersions

Tatsuya Kubokawa[*], Shonosuke Sugasawa[†], Malay Ghosh[‡], and Sanjay Chaudhuri[§]
University of Tokyo, University of Florida and National University of Singapore

January 5, 2015

Abstract

The paper concerns small-area estimation in the heteroscedastic nested error regression (HNER) model which assumes that the within-area variances are different among areas. Although HNER is useful for analyzing data where the within-area variation changes from area to area, it is difficult to provide good estimates for the error variances because of small samples sizes for small-areas. To fix this difficulty, we suggest a random dispersion HNER model which assumes a prior distribution for the error variances. The resulting Bayes estimates of small area means provide stable shrinkage estimates even for small sample sizes. Next we propose an empirical Bayes procedure for estimating the small area means. For measuring uncertainty of the empirical Bayes estimators, we use the conditional and unconditional mean squared errors (MSE) and derive their second-order approximations. It is interesting to note that the difference between the two MSEs appears in the first-order terms while the difference appears in the second-order terms for classical normal linear mixed models. Second-order unbiased estimators of the two MSEs are given with an application to the posted land price data.

Key words and phrases: Asymptotic approximation, conditional mean squared error, empirical Bayes, parametric bootstrap, second-order approximation, second-order unbiased estimate, small area estimation.

1 Introduction

Linear mixed (LM) models and the model-based estimators including empirical Bayes estimator (EB) or empirical best linear unbiased predictor (EBLUP) have been studied quite extensively in the literature from both theoretical and applied points of view. For a good review and account on this topic, see Ghosh and Rao (1994), Pfeffermann (2002), Rao (2003) and Datta (2009). Of these, the nested error regression (NER) model introduced by Battese, Harter and Fuller (1988) has been used as a unit-level model. The NER model with m small-areas assumes that the data $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^T$ are taken from the i-th small-area, $i = 1, \ldots, m$, where $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ are mutually

[*]Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN. E-Mail: tatsuya@e.u-tokyo.ac.jp
[†]Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, E-Mail: shonosuke622@gmail.com
[‡]Department of Statistics, University of Florida, 102 Griffin-Floyd Hall, Gainesville, Florida 32611. E-Mail: ghoshm@stat.ufl.edu
[§]Department of Statistics and Applied Probability, National University of Singapore, Block S16, Level 7, 6 Science Drive 2, SINGAPORE 117546. E-Mail: stasc@nus.edu.sg

independent. It is further assumed that y_{ij} is normally distributed with $E[y_{ij}] = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}$, $Var(y_{ij}) = \sigma_y^2$ and $Corr(y_{ij}, y_{ik}) = \rho$, $j \neq k$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$, σ_y^2 and ρ are unknown parameters, $\boldsymbol{x}_{ij}$'s are known vectors of covariates, and $Corr(y_{ij}, y_{ik})$ denotes the correlation coefficient of y_{ij} and y_{ik}.

The NER model can be expressed as a random effects model with

$$y_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + v_i + \varepsilon_{ij}, \quad i = 1, \ldots, m;\ j = 1, \ldots, n_i, \tag{1.1}$$

where v_i's and ε_{ij}'s are mutually independent with $v_i \sim \mathcal{N}(0, \lambda\sigma^2)$ and $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. Then $Var(v_i)/Var(\varepsilon_{ij}) = \lambda$, and that σ_y^2 and ρ correspond to

$$\sigma_y^2 = (1 + \lambda)\sigma^2 \quad \text{and} \quad \rho = \lambda/(1 + \lambda).$$

Jiang and Nguyen (2012) illustrated that the within-area sample variances change dramatically from small-area to small-area for the data given in Battese, *et al.* (1988). Figure 1, given in Section 5 in this paper, also indicates variability of the within-area variances. Jiang and Nguyen (2012) assumed that the variance $E[(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta})^2]$ is proportional to σ_i^2, which depends on the area i. Since $E[(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta})^2] = E[(v_i + \varepsilon_{ij})^2] = Var(v_i) + Var(\varepsilon_{ij})$, this assumption implies that $Var(v_i) + Var(\varepsilon_{ij}) = C\sigma_i^2$ for some constant C. If we let $Var(\varepsilon_{ij}) = \sigma_i^2$, we can see that $\{Var(v_i)/Var(\varepsilon_{ij}) + 1\}Var(\varepsilon_{ij}) = \{Var(v_i)/Var(\varepsilon_{ij}) + 1\}\sigma_i^2 = C\sigma_i^2$, which means that

$$Var(v_i)/Var(\varepsilon_{ij}) = C - 1.$$

That is, $Var(v_i)/Var(\varepsilon_{ij})$ is a constant.

Using the same notation as in the NER model, we write $Var(v_i)/Var(\varepsilon_{ij}) = \lambda$. Thus, the heteroscedastic nested error regression (HNER) model suggested by Jiang and Nguyen (2012) is the model given in (1.1) with

$$Var(v_i) = \lambda\sigma_i^2 \quad \text{and} \quad Var(\varepsilon_{ij}) = \sigma_i^2. \tag{1.2}$$

In the HNER model, Jiang and Nguyen (2012) showed that the maximum likelihood (ML) estimators of $\boldsymbol{\beta}$ and λ are consistent for large m and that the resulting empirical Bayes (EB) estimator of $\xi_i = \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta} + v_i$ ($\overline{\boldsymbol{x}}_i = n_i^{-1}\sum_{j=1}^{n_i}\boldsymbol{x}_{ij}$) estimates the Bayes estimator consistently since the Bayes estimator does not depend on $\sigma_1^2, \ldots, \sigma_m^2$, but on $\boldsymbol{\beta}$ and λ. This is quite interesting, because the number of unknown variances σ_i^2's increases as m tends to infinity. However, the posterior variance of v_i given $(y_{i1}, \ldots, y_{i,n_i})$ is

$$Var(v_i|y_{i1}, \ldots, y_{i,n_i}) = \sigma_i^2\lambda/(1 + n_i\lambda),$$

which implies that the mean squared error (MSE) of the EB estimator of ξ_i depends on σ_i^2. Then, we need to estimate σ_i^2 for estimating the MSE of the EB estimator of ξ_i. Since the sample sizes n_i is small in the small-area estimation, we cannot provide good estimates for σ_i^2 with reasonable precision.

In this paper, we propose a random dispersion HNER (RHNER) model which assumes that the areawise precisions $(\sigma_i^2)^{-1}$, $i = 1, \ldots, m$, are mutually independent gamma random variables. The resulting Bayes estimator of σ_i^2 gives stable shrinkage estimates of small area means even for $n_i - p = 0$.

For measuring uncertainty of the empirical Bayes estimator $\hat{\xi}_i^{EB}$ of ξ_i, we use the conditional and unconditional mean squared errors (MSE) defined by

$$\begin{aligned} cMSE(\omega; \hat{\xi}_i^{EB}|\boldsymbol{y}_i) =& E[(\hat{\xi}_i^{EB} - \xi_i)^2|\boldsymbol{y}_i], \\ MSE(\omega; \hat{\xi}_i^{EB}) =& E[(\hat{\xi}_i^{EB} - \xi_i)^2]. \end{aligned}$$

When data of the small area of interest are observed as $\boldsymbol{y}_i$ and one wants to know the prediction error of the EB estimators based on these data, the conditional mean squared error (cMSE) given $\boldsymbol{y}_i$ is used instead of the conventional unconditional MSE. Booth and Hobert (1998) showed that the difference between the cMSE and MSE is quite small and appears in the second-order terms in classical normal linear mixed models. In this article, however, we show that the difference appears in the leading or the first-order terms for the RHNER model.

2 HNER Models with Random Dispersions

2.1 Setup of models and predictors

We begin with the model given in (1.1) and (1.2). For stable estimators of σ_i^2's, we need sufficient amount of data from each area. Since n_i's are typically small, σ_i^2 cannot usually be estimated with reasonable precision. To give more stable estimators for σ_i^2, we assume a prior distribution for σ_i^2. Let $\eta_i = 1/\sigma_i^2$. It is assumed that $\eta_1, \ldots, \eta_m$ are independent and identically distributed with common pdf

$$\pi(\eta_i|\tau_1, \tau_2) \sim \mathcal{G}a(\tau_1/2, 2/\tau_2), \tag{2.1}$$

a gamma distribution with mean τ_1/τ_2 and variance $2\tau_1/\tau_2^2$. The HNER model given in (1.1) and (1.2) with the random dispersion (2.1) is called a *Random Heteroscedastic Nested Error Regression* (RHNER) model.

Let $\boldsymbol{y} = (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_m^T)^T$, $\boldsymbol{X}_i^T = (\boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,n_i})$ and $\boldsymbol{X} = (\boldsymbol{X}_1^T, \ldots, \boldsymbol{X}_m^T)$. All the unknown parameters are denoted by $\omega = (\boldsymbol{\beta}, \lambda, \boldsymbol{\tau})$ for $\boldsymbol{\tau} = (\tau_1, \tau_2)$. Then, the RHNER model is given by

$$\begin{aligned} \boldsymbol{y}_i|v_i, \eta_i \sim& \mathcal{N}_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{j}_{n_i}v_i, \eta_i^{-1}\boldsymbol{I}_{n_i}), \\ v_i|\eta_i \sim& \mathcal{N}(0, \lambda\eta_i^{-1}), \\ \eta_i \sim& \mathcal{G}a(\tau_1/2, 2/\tau_2). \end{aligned} \tag{2.2}$$

The conditional distribution of v_i given $\boldsymbol{y}_i$ and η_i is $\mathcal{N}(\hat{v}_i, \lambda\eta_i^{-1}/(n_i\lambda + 1))$, where

$$\hat{v}_i = \hat{v}_i(\boldsymbol{\beta}, \lambda) = \frac{n_i\lambda}{n_i\lambda + 1}(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta}). \tag{2.3}$$

It is noted that $\hat{v}_i = E[v_i|\boldsymbol{y}_i]$ does not depend on η_i or σ_i^2.

In this paper, we consider the problem of predicting the mixed quantity

$$\xi_i = \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta} + v_i, \quad i = 1, \ldots, m.$$

The conditional expectation of ξ_i given $\boldsymbol{y}_i$ and η_i is

$$\hat{\xi}_i^B(\boldsymbol{\beta}, \lambda) = E[\xi_i|\boldsymbol{y}_i, \sigma_i^2] = \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta} + \hat{v}_i(\boldsymbol{\beta}, \lambda).$$

This is interpreted as the Bayes estimator of ξ under squared error loss. Since it does not depend on η_i, the estimator $\hat{\xi}_i^B(\boldsymbol{\beta}, \lambda)$ continues to be the conditional expectation of ξ_i given $\boldsymbol{y}_i$ after integrating out the η_i, that is the Bayes estimator of ξ_i is the same in the two situations. However, the empirical Bayes estimators, which substitute estimators of $\boldsymbol{\beta}$ and λ into $\hat{\xi}_i^B(\boldsymbol{\beta}, \lambda)$, are different between the HNER and RHNER models.

In the HNER model, we need to estimate $(m + p + 1)$ parameters $\boldsymbol{\beta}$, λ and $\sigma_1^2, \ldots, \sigma_m^2$. Noting that the number of parameters increases as m increases and that n_is are bounded in small-area estimation, we are faced with the problem of consistency and instability of the estimators of σ_i^2. In this situation, Jiang and Nguyen (2012) established the surprising result that the MLEs of $\boldsymbol{\beta}$ and λ are consistent, which lead to the consistency of the EB estimator $\hat{\xi}_i^B(\widehat{\boldsymbol{\beta}}, \hat{\lambda})$. However, there are no consistent estimators of the σ_i^2. This problem can be fixed if instead the RHNER model is used. In fact, the parameters we need to estimate in the RHNER model are $\boldsymbol{\beta}$, λ, τ_1 and τ_2, and we can provide their consistent estimators.

2.2 A motivation from estimation of dispersions

We give more detailed motivation from the estimation of the dispersion parameters σ_i^2 in the HNER model. We first treat the simple case that $\boldsymbol{\beta} = \mathbf{0}$ and $n_1 = \cdots = n_m = n$ in (1.1). Let $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_m^2)^T$ and $\gamma = 1/(1 + n\lambda)$. It follows from the equation (4) of Jiang and Nguyen (2012) that the log-likelihood is then

$$L^H(\gamma, \boldsymbol{\sigma}^2) = \sum_{i=1}^m \Big[-n \log \sigma_i^2 + \log \gamma - \Big\{ \sum_{j=1}^n (y_{ij} - \overline{y}_i)^2 + n\gamma \overline{y}_i^2 \Big\} / \sigma_i^2 \Big] + K,$$

where K is a generic constant. Differentiating $L^H(\gamma, \boldsymbol{\sigma}^2)$ with respect to γ and σ_i^2's, we see that the maximum likelihood (ML) estimators, $\hat{\gamma}^H$ and $\hat{\sigma}_{(H)i}^2$, of γ and σ_i^2's are solutions of the equations

$$\begin{aligned} \hat{\gamma}^H =& \frac{m}{\sum_{i=1}^m n\overline{y}_i^2/\hat{\sigma}_{(H)i}^2}, \\ \hat{\sigma}_{(H)i}^2 =& \frac{1}{n} \Big\{ \sum_{j=1}^n (y_{ij} - \overline{y}_i)^2 + n\hat{\gamma}^H \overline{y}_i^2 \Big\}. \end{aligned} \tag{2.4}$$

It is interesting to point out that $\hat{\sigma}_{(H)i}^2$ is an asymptotically unbiased estimator of σ_i^2. For the proof, we can use the fact that $n\overline{y}_i^2$ and $\sum_{j=1}^n (y_{ij} - \overline{y}_i)^2$ are mutually independent with $n\overline{y}_i^2/(1+n\lambda) \sim \sigma_i^2 \chi_1^2$ and $\sum_{j=1}^n (y_{ij} - \overline{y}_i)^2 \sim \sigma_i^2 \chi_{n-1}^2$. Then, from the consistency of $\hat{\gamma}^H$, it follows that $E[\hat{\sigma}_{(H)i}^2]$ converges to

$$\frac{1}{n} E\Big[\sum_{j=1}^n (y_{ij} - \overline{y}_i)^2 + n\gamma \overline{y}_i^2 \Big] = \frac{\sigma_i^2}{n} E[\chi_{n-1}^2 + \gamma(1 + n\lambda)\chi_1^2] = \sigma_i^2,$$

which shows that $\hat{\sigma}_{(H)i}^2$ is an asymptotically unbiased estimator of σ_i^2.

Although $\hat{\sigma}_{(H)i}^2$ is asymptotically unbiased, it is clear that $\hat{\sigma}_{(H)i}^2$ is not consistent when $m \to \infty$, but n is bounded. Thus, we need to modify $\hat{\sigma}_{(H)i}^2$ when n is small. For example, we look at the empirical Bayes estimator of ξ_i. In the simple case we treat here, we have $\xi_i = v_i$, and the EB estimator of ξ_i is given by

$$\hat{\xi}_i^H = (1 - \hat{\gamma}^H)\overline{y}_i = \Big\{ 1 - \frac{m}{\sum_{i=1}^m n\overline{y}_i^2/\hat{\sigma}_{(H)i}^2} \Big\} \overline{y}_i,$$

from (2.4). This is a natural shrinkage estimator, and it is reasonable for large m since $\hat{\gamma}^H$ is consistent. When m is not large, however, there is a concern about the precision of the estimator $\hat{\sigma}_{(H)i}^2$. Since $\sum_{j=1}^n (y_{ij} - \overline{y}_i)^2 \le n\hat{\sigma}_{(H)i}^2 \le \sum_{j=1}^n y_{ij}^2$, it is seen that

$$\frac{\overline{y}_i^2}{\sum_{j=1}^n y_{ij}^2/n} \le \frac{\overline{y}_i^2}{\hat{\sigma}_{(H)i}^2} \le \frac{\overline{y}_i^2}{T_i/n},$$

for $T_i = \sum_{j=1}^n (y_{ij} - \overline{y}_i)^2$. When n is small, clearly $\overline{y}_i^2/\hat{\sigma}_{(H)i}^2$ has a large variation, which leads to the instability of the empirical Bayes estimator $\hat{\xi}_i^H$. Although the simple case of equal replications n is considered here, in the survey data we need to handle the case of small sample sizes n_i's for some small-areas, and the estimators $\hat{\sigma}_{(H)i}^2$'s would not have enough degrees of freedom.

To overcome this drawback, we need to stabilize $\hat{\sigma}_{(H)i}^2$ by shrinking it to a point. The random dispersion model is helpful for this purpose. Assume that $\eta_i = 1/\sigma_i^2$ has a distribution $\mathcal{G}a(\tau_1/2, 2/\tau_2)$. Since $T_i\eta_i = T_i/\sigma_i^2 \sim \chi_{n-1}^2$, from the joint distribution of (T_i, η_i), the posterior mean of σ_i^2 given T_i is

$$E[\sigma_i^2 | T_i] = (T_i + \tau_2)/(n - 1 + \tau_1).$$

When $\hat{\tau}_1$ and $\hat{\tau}_2$ are estimators of τ_1 and τ_2 based on the statistics $T_1, \ldots, T_m$, it is reasonable to estimate σ_i^2 by

$$\hat{\sigma}^2_{(RH)i} = (T_i + \hat{\tau}_2)/(n - 1 + \hat{\tau}_1).$$

Clearly, $\hat{\sigma}^2_{(RH)i}$ is more stable than the unbiased estimator $T_i/(n-1)$ when n is small. Replacing $\hat{\sigma}^2_{(H)i}$ in $\hat{\xi}_i^H$ with a shrinkage estimator like $\hat{\sigma}^2_{(RH)i}$, one can get the more stabilized empirical Bayes estimator

$$\hat{\xi}_i^{RH} = \Big\{1 - \frac{m}{\sum_{i=1}^m n\overline{y}_i^2/\hat{\sigma}^2_{(RH)i}}\Big\}\overline{y}_i.$$

Another need for a consistent estimator of σ_i^2 appears in evaluation of uncertainty of the empirical Bayes estimator $\hat{\xi}_i^H$. When the mean squared error is used for measuring the uncertainty, the MSE of $\hat{\xi}_i^H$, denoted by $E[(\hat{\xi}_i^H - \xi_i)^2]$ converges to

$$E[\mathrm{Var}(v_i|\boldsymbol{y}_i)] = \sigma_i^2\lambda/(1+n\lambda) = \sigma_i^2(1-\gamma)/n$$

for large m. In order to estimate the uncertainty of $\hat{\xi}_i^H$, we want to estimate the leading term of the MSE consistently. Since $\hat{\sigma}^2_{(H)i}$ is not consistent, however, we cannot provide any consistent estimator of the leading term in the MSE of $\hat{\xi}_i^H$ in the HNER model. This drawback is overcome in the RHNER model.

3 Predictors and Asymptotic Properties of MSE

3.1 MLE of parameters and the empirical Bayes estimator

We now return back to the RHNER model given in (2.2). When λ and $\boldsymbol{\beta}$ are known, the best predictor or the Bayes estimator of $\xi_i = \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta} + v_i$ is given by

$$\begin{aligned}\hat{\xi}_i^B =& \hat{\xi}_i^B(\boldsymbol{\beta}, \lambda) = E[\xi_i|\boldsymbol{y}_i] \\ =& \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta} + (1-\gamma_i)(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta}),\end{aligned} \tag{3.1}$$

where $\gamma_i = \gamma_i(\lambda) = 1/(n_i\lambda + 1)$. In our case since λ and $\boldsymbol{\beta}$ are unknown, we need to estimate them from the marginal distributions of the $\boldsymbol{y}_i$. We provide the maximum likelihood (ML) estimators for unknown parameters $\omega = (\boldsymbol{\beta}, \lambda, \tau_1, \tau_2)$.

The marginal likelihood of $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m)^T$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_m)^T$ after integrating out the full joint likelihood with respect to v_i's can be expressed as

$$\begin{aligned} f(\boldsymbol{y}, \boldsymbol{\eta}|\omega) =& \prod_{i=1}^m \Big\{ \frac{\eta_i^{n_i/2}}{(2\pi)^{n_i/2}\sqrt{n_i\lambda+1}} \exp\Big[-\frac{\eta_i}{2}\{\sum_{j=1}^{n_i}(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta})^2 - \frac{n_i^2\lambda}{n_i\lambda+1}(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2\}\Big] \\ & \times \pi(\eta_i|\tau_1, \tau_2)\Big\} \\ =& \prod_{i=1}^m \Big\{ \frac{\tau_2^{\tau_1/2}\eta_i^{(n_i+\tau_1)/2-1}2^{-(n_i+\tau_1)/2}}{\pi^{n_i/2}\Gamma(\tau_1/2)\sqrt{n_i\lambda+1}} \exp\Big[-\frac{\eta_i}{2}\{Q_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda) + \tau_2\}\Big]\Big\}, \end{aligned} \tag{3.2}$$

where

$$\begin{aligned} Q_i = Q_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda) =& \sum_{j=1}^{n_i}(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta})^2 - \frac{n_i^2\lambda}{n_i\lambda+1}(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2 \\ =& \sum_{j=1}^{n_i}\{(y_{ij} - \overline{y}_i) - (\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)^T\boldsymbol{\beta}\}^2 + n_i\gamma_i(\lambda)(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2, \end{aligned} \tag{3.3}$$

where $\gamma_i = \gamma_i(\lambda) = 1/(n_i\lambda + 1)$. Integrating out the joint density $f(\boldsymbol{y}, \boldsymbol{\eta}|\omega)$ in (3.2) with respect to $\boldsymbol{\eta}$ yields the marginal likelihood of $\boldsymbol{y}$ given by

$$f(\boldsymbol{y}|\omega) = \prod_{i=1}^{m} \Big\{ \frac{\tau_2^{\tau_1/2}\Gamma((n_i+\tau_1)/2)}{\pi^{n_i/2}\sqrt{n_i\lambda+1}\Gamma(\tau_1/2)} \{Q_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda) + \tau_2\}^{-(n_i+\tau_1)/2} \Big\}. \tag{3.4}$$

Let $L = L(\boldsymbol{\beta}, \lambda, \tau_1, \tau_2) = \log f(\boldsymbol{y}|\omega)$. Then,

$$\begin{aligned}2L =& -n_i \log \pi + m\tau_1 \log \tau_2 + 2\sum_{i=1}^{m} \psi(\frac{n_i+\tau_1}{2}) - 2m\psi(\frac{\tau_1}{2}) \\ & - \sum_{i=1}^{m} \log(n_i\lambda + 1) - \sum_{i=1}^{m} (n_i + \tau_1)\log(Q_i + \tau_2),\end{aligned}$$

where $\psi(a) = \log(\Gamma(a))$. Let $L_{\boldsymbol{\beta}}$, L_λ, L_{τ_1} and L_{τ_2} be the derivatives of L with respect to $\boldsymbol{\beta}$, λ, τ_1 and τ_2. Then,

$$\begin{aligned}2L_{\boldsymbol{\beta}} =& -\sum_{i=1}^{m} \frac{n_i+\tau_1}{Q_i+\tau_2}\frac{\partial Q_i}{\partial \boldsymbol{\beta}}, \\ 2L_\lambda =& -\sum_{i=1}^{m} \frac{n_i+\tau_1}{Q_i+\tau_2}\frac{\partial Q_i}{\partial \lambda} - \sum_{i=1}^{m} n_i\gamma_i, \\ 2L_{\tau_1} =& \sum_{i=1}^{m} \log\Big(\frac{\tau_2}{Q_i+\tau_2}\Big) + \sum_{i=1}^{m} \Big\{\psi'(\frac{n_i+\tau_1}{2}) - \psi'(\frac{\tau_1}{2})\Big\}, \\ 2L_{\tau_2} =& m\frac{\tau_1}{\tau_2} - \sum_{i=1}^{m} \frac{n_i+\tau_1}{Q_i+\tau_2},\end{aligned} \tag{3.5}$$

where $\partial Q_i/\partial\lambda = -n_i^2\gamma_i^2(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2$ for $\partial\gamma_i/\partial\lambda = -n_i\gamma_i^2$, and

$$\frac{\partial Q_i}{\partial \boldsymbol{\beta}} = -2\sum_{j=1}^{n_i} \{(y_{ij} - \overline{y}_i) - (\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)^T\boldsymbol{\beta}\}(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i) - 2n_i\gamma_i(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})\overline{\boldsymbol{x}}_i. \tag{3.6}$$

The MLEs of $\boldsymbol{\beta}$, λ, τ_1 and τ_2 are the solution of the simultaneous equations $L_{\boldsymbol{\beta}} = 0$, $L_\lambda = 0$, $L_{\tau_1} = 0$ and $L_{\tau_2} = 0$, and the MLEs are denoted by $\widehat{\boldsymbol{\beta}}$, $\hat{\lambda}$, $\hat{\tau}_1$ and $\hat{\tau}_2$. The empirical Bayes estimator of $\xi_i = \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta} + v_i$ is provided by

$$\hat{\xi}_i^{EB} = \hat{\xi}_i^B(\widehat{\boldsymbol{\beta}}, \hat{\lambda}) = \overline{\boldsymbol{x}}_i^T\widehat{\boldsymbol{\beta}} + (1 - \hat{\gamma}_i)(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\widehat{\boldsymbol{\beta}}), \tag{3.7}$$

where $\hat{\gamma}_i = \gamma_i(\hat{\lambda}) = 1/(n_i\hat{\lambda} + 1)$.

3.2 Asymptotic properties of MLE

To evaluate the mean squared errors of the empirical Bayes estimator $\hat{\xi}_i^{EB}$ asymptotically, we need to derive asymptotic variances and covariances of the MLE when m tends to infinity. To derive asymptotic properties of the MLE, we assume the following:

(A1) The sample sizes n_i's are bounded below and above as $\underline{n} \le n_i \le \overline{n}$ for constants $\underline{n}$ and $\overline{n}$. The elements of $\boldsymbol{X}$ are uniformly bounded, $\boldsymbol{X}^T\boldsymbol{X}$ is positive definite and $\boldsymbol{X}^T\boldsymbol{X}/m$ converges to a positive definite matrix.

Since their asymptotic variances and covariances are expressed by the Fisher information matrix, we begin by deriving the Fisher information. Let $\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ be the Fisher information matrix of $\boldsymbol{\beta}$. The Fisher information matrix of $\boldsymbol{\theta} = (\lambda, \tau_1, \tau_2)^T$ and the inverse are denoted by

$$\boldsymbol{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{pmatrix} I_{\lambda\lambda} & I_{\lambda\tau_1} & I_{\lambda\tau_2} \\ I_{\lambda\tau_1} & I_{\tau_1\tau_1} & I_{\tau_1\tau_2} \\ I_{\lambda\tau_2} & I_{\tau_1\tau_2} & I_{\tau_2\tau_2} \end{pmatrix} \quad \text{and} \quad \boldsymbol{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} = \begin{pmatrix} I^{\lambda\lambda} & I^{\lambda\tau_1} & I^{\lambda\tau_2} \\ I^{\lambda\tau_1} & I^{\tau_1\tau_1} & I^{\tau_1\tau_2} \\ I^{\lambda\tau_2} & I^{\tau_1\tau_2} & I^{\tau_2\tau_2} \end{pmatrix}.$$

Then, exact expressions of the Fisher information matrices can be derived in the following theorem. The proof is given in the Appendix.

Theorem 3.1 *The Fisher information of $\boldsymbol{\beta}$ is given by*

$$\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \frac{\tau_1}{\tau_2}\sum_{i=1}^m \frac{n_i+\tau_1}{n_i+\tau_1+2}\Big\{\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)^T + n_i\gamma_i\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i^T\Big\}.$$

Also, $\boldsymbol{I}_{\boldsymbol{\beta}\lambda} = \boldsymbol{0}$, $\boldsymbol{I}_{\boldsymbol{\beta}\tau_1} = \boldsymbol{0}$ and $\boldsymbol{I}_{\boldsymbol{\beta}\tau_2} = \boldsymbol{0}$. The elements of $2\boldsymbol{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ are given by

$$\begin{aligned}
2I_{\lambda\lambda} =& \sum_{i=1}^m \frac{(n_i+\tau_1-1)n_i^2\gamma_i^2}{n_i+\tau_1+2}, \quad & 2I_{\lambda\tau_1} =& -\sum_{i=1}^m \frac{n_i\gamma_i}{n_i+\tau_1}, \\
2I_{\lambda\tau_2} =& \frac{\tau_1}{\tau_2}\sum_{i=1}^m \frac{n_i\gamma_i}{n_i+\tau_1+2}, \quad & 2I_{\tau_1\tau_1} =& \frac{1}{2}\sum_{i=1}^m \{\psi''(\frac{\tau_1}{2}) - \psi''(\frac{n_i+\tau_1}{2})\}, \\
2I_{\tau_1\tau_2} =& -\frac{1}{\tau_2}\sum_{i=1}^m \frac{n_i}{n_i+\tau_1}, \quad & 2I_{\tau_2\tau_2} =& \frac{\tau_1}{\tau_2^2}\sum_{i=1}^m \frac{n_i}{n_i+\tau_1+2}.
\end{aligned}$$

It follows from Theorem 3.1 and assumption (A1) that $m^{-1}\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} = O(1)$ and $m^{-1}\boldsymbol{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} = O(1)$, and the limiting values of these quantities are away from zero. The following theorem is essential for approximating the MSE of $\hat{\xi}_i^{EB}$ asymptotically. The proof is given in the Appendix.

Theorem 3.2 *Assume condition* (A1). *Then, for $\widehat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\tau}_1, \hat{\tau}_2)^T$, it holds that*

$$\begin{aligned}
E[(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^T|\boldsymbol{y}_i] =& (\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}})^{-1} + O_p(m^{-3/2}), \\
E[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T|\boldsymbol{y}_i] =& (\boldsymbol{I}_{\boldsymbol{\theta}\boldsymbol{\theta}})^{-1} + O_p(m^{-3/2}), \\
E[(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T|\boldsymbol{y}_i] =& O_p(m^{-3/2})0
\end{aligned} \tag{3.8}$$

This implies that $\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}|\boldsymbol{y}_i = O_p(m^{-1/2})$ and $\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}|\boldsymbol{y}_i = O_p(m^{-1/2})$. Also, the conditional biases $E[\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}|\boldsymbol{y}_i] = O(m^{-1})$ and $E[\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}|\boldsymbol{y}_i] = O(m^{-1})$.

4 Measures of Uncertainty of the Empirical Bayes Estimator

4.1 Second-order approximation of the conditional and unconditional MSEs

We shall derive a second-order approximation of the MSE of the empirical Bayes (EB) estimator and its second-order unbiased estimator. Recall that we want to predict $\xi_i = \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta} + v_i$ with EB $\hat{\xi}_i^{EB} = \hat{\xi}_i^B(\widehat{\boldsymbol{\beta}}, \hat{\lambda}) = \overline{\boldsymbol{x}}_i^T\widehat{\boldsymbol{\beta}} + \hat{v}_i(\widehat{\boldsymbol{\beta}}, \hat{\lambda})$. For measuring uncertainty of EB, we use the conditional and unconditional mean squared errors (MSE) defined by

$$\begin{aligned}
cMSE(\omega; \hat{\xi}_i^{EB}|\boldsymbol{y}_i) =& E[(\hat{\xi}_i^{EB} - \xi_i)^2|\boldsymbol{y}_i], \\
MSE(\omega; \hat{\xi}_i^{EB}) =& E[(\hat{\xi}_i^{EB} - \xi_i)^2].
\end{aligned}$$

The conditional and unconditional MSEs can be decomposed as

$$\begin{aligned} cMSE(\omega; \hat{\xi}_i^{EB}|\boldsymbol{y}_i) =& E[\{\xi_i - \hat{\xi}_i^B(\boldsymbol{\beta}, \lambda)\}^2|\boldsymbol{y}_i] + E[\{\hat{\xi}_i^B(\widehat{\boldsymbol{\beta}}, \hat{\lambda}) - \hat{\xi}_i^B(\boldsymbol{\beta}, \lambda)\}^2|\boldsymbol{y}_i] \\ =& g_1^c(\omega|\boldsymbol{y}_i) + g_2^c(\omega|\boldsymbol{y}_i), \quad \text{(say)} \end{aligned} \tag{4.1}$$

and

$$\begin{aligned} MSE(\omega; \hat{\xi}_i^{EB}) =& E[\{\xi_i - \hat{\xi}_i^B(\boldsymbol{\beta}, \lambda)\}^2] + E[\{\hat{\xi}_i^B(\widehat{\boldsymbol{\beta}}, \hat{\lambda}) - \hat{\xi}_i^B(\boldsymbol{\beta}, \lambda)\}^2] \\ =& g_1(\omega) + g_2(\omega). \quad \text{(say)} \end{aligned} \tag{4.2}$$

The first term $g_1^c(\omega|\boldsymbol{y}_i)$ is the posterior variance of ξ_i given $\boldsymbol{y}$, namely,

$$g_1^c(\omega|\boldsymbol{y}_i) = Var(\xi_i|\boldsymbol{y}_i) = \frac{\lambda}{n_i\lambda + 1} E[\eta_i^{-1}|\boldsymbol{y}_i] = \frac{\lambda}{n_i\lambda + 1} \frac{Q_i + \tau_2}{n_i + \tau_1 - 2}, \tag{4.3}$$

where Q_i is given in (3.3). Similarly, $g_1(\omega)$ is given by

$$g_1(\omega) = E[Var(\xi_i|\boldsymbol{y}_i)] = \frac{\lambda}{n_i\lambda + 1} E[\eta_i^{-1}] = \frac{\lambda}{n_i\lambda + 1} \frac{\tau_2}{\tau_1 - 2}. \tag{4.4}$$

Noting that $g_1^c(\omega|\boldsymbol{y}_i) = O_p(1)$, $g_2^c(\omega|\boldsymbol{y}_i) = O_p(m^{-1})$, $g_1(\omega) = O(1)$ and $g_2(\omega) = O(m^{-1})$, we can see that the difference between the cMSE and MSE appears in the leading or the first-order terms. This is an interesting fact, because the difference is small and appears in the second-order terms in the classical normal theory mixed models as demonstrated by Booth and Hobert (1998). They also showed that the difference is significant and appears in the first-order terms for distributions far from normality. Noting that the random dispersion model (2.2) is not a normal distribution, but close to a t-distribution, we observe that the above fact coincides with their assertion.

In the case of the HNER model, $Var(\xi_i|\boldsymbol{y}_i)$ is identical to $E[Var(\xi_i|\boldsymbol{y}_i)]$ since y_i has a normal distribution, and is given by $\sigma_i^2\lambda/(n_i\lambda + 1)$. Thus, it should be noted that we cannot estimate the first-order term $\sigma_i^2\lambda/(n_i\lambda + 1)$ consistently in the HNER model, since n_i is bounded. However, we can estimate $g_1^c(\omega|\boldsymbol{y}_i)$ and $g_1(\omega)$ consistently in the RHNER model (2.2) since λ, τ_1 and τ_2 are estimated consistently.

Theorem 4.1 *Under assumption* (A1)*, the conditional MSE of* $\hat{\xi}_i^{EB}$ *is approximated as*

$$\begin{aligned} cMSE(\lambda, \boldsymbol{\tau}; \hat{\xi}_i^{EB}|\boldsymbol{y}_i) =& \frac{1 - \gamma_i}{n_i} \frac{Q_i + \tau_2}{n_i + \tau_1 - 2} + \gamma_i^2 \overline{\boldsymbol{x}}_i^T \left(\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}\right)^{-1} \overline{\boldsymbol{x}}_i \\ &+ n_i^2 \gamma_i^4 (\overline{y}_i - \overline{\boldsymbol{x}}_i^T \boldsymbol{\beta})^2 I^{\lambda\lambda} + O_p(m^{-3/2}), \end{aligned} \tag{4.5}$$

for $\gamma_i = 1/(n_i\lambda + 1)$*, and the unconditional MSE is approximated as*

$$\begin{aligned} MSE(\lambda, \boldsymbol{\tau}; \hat{\xi}_i^{EB}) =& \frac{1 - \gamma_i}{n_i} \frac{\tau_2}{\tau_1 - 2} + \gamma_i^2 \overline{\boldsymbol{x}}_i^T \left(\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}\right)^{-1} \overline{\boldsymbol{x}}_i \\ &+ n_i \gamma_i^3 \frac{\tau_2}{\tau_1 - 2} I^{\lambda\lambda} + O(m^{-3/2}). \end{aligned} \tag{4.6}$$

4.2 Second-order unbiased estimators of the conditional and unconditional MSEs

We now derive second-order unbiased estimators of the unconditional and conditional MSEs. Since it is hard to derive second-order biases of the MLEs of $\boldsymbol{\beta}$, λ, τ_1 and τ_2, we could not provide analytical second-order unbiased estimators of the MSEs. Instead, we use the parametric bootstrap methods, which provide second-order unbiased MSE estimators.

We begin by treating the unconditional case. The parametric bootstrap sample in this case is denoted as

$$y_{ij}^* = \boldsymbol{x}_{ij}^T \widehat{\boldsymbol{\beta}} + v_i^* + \varepsilon_{ij}^*, \quad i = 1, \ldots, m; \ j = 1, \ldots, n_i, \tag{4.7}$$

where v_i^*'s and ε_{ij}^*'s are conditionally mutually independent given η_i^*'s and

$$\begin{aligned} v_i^* | \eta_i^* \sim & \mathcal{N}(0, \hat{\lambda}/\eta_i^*), \\ \varepsilon_{ij}^* | \eta_i^* \sim & \mathcal{N}(0, 1/\eta_i^*), \\ \eta_i^* \sim & \sim \mathcal{G}a(\hat{\tau}_1/2, 2/\hat{\tau}_2). \end{aligned} \tag{4.8}$$

The estimator of the unconditional MSE, $MSE(\lambda, \boldsymbol{\tau}; \hat{\xi}_i^{EB})$, is given by

$$mse^*(\hat{\xi}_i^{EB}) = \hat{g}_1^* + \hat{g}_2^*,$$

where

$$\begin{aligned} \hat{g}_1^* =& 2g_1(\hat{\lambda}, \widehat{\boldsymbol{\tau}}) - E_*[g_1(\hat{\lambda}^*, \widehat{\boldsymbol{\tau}}^*)], \\ \hat{g}_2^* =& \hat{\gamma}_i^2 E^*[\{\overline{\boldsymbol{x}}_i^T (\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})\}^2] + n_i \hat{\gamma}_i^3 \frac{\hat{\tau}_2}{\hat{\tau}_1 - 2} E^*[(\hat{\lambda}^* - \hat{\lambda})^2]. \end{aligned}$$

Proposition 4.1 *Assume the condition* (A1)*. Then,*

$$E[mse^*(\hat{\xi}_i^{EB})] = MSE(\lambda, \boldsymbol{\tau}; \hat{\xi}_i^{EB}) + O(m^{-3/2}).$$

We next consider the conditional case. Keeping $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^T$ fixed, a bootstrap sample $\boldsymbol{y}_k = (y_{k1}, \ldots, y_{kn_k})^T$ is generated from (4.7) for $k \neq i$. Noting that $\boldsymbol{y}_i$ is fixed, we construct the estimators $\widehat{\boldsymbol{\beta}}_{(i)}^*$, $\hat{\lambda}_{(i)}^*$, $\hat{\tau}_{1(i)}^*$ and $\hat{\tau}_{2(i)}^*$ from $\boldsymbol{y}_i$ and the bootstrap sample

$$\boldsymbol{y}_1^*, \ldots, \boldsymbol{y}_{i-1}^*, \boldsymbol{y}_i, \boldsymbol{y}_{i+1}^*, \ldots, \boldsymbol{y}_m^* \tag{4.9}$$

with the same technique as used to obtain the estimator $\widehat{\boldsymbol{\beta}}$, $\hat{\lambda}$, $\hat{\tau}_1$ and $\hat{\tau}_2$. Let $E_*[\cdot|\boldsymbol{y}_i]$ be the expectation with regard to the bootstrap sample (4.9). The conditional MSE is given by $cMSE(\omega; \hat{\xi}_i^{EB}|\boldsymbol{y}_i) = g_1^c(\omega|\boldsymbol{y}_i) + g_2^c(\omega|\boldsymbol{y}_i)$, where $g_1^c(\omega|\boldsymbol{y}_i) = E[\{\xi_i - \hat{\xi}_i^B(\boldsymbol{\beta}, \lambda)\}^2|\boldsymbol{y}_i]$ and $g_2^c(\omega|\boldsymbol{y}_i) = E[\{\hat{\xi}_i^B(\widehat{\boldsymbol{\beta}}, \hat{\lambda}) - \hat{\xi}_i^B(\boldsymbol{\beta}, \lambda)\}^2|\boldsymbol{y}_i]$ from (4.1). Since $g_1^c(\omega|\boldsymbol{y}_i) = n_i^{-1}(1 - \gamma_i(\lambda))(Q_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda) + \tau_2)/(n_i + \tau_1 - 2)$ from (4.3), a second-order unbiased estimator of $g_1^c(\omega|\boldsymbol{y}_i)$ is given by

$$\hat{g}_1^{c*} = 2g_1(\boldsymbol{y}_i, \widehat{\boldsymbol{\beta}}, \hat{\lambda}, \widehat{\boldsymbol{\tau}}) - E_* \left[g_1(\boldsymbol{y}_i, \widehat{\boldsymbol{\beta}}_{(i)}^*, \hat{\lambda}_{(i)}^*, \widehat{\boldsymbol{\tau}}_{(i)}^*)|\boldsymbol{y}_i \right].$$

Then, it can be verified that $E[\hat{g}_1^{c*}|\boldsymbol{y}_i] = g_1^c(\omega|\boldsymbol{y}_i) + o_p(m^{-1})$. $g_2^c(\omega|\boldsymbol{y}_i)$, is estimated via parametric bootstrap as

$$\hat{g}_2^{c*} = E^* \left[\{\hat{\xi}_i^{B*}(\widehat{\boldsymbol{\beta}}_{(i)}^*, \hat{\lambda}_{(i)}^*) - \hat{\xi}_i^{B*}(\widehat{\boldsymbol{\beta}}, \hat{\lambda})\}^2 |\boldsymbol{y}_i \right].$$

Thus,

$$cmse^*(\hat{\xi}_i^{EB}|\boldsymbol{y}_i) = \hat{g}_1^{c*} + \hat{g}_2^{c*}. \tag{4.10}$$

Theorem 4.2 *Under the condition* (A1)*, the estimator* (4.10) *is a second-order unbiased estimator of cMSE, namely*

$$E[cmse^*(\hat{\xi}_i^{EB}|\boldsymbol{y}_i)|\boldsymbol{y}_i] = cMSE(\omega; \hat{\xi}_i^{EB}|\boldsymbol{y}_i) + o_p(m^{-1}).$$

5 Application to PLP data in Japan

We now investigate empirical performances of the suggested model, the empirical Bayes estimator and the second-order unbiased estimators of the conditional and unconditional MSEs through analysis of real data. The data used here originates from the posted land price data along the Keikyu train line in 2001. This train line connects the suburbs in the Kanagawa prefecture to the Tokyo metropolitan area. Those who live in the suburbs in the Kanagawa prefecture take this line to work or study in Tokyo everyday. Thus, it is expected that the land price depends on the distance from Tokyo. The posted land price data are available for 52 stations on the Keikyu train line, and we consider each station as a small area, namely, $m = 52$. For the i-th station, data of n_i land spots are available, where n_i varies around 4 and some areas have only one observation.

To investigate variability in each area, the boxplots are drawn for all areas. For nine selected areas among areas with more than 4 observations, we draw the boxplots in Figure 1, which clearly indicates that the posted land price has the large within-area variation and the conventional NER model (which assumes homogeneity of variance) does not seem to be appropriate.

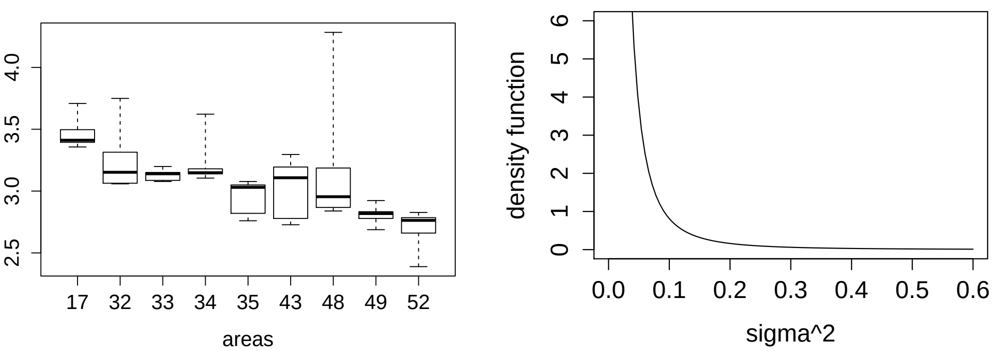


Figure 1: Boxplots of the Posted Land Price Data for Selected Areas (left) and the Estimated Density Function of $\sigma_i^2 = 1/\eta_i$ (right).

For $j = 1, \ldots, n_i$, y_{ij} denotes the value which is transformed by logarithm from the posted land price (Yen/10,000) for the unit meter squares of the j-th spot, T_i is the time to take from the nearby station i to the Tokyo station around 8:30 in the morning, D_{ij} is the value of geographical distance from the spot j to the station i and FAR_{ij} denotes the floor-area ratio, or ratio of building volume to lot area of the spot j. These values of T_i, D_{ij} and FAR_{ij} are transformed by logarithm. Since these data have within-area variability as indicated in Figure 1 (left), we use the RHNER model

$$y_{ij} = \beta_0 + FAR_{ij}\beta_1 + T_i\beta_2 + D_{ij}\beta_3 + v_i + \varepsilon_{ij}, \tag{5.1}$$

where v_i and ε_{ij} are mutually independent and distributed as $\mathcal{N}(0, \lambda\sigma_i^2)$ and $\mathcal{N}(0, \sigma_i^2)$, and $\eta_i (= 1/\sigma_i^2)$ is independently distributed as $\Gamma(\tau_1/2, 2/\tau_2)$.

The estimates of the parameters $(\beta_0, \beta_1, \beta_2, \beta_3, \lambda, \tau_1, \tau_2)$ are

$$\widehat{\beta}_0 = 5.69,\ \widehat{\beta}_1 = 0.11,\ \widehat{\beta}_2 = -0.63,\ \widehat{\beta}_3 = -0.08,\ \hat{\lambda} = 0.22,\ \hat{\tau}_1 = 2.93,\ \hat{\tau}_2 = 0.04.$$

It is interesting to point out that the estimated regression function is a decreasing function of T_i and D_{ij}, which means that the land price y_{ij} tends to decrease as the time from Tokyo or distance from nearest station increases. Since $\hat{\tau}_1 = 2.93$ and $\hat{\tau}_2 = 0.04$, the distribution of η_i has a large mean about 73 and a heavy tail. Since the estimated value of τ_1 is smaller than 4, the variance of η_i or σ_i^2 does not exist, which agrees the observation that the posted land price data has great variability as indicated by the boxplots in Figure 1. Figure 1 (right) draws the estimated density function of $\sigma_i^2 = 1/\eta_i$ where η_i has $\Gamma(\hat{\tau}_1/2, 2/\hat{\tau}_2)$, so that the distribution of σ_i^2 has a small mean, but a heavy tail.

The predicted values of $\overline{\boldsymbol{x}}_i'\boldsymbol{\beta} + v_i$ and their conditional and unconditional MSE estimates, which can be obtained based on $1,000$ bootstrap samples, are given in Table 4. It is revealed from Table 4 that the estimates of the unconditional MSE get smaller as n_i gets larger. On the other hands, the estimates of the conditional MSE do not have a similar property, because the conditional MSE is affected by not only n_i but also the observed values as indicated in Table 3. It is interesting to point out that, in area 48, the estimated conditional MSE is relatively large while the estimated unconditional MSE is not large. Noting that this area has great variability as shown in Figure 1, it seems that the conditional MSE can capture the variability of areas.

Table 1: Values of EBLUP and Estimates of Unconditional and Conditional MSES for Selected fifteen Areas. (Estimates of MSE and cMSE are multiplied by 100)

Area	n_i	EBLUP	$\widehat{\text{MSE}}$	$\widehat{\text{cMSE}}$
1	1	4.02	5.19	0.58
4	2	3.91	4.34	0.49
5	5	3.96	3.13	2.31
8	3	3.86	3.83	0.33
17	7	3.50	2.66	1.04
25	7	3.39	2.65	1.37
26	4	3.45	3.42	1.88
32	6	3.22	2.86	2.68
33	8	3.12	2.48	1.90
34	11	3.16	2.09	1.10
35	7	2.99	2.65	3.58
43	6	3.02	2.86	3.73
48	6	3.07	2.86	5.11
49	10	2.82	2.21	2.69
52	6	2.76	2.87	6.55

References

[1] 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.

[2] Booth, J. S. and Hobert, P. (1998). Standard errors of prediction in generalized linear mixed models. *J. Amer. Statist. Assoc.*, **93**, 262 - 272.

[3] Butar, F.B. and Lahiri, P. (2003). On measures of uncertainty of empirical Bayes small-area estimators. *J. Statist. Plan. Inf.*, **112**, 63-76.

[4] Datta, G.S. (2009). Model-based approach to small area estimation. In *Handbook of Statistics*, **29B**. Editors: D. Pfeffermann and C.R. Rao. North Holland, New York, pp 251-288.

[5] 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.

[6] Fay, R.E. and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*, **74**, 269-277.

[7] Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statist. Science*, **9**, 55-93.

[8] Jiang, J. and Nguyen, T. (2012). Small area estimation via heteroscedastic nested-error regression. *Canad. J. Statist.*, **40**, 588-603.

[9] Mardia, K.V. and Marshall, R.J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, **71**, 135-146.

[10] Pfeffermann, D. (2002). Small area estimation: new developments and directions. *International Statistical Institute Review*, **70**, 125-143.

[11] 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.

[12] Rao, J.N.K. (2003). *Small Area Estimation.* Wiley.

[13] Sweeting, T.J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.*, **8**, 1375-1381.

乗法モデルとベンチマーク問題

川久保 友超 [1]
with Malay Ghosh[2], 久保川 達也 [3]

[1] 東京大学・経済・D2

[2]University of Florida

[3] 東京大学

2015 年 1 月 30 日
科研コンファレンス「経済統計・政府統計の数理的基礎と応用」
東京大学

はじめに

小地域推定

- ある予算を市町村（小地域）ごとに配分する際に，政策の根拠として各小地域の所得の平均を知りたいとする.
- しかし予算制約などの都合で小地域レベルでは十分なサンプルを得られない場合，標本平均では推定誤差が大きく，不適切な予算配分が行われる危険性がある.

→小地域ごとの推定精度を上げたい

はじめに

Fay–Herriot model（Fay and Herriot, 1979）

- 小地域 i の興味ある特性の direct estimator（標本平均など）を z_i とし，z_i を以下のような 2 段階のモデリングを行う．

$$\begin{aligned} Level\ 1: &\ (sampling\ \ model): \ z_i = \phi_i + \varepsilon_i; \\ Level\ 2: &\ (linking\ \ model): \ \phi_i = \boldsymbol{x}_i^t \boldsymbol{\beta} + v_i, \end{aligned}$$

  ε_i は sampling error，v_i は地域効果（random effect）で，
  $\varepsilon_i \sim \mathcal{N}(0, d_i)$, $v_i \sim \mathcal{N}(0, \tau^2)$，ただし d_i は既知．

- モデルにもとづいて ϕ_i を推定する．
  → model based estimator:

$$\hat{\phi}_i = \frac{\hat{\tau}^2}{d_i + \hat{\tau}^2} z_i + \frac{d_i}{d_i + \hat{\tau}^2} \boldsymbol{x}_i^t \widehat{\boldsymbol{\beta}}(\hat{\tau}^2)$$

- direct estimator z_i を $\boldsymbol{x}_i^t \boldsymbol{\beta}$ の方向に縮小する作用がはたらき，小地域の推定が安定する．

はじめに

ベンチマーク法

- model based estimator の総和を，direct estimator の総和（ベンチマーク）に一致させる． → benchmarked estimator
  - model misspecification に対して頑健．
  - 官庁統計における要請（統計の整合性の観点から）．
- 制約付きベイズ推定量（Ghosh, 1992）が小地域の benchmarked estimator としてよく用いられる．

はじめに

本研究の概要

- 所得など正の値のみをとり歪んでいるデータに対し，対数変換したものが Fay–Herriot モデルとなるような乗法モデルとして，transformed Fay–Herriot model（Slud and Maiti, 2006）を考える．
- weighted Kullback–Leibler loss のもとで，benchmarked estimator を導出する．benchmarked estimator は non-benchmarked model based estimator の定数倍という自然なかたちとなる．
- 推定量のリスクの漸近不偏推定量を構成する．

モデルとベンチマーク制約

transformed Fay–Herriot model

- 各地域 $i = 1, \ldots, m$ に対して，正の direct estimarot y_i が乗法モデル $y_i = \theta_i \eta_i$ にしたがっている．推定の対象は θ_i.
- $z_i = \log y_i$，$\phi_i = \log \theta_i$，$\varepsilon_i = \log \eta_i$ とすると，z_i が以下の Fay–Herriot model にしたがう．

$$
\begin{aligned}
z_i =& \phi_i + \varepsilon_i, \quad \phi_i = \boldsymbol{x}_i^t \boldsymbol{\beta} + v_i, \\
v_i \sim& \mathcal{N}(0, \tau^2), \quad \varepsilon_i \sim \mathcal{N}(0, d_i)
\end{aligned}
$$

  d_i は既知の定数，τ^2 は未知パラメータ，$v_1, \ldots, v_m, \varepsilon_1, \ldots, \varepsilon_m$ は互いに独立．
- $\boldsymbol{\beta}$ は事前分布 $\boldsymbol{\beta} \sim uniform(\mathbb{R}^p)$ をもつとする（階層ベイズモデル）．
- θ_i をベイズ推定し，階層ベイズ（HB）推定量 $\hat{\theta}_i^{\mathrm{HB}}$ を得る．
  → （制約無し）model based estimator
- ベンチマーク制約のもとで事後リスク最小化を行い，制約付き階層ベイズ推定量 $\hat{\theta}_i^{\mathrm{CHB}}$ を得る．
  → benchmarked estimator

モデルとベンチマーク制約

ベンチマーク制約

$$\sum_{i=1}^{m} w_i \hat{\theta}_i = \sum_{i=1}^{m} w_i y_i$$

- 上式の制約下での事後リスク最小化問題の解が，制約付き階層ベイズ（CHB）推定量.
- $L(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ を損失関数とすると，ラグランジュ関数は以下で定義される.

$$LM(\widehat{\boldsymbol{\theta}}, \lambda) = E[L(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})|\mathbf{y}] + \lambda \Big\{ \sum_{i=1}^{m} w_i \hat{\theta}_i - \sum_{i=1}^{m} w_i y_i \Big\},$$

- 推定量は，損失関数の取り方に依存する.
- θ_i の推定に対して適切で，かつ CHB 推定量が陽に書け自然なかたちになる損失関数を提案したい.

モデルとベンチマーク制約

損失関数の選択

- 損失関数の候補をいくつか考える.

$$L_{\mathrm{Q}}(\widehat{\boldsymbol{\theta}},\boldsymbol{\theta}) = \sum_{i=1}^{m} \xi_i(\hat{\theta}_i - \theta_i)^2,$$

$$L_{\mathrm{TQ}}(\widehat{\boldsymbol{\theta}},\boldsymbol{\theta}) = \sum_{i=1}^{m} \xi_i(\log\hat{\theta}_i - \log\theta_i)^2,$$

$$L_{\mathrm{KL}}^{\xi}(\widehat{\boldsymbol{\theta}},\boldsymbol{\theta}) = \sum_{i=1}^{m} \xi_i\left\{\hat{\theta}_i/\theta_i - \log(\hat{\theta}_i/\theta_i) - 1\right\},$$

- L_{Q}：制約無しベイズ推定量は，事後平均 $E(\theta_i|\boldsymbol{y})$．しかし，CB 推定量が負値をとりうる．また尺度母数 θ_i の推定に二乗損失は適当でない．
- L_{TQ}：CB 推定量が陽に書けない．
- L_{KL}^{ξ}：制約無しベイズ推定量は，調和平均 $1/E(\theta_i^{-1}|\boldsymbol{y})$．

モデルとベンチマーク制約

weighted Kullback–Leibler loss

$$L_{\mathrm{KL}}^{w}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \sum_{i=1}^{m} w_i \theta_i \left\{ \hat{\theta}_i / \theta_i - \log(\hat{\theta}_i / \theta_i) - 1 \right\}$$

- weighted KL loss のもとでの θ_i の（制約無し）階層ベイズ推定量は，$\hat{\theta}_i^{\mathrm{HB}}(\tau^2) = E(\theta_i | \boldsymbol{y})$（事後平均）となる.
- 以下の $\hat{\theta}_i^{\mathrm{CHB}}$ は，weighted KL loss とベンチマーク制約のもとでの，制約付き階層ベイズ推定量.

$$\hat{\theta}_i^{\mathrm{CHB}}(\tau^2) = \hat{\theta}_i^{\mathrm{HB}}(\tau^2) \frac{\sum_{j=1}^{m} w_j y_j}{\sum_{j=1}^{m} w_j \hat{\theta}_j^{\mathrm{HB}}(\tau^2)}.$$

- τ^2 を推定量 $\hat{\tau}^2$ でおきかえて，
  - $\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2) \to$（制約無し）model based estimator
  - $\hat{\theta}_i^{\mathrm{CHB}}(\hat{\tau}^2) \to$ benchmarked estimator

リスクの推定

制約無し model based estimator の推定リスク

- weighted KL loss にもとづいた，$\hat{\theta}_i^{\mathrm{HB}}$ の推定リスクの 2 次漸近不偏推定量（$O(m^{-1})$ の項まで）を構成する．
- $\hat{\theta}_i^{\mathrm{HB}}(\tau^2) = E(\theta_i|\boldsymbol{y}, \tau^2)$ に気を付けると，リスクは以下に分解される．

$$
\begin{aligned}
&R_\omega\{\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2)\} \\
=&E[\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2) - \theta_i - \theta_i \log\{\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2)/\theta_i\}] \\
=&E[-\theta_i \log\{\hat{\theta}_i^B/\theta_i\}] + E[\hat{\theta}_i^{\mathrm{HB}}(\tau^2) - \theta_i - \theta_i \log\{\hat{\theta}_i^{\mathrm{HB}}(\tau^2)/\hat{\theta}_i^B\}] \\
&+ E[\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2) - \hat{\theta}_i^{\mathrm{HB}}(\tau^2) - \theta_i \log\{\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2)/\hat{\theta}_i^{\mathrm{HB}}(\tau^2)\}] \\
=&I_1 + I_2 + I_3 \ (\mathrm{say}).
\end{aligned}
$$

ただし，$\hat{\theta}_i^{\mathrm{B}} = \hat{\theta}_i^{\mathrm{B}}(\boldsymbol{\omega}) = E(\theta_i|\boldsymbol{y}, \boldsymbol{\omega})$，$\boldsymbol{\omega} = (\boldsymbol{\beta}^t, \tau^2)^t$.

- I_1，I_2 は解析的な評価が比較的容易だが，I_3 は超パラメータ τ^2 の推定リスクを測っており評価が難しい．

→解析的な手法（テイラー展開）と数値的な手法（parametric bootstrap）

リスクの推定

制約無し model based estimator の推定リスク（解析的手法）

- $I_1 = g_{1i}(\omega)$, $I_2 = g_{2i}(\omega) + O(m^{-2})$, $I_3 = g_{3i}(\omega) + O(m^{-3/2})$ となり，

$$R_\omega\{\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2)\} = g_{1i}(\omega) + g_{2i}(\omega) + g_{3i}(\omega) + O(m^{-3/2}),$$

  とリスクが近似できる．ただし，
  $g_{1i}(\omega) = O(1)$, $g_{2i}(\omega) = O(m^{-1})$, $g_{3i}(\omega) = O(m^{-1})$.
- $g_{1i}(\widehat{\omega})$ は 2 次バイアス $b_{1i}(\omega) = O(m^{-1})$ が生じる．

$$E\{g_{1i}(\widehat{\omega})\} = g_{1i}(\omega) + b_{1i}(\omega) + O(m^{-3/2})$$

- これらの表現がすべて解析的に得られる場合，

$$\widehat{R}_\omega\{\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2)\} = g_{1i}(\widehat{\omega}) - b_{1i}(\widehat{\omega}) + g_{2i}(\widehat{\omega}) + g_{3i}(\widehat{\omega})$$

  としてリスクの 2 次漸近不偏推定量を構成できる．

リスクの推定

制約無し model based estimator の推定リスク（ブートストラップ）

- 推定量で構成した以下のモデルからブートストラップ標本を得る.
- $i = 1, \ldots, m$ に対して，正の観測 y_i^* が乗法モデル $y_i^* = \theta_i^* \eta_i^*$ にしたがっている.
- $z_i^* = \log y_i^*, \phi_i^* = \log \theta_i^*, \varepsilon_i^* = \log \eta_i^*$ とすると，z_i^* が以下の対数線形モデルにしたがう.

$$z_i^* = \phi_i^* + \varepsilon_i^*, \quad \phi_i^* = \boldsymbol{x}_i^t \widehat{\boldsymbol{\beta}}(\hat{\tau}^2) + u_i^*,$$
$$u_i^* \sim \mathcal{N}(0, \hat{\tau}^2), \quad \varepsilon_i^* \sim \mathcal{N}(0, d_i)$$

ただし，$\widehat{\boldsymbol{\beta}}(\hat{\tau}^2)$ と $\hat{\tau}^2 = \hat{\tau}^2(\boldsymbol{y})$ は，データ $\boldsymbol{y}$ にもとづいた推定量.

- $l_1^* = \{g_{1i}(\widehat{\boldsymbol{\omega}})\}^2 / E_*\{g_{1i}(\widehat{\boldsymbol{\omega}}^*)\}$，$l_2^* = g_{2i}(\widehat{\boldsymbol{\omega}})$，さらに

$$l_3^* = E_*[\hat{\theta}_i^{\mathrm{HB}*}(\hat{\tau}^{2*}) - \hat{\theta}_i^{\mathrm{HB}*}(\hat{\tau}^2) - \hat{\theta}_i^{\mathrm{B}*}(\widehat{\boldsymbol{\omega}}) \log\{\hat{\theta}_i^{\mathrm{HB}*}(\hat{\tau}^{2*}) / \hat{\theta}_i^{\mathrm{HB}*}(\hat{\tau}^2)\}]$$

としたとき，$\widehat{R}^*\{\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2)\} = l_1^* + l_2^* + l_3^*$ がリスクの 2 次漸近不偏推定量.

リスクの推定

benchmarked estimator の推定リスク

- benchmarked estimator $\hat{\theta}_i^{\mathrm{CHB}}(\hat{\tau}^2)$ のリスクを推定する.

$$\hat{\theta}_i^{\mathrm{CHB}}(\hat{\tau}^2) = \hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2) \frac{\sum_{j=1}^m w_j y_j}{\sum_{j=1}^m w_j \hat{\theta}_j^{\mathrm{HB}}(\hat{\tau}^2)}.$$

- $\hat{\theta}_i^{\mathrm{CHB}}(\hat{\tau}^2)$ のリスクは $\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2)$ のリスクを用いて，以下のように表現できる.

$$\begin{aligned} R_\omega\{\hat{\theta}_i^{\mathrm{CHB}}(\hat{\tau}^2)\} =& R_\omega\{\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2)\} - E\left[\hat{\theta}_i^{\mathrm{B}}(\omega) \log\left(\frac{\sum_{j=1}^m w_j y_j}{\sum_{j=1}^m w_j \hat{\theta}_j^{\mathrm{HB}}(\hat{\tau}^2)}\right)\right] \\ & + E\left[\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2)\left\{\frac{\sum_{j=1}^m w_j y_j}{\sum_{j=1}^m w_j \hat{\theta}_j^{\mathrm{HB}}(\hat{\tau}^2)} - 1\right\}\right] \\ =& R_\omega\{\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2)\} - K_1 + K_2. \end{aligned}$$

- $\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2)$ と $\hat{\theta}_i^{\mathrm{CHB}}(\hat{\tau}^2)$ のリスクの差 $-K_1 + K_2$ を推定すればよい.

リスクの推定

benchmarked estimator の推定リスク（つづき）

- $-K_1 + K_2 = E[\widehat{K}] + K_3$ と書き換えられる．ただし，

$$\widehat{K} = \hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2)\left\{\frac{\sum_{j=1}^m w_j y_j}{\sum_{j=1}^m w_j \hat{\theta}_j^{\mathrm{HB}}(\hat{\tau}^2)} - \log\left(\frac{\sum_{i=1}^m w_i y_i}{\sum_{j=1}^m w_j \hat{\theta}_j^{\mathrm{HB}}(\hat{\tau}^2)}\right) - 1\right\},$$

$$K_3 = E\left[\left\{\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2) - \hat{\theta}_i^{\mathrm{B}}(\boldsymbol{\omega})\right\}\log\left\{\frac{\sum_{i=1}^m w_i y_i}{\sum_{j=1}^m w_j \hat{\theta}_j^{\mathrm{HB}}(\hat{\tau}^2)}\right\}\right].$$

- $\widehat{K}$ は $E[\widehat{K}]$ の不偏推定量だから，$K_3 = O(m^{-1})$ をパラメトリック・ブートストラップ法で推定する→ K_3^* を得る．
- $\widehat{R}^*\{\hat{\theta}_i^{\mathrm{CHB}}(\hat{\tau}^2)\} = \widehat{R}^*\{\hat{\theta}_i^{\mathrm{HB}}(\hat{\tau}^2)\} + \widehat{K} + K_3^*$ が $\hat{\theta}_i^{\mathrm{CHB}}(\hat{\tau}^2)$ のリスクの 2 次漸近不偏推定量．

実データにもとづく例

小地域の教育支出の平均推定

- 2011 年 11 月の各県庁所在地の教育支出の平均を推定する.
- 家計調査「教育支出」を direct estimator とする.
- 家計調査は毎月行われるが，小地域ではサンプル数が小さく推定が安定しない.
- そこで transformed Fay–Herriot model をあてはめ，（制約無しの）model based estimate, benchmarked estimate をそれぞれ求める．さらにリスクの推定も行う.
- 補助情報（モデルの説明変数）として，5 年に 1 度行われる全国消費実態調査の 2009 年の「教育支出」のデータを用いる．全国消費実態調査は大規模な調査のため，推定の精度が高い.

実データにもとづく例

- 各県庁所在地をあらわす添え字 $i = 1, \ldots, m, (m = 47)$ に対して，i 番目の都市の 2011 年 11 月の家計調査「教育支出」データ（単位は千円）を y_i とする（direct estimator），
- 説明変数として，i 番目の都市の 2009 年全国消費実態調査「教育支出」データを x_{i1} とする．
- $z_i = \log y_i$ に対し，以下のモデルをあてはめる．

$$
\begin{aligned}
z_i =& \phi_i + \varepsilon_i, \quad \phi_i = \boldsymbol{x}_i^t \boldsymbol{\beta} + u_i, \\
u_i \sim& \mathcal{N}(0, \tau^2), \quad \varepsilon_i \sim \mathcal{N}(0, d_i)
\end{aligned}
$$

  ただし $\boldsymbol{x}_i^t = (1, x_{i1})$ とする．
- モデル上は誤差項の分散 d_i は既知だが，同じ都市の過去 10 年の 11 月の「教育支出」の標本分散で推定する．
- i 番目の都市の標本数を n_i とし，ウェイトは標本数で重みをつけた $w_i = n_i / \sum_{j=1}^{47} n_j$ を用いる．

実データにもとづく例

推定結果

- $\sum_{i=1}^{m} w_i y_i / \sum_{j=1}^{m} w_j \hat{\theta}_j^{\mathrm{HB}}(\hat{\tau}^2)$ の値は 1.0876 となった．よって，HB 推定値を 1.0876 倍すれば，CHB 推定値が得られる．
- direct estimator y_i のリスクも推定した．
- CHB 推定量は制約付きの事後リスク最小解なので，制約無しの事後リスク最小解である HB よりも，リスクは大きくなる．しかしほとんどの地域で y_i のリスクよりは小さい．

県	n_i	d_i	y_i	HB	CHB	$\widehat{R}_{\mathrm{HB}}$	$\widehat{R}^*_{\mathrm{HB}}$	$\widehat{R}^*_{\mathrm{CHB}}$	$\widehat{R}_{\boldsymbol{y}}$
茨城	95	9.56	8.10	8.89	9.65	19.5	19.1	22.4	45.6
栃木	95	26.96	10.03	9.48	10.29	23.6	23.4	27.3	134.7
群馬	94	8.41	5.21	7.71	8.37	19.7	19.4	22.3	42.0
埼玉	95	3.93	12.33	12.73	13.83	19.9	19.7	25.0	26.0
千葉	94	37.83	30.71	13.10	14.22	32.1	31.5	37.0	240.3
東京	386	2.30	15.45	14.45	15.68	12.4	12.7	18.3	14.2
神奈川	142	15.99	23.25	14.06	15.27	30.0	29.3	35.0	98.4

Table : HB 推定値，CHB 推定値とそれらのリスクの推定値

- Fay, R.E. and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*, **74**, 269-277.
- Ghosh, M. (1992). Constrained Bayes estimation with applications. *J. American Statist. Assoc.*, **87**, 533-540.
- Slud, E.V. and Maiti, T. (2006). Mean-squared error estimation in transformed Fay-Herriot models. *J. Royal Statist. Soc.*, **B 68**, 239-257.

データ変換と小地域推定

菅澤 翔之助

東京大学大学院経済学研究科修士 2 年

2015 年 1 月 30 日

経済統計・政府統計の基礎と応用

小地域推定 (設定・目的・方法)

設定

m 個のクラスター (地域) があり，各地域内で n_i 個 ($i = 1, \ldots, m$) のサンプル $(y_{ij}, \mathbf{x}'_{ij})$, $j = 1, \ldots, n_i$ が得られている.

クラスター数 m は大きいが n_i はあまり大きくない.

目的

各地域の平均を精度良く推定したい.

($\bar{y}_i$ で推定した場合は n_i が小さい故に不安定な推定量になってしまう．)

手法

他の地域の情報や共変量の情報を「うまく」用いる.

具体的には Nested Error Regression Model(NERM) と呼ばれる以下のモデルを用いる.

$$y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + v_i + \varepsilon_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, m$$

$v_i \sim N(0, \tau^2)$: 変量効果, $\varepsilon_{ij} \sim N(0, \sigma^2)$: 誤差項, v_i, ε_{ij} は互いに独立.

個々の地域の特徴を v_i で表現し全体としての特徴を $\boldsymbol{\beta}$ で表現している.

小地域推定 (設定・目的・方法)

前スライドのモデルをもとに

$$\mu_i = \bar{\mathbf{x}}_i' \boldsymbol{\beta} + v_i$$

の予測量を構成すると

$$\hat{\mu}_i = \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}} + \frac{n_i \hat{\tau}^2}{\hat{\sigma}^2 + n_i \hat{\tau}^2} (\bar{y}_i - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}})$$

となる.

これを各地域の平均の推定量として用いることで安定した推定を行うことができる.

線形混合モデル

実は NERM は線形混合モデル (LMM) と呼ばれるモデルのクラスに属する．

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon},$$
$$\mathbf{b} \sim N_r(\mathbf{0}, \mathbf{G}), \quad \boldsymbol{\varepsilon} \sim N_N(\mathbf{0}, \mathbf{R})$$

このモデルにおいて $\mu = \mathbf{c}'\boldsymbol{\beta} + \mathbf{d}'\mathbf{b}$ の予測量は

$$\hat{\mu} = \mathbf{c}'\hat{\boldsymbol{\beta}} + \mathbf{d}'\hat{\mathbf{G}}\mathbf{Z}'\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

で与えられる．(EBLUP: 経験最良不偏予測量)

問題提起

小地域推定に限らず地域毎 (クラスター毎) に観測されたデータにおいて y_{ij} が正の値となるデータは頻出.

LMM は観測 $\mathbf{y}$ に正規分布を仮定しているため，正のデータに対してそのまま適用するのは妥当でない.

そのような状況では対数変換を用いて $\log y_{ij}$ に対して LMM を当てはめるが一般的.

本研究の目的

Box-Cox(BC) 変換を用いた LMM を導入し，そのモデルに対する推定および予測を考え，小地域推定へ応用する．

Box-Cox 変換

$$h(x,\lambda) = (x^{\lambda} - 1)/\lambda, \quad \lambda \neq 0, \qquad h(x,0) = \log x$$

$\Rightarrow$ 対数変換 ($\lambda = 0$) と恒等変換 ($\lambda = 1$) を含む.

変換パラメータもデータから推定することによって柔軟なモデリングが可能になる.

本研究の目的

具体的な研究の目的は以下の通り.

- BC 変換を取り入れた LMM を提案する.

  特に小地域推定で NERM を用いた応用を視野にいれている.

- すべてのパラメータの一致推定量を構成する.

  実は BC 変換で最尤法により λ を推定すると一致性を持たないのでその点を解消する.

- 予測量を構成し，その不確実性を測るために予測区間を構成する.

  予測量のリスクを測ることは応用上非常に重要.

BC-LMM

次のような Box-Cox transformed LMM (BC-LMM) を考える.

$$h(\mathbf{y}_i, \lambda) = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, m$$

$\mathbf{b}_1, \ldots, \mathbf{b}_N \in \mathbb{R}^q$ および $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_N \in \mathbb{R}^{n_i}$ は互いに独立に
$\mathbf{b}_i \sim N_q(0, \mathbf{R}(\boldsymbol{\psi}))$, $\boldsymbol{\varepsilon}_i \sim N_{n_i}(0, \mathbf{G}_i(\boldsymbol{\psi}))$.

$\boldsymbol{\psi}$ は分散パラメータ.

$q = 1$ として $\mathbf{Z}_i = \mathbf{1}_{n_i}$ とすると以下のような Box-Cox 変換を取り入れた NERM(BC-NERM) が得られる.

$$h(y_{ij}, \lambda) = \mathbf{x}_{ij}' \boldsymbol{\beta} + v_i + \varepsilon_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, m$$

パラメータ推定

λ が所与のもとでは通常の LMM の理論が使える.

$\Rightarrow \boldsymbol{\beta}, \psi$ ともに λ の関数として $\hat{\boldsymbol{\beta}}(\lambda), \hat{\psi}(\lambda)$ として表せる.

変換パラメータ λ の推定が本質的.

以下の方程式の解として推定.

$$\sum_{i=1}^{m} \sum_{j=1}^{n_i} \left\{ h(y_{ij}, \lambda) - \mathbf{x}'_{ij} \hat{\boldsymbol{\beta}}(\lambda) \right\}^3 = 0,$$

「なせ変換するか？—分布を正規分布に近付けるため」

ということを考えると自然な推定手法.

パラメータ推定

λ が推定できれば $\boldsymbol{\beta}, \boldsymbol{\psi}$ は plug-in して $\hat{\boldsymbol{\beta}}(\hat{\lambda}), \hat{\boldsymbol{\psi}}(\hat{\lambda})$ と推定できる.

以上のように得られた推定量に対して以下が示せる.

$\theta = (\boldsymbol{\beta}', \boldsymbol{\psi}', \lambda)'$ とすると

$$\hat{\theta} - \theta = O_p(m^{-1/2}), \quad E[\hat{\theta} - \theta] = O(m^{-1})$$

λ の推定に対して ML を用いるとこの結果は得られない.

予測

以下の量の予測を考える.

$$h^{-1}(\mu,\lambda)=(\lambda\mu+1)^{1/\lambda},\qquad \mu=\mathbf{c}_1'\boldsymbol{\beta}+\mathbf{c}_2'\mathbf{b},$$

ただし $\mathbf{b}=(\mathbf{b}_1',\ldots,\mathbf{b}_m')'$ で $\mathbf{c}_1,\mathbf{c}_2$ は既知の定数ベクトル.

(固定効果 β と変量効果 b の線形結合 μ を逆変換した量)

μ の EBLUP は

$$\hat{\mu}^{\mathrm{EBLUP}}=\mathbf{c}_1'\hat{\boldsymbol{\beta}}+\mathbf{c}_2'\hat{\mathbf{b}}^{\mathrm{BP}}(\hat{\psi},\hat{\lambda}),$$

ただし

$$\hat{\mathbf{b}}_i^{\mathrm{BP}}(\psi,\lambda)=\mathbf{R}(\psi)\mathbf{Z}_i'\boldsymbol{\Sigma}_i(\psi)^{-1}\left\{h(\mathbf{y}_i,\lambda)-\mathbf{X}_i\boldsymbol{\beta}\right\}.$$

$\Rightarrow h^{-1}(\mu,\lambda)$ を $h^{-1}(\hat{\mu}^{\mathrm{EBLUP}},\hat{\lambda})$ で予測する.

予測区間

前スライドの予測量 $h^{-1}(\hat{\mu}^{\mathrm{EBLUP}}, \hat{\lambda})$ に対してそのリスクを見積もるための予測区間を構成する.

⇒ plug-in による naive な方法 —— $O(m^{-1})$ の精度.

⇒ parametric bootstrap による方法 —— $O(m^{-3/2})$ の精度.

以下の分布を近似するのが本質的.

$$T = \hat{\sigma}^{-1}\left\{\frac{(\lambda\mu + 1)^{\hat{\lambda}/\lambda} - 1}{\hat{\lambda}} - \hat{\mu}^{\mathrm{EBLUP}}\right\}$$

$$\sigma^2 = \mathbf{c}_2'\mathrm{diag}\,(\mathbf{V}_1, \ldots, \mathbf{V}_m)\mathbf{c}_2, \quad \mathbf{V}_i = \mathbf{R}(\psi) - \mathbf{R}(\psi)\mathbf{Z}_i'\boldsymbol{\Sigma}_i^{-1}(\psi)\mathbf{Z}_i\mathbf{R}(\psi).$$

この統計量は $m \to \infty$ で $N(0, 1)$ に収束する (ように構成されている).

予測区間

前スライドの T の分布の分位点 q_L, q_U がわかれば

$$I_m = \left[\left\{ \hat{\lambda}(\hat{\mu}^{\mathrm{EBLUP}} + q_1 \hat{\sigma}) + 1 \right\}^{1/\hat{\lambda}}, \left\{ \hat{\lambda}(\hat{\mu}^{\mathrm{EBLUP}} + q_2 \hat{\sigma}) + 1 \right\}^{1/\hat{\lambda}} \right]$$

のように予測区間が求まる．

T の分布を N(0,1) で近似 $\Rightarrow O(m^{-1})$ の精度

T の分布を parametric bootstrap で近似 $\Rightarrow O(m^{-3/2})$ の精度

Simulation

提案した推定量の MSE を計算し,ML で λ を推定したケースとの比較.

データ生成過程：

$$h(y_{ij}, \lambda) = \beta_0 + \beta_1 x_{ij} + v_i + \varepsilon_{ij}, \quad i = 1, \ldots, 15, \quad j = 1, \ldots, 5$$

$v_i \sim N(0, \boldsymbol{\sigma}_1^2)$, $\varepsilon_{ij} \sim N(0, \boldsymbol{\sigma}_2^2)$, $\boldsymbol{\sigma}_1 = 1, \ \ \boldsymbol{\sigma}_2 = 1.5$

$x_{ij} \sim U(4, 8)$. (最初に発生させて各 run では固定)

$\beta_0 = 1, \ \beta_1 = 2.$

Simulation

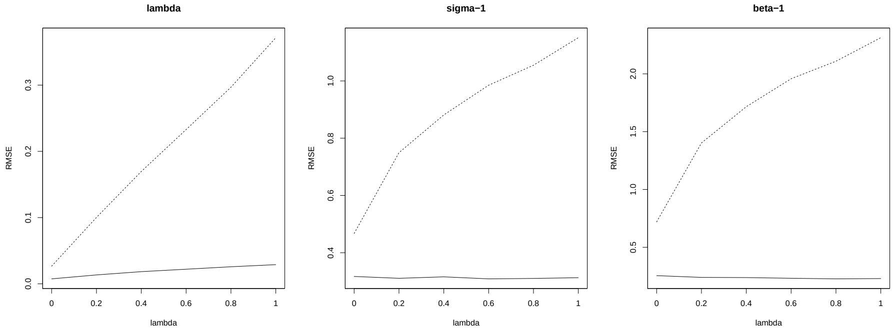


Figure: 5000 回の simulation から計算した推定量の MSE. 実線が提案推定量であり破線が λ を ML で推定したケース.

λ の ML による推定は精度悪く，それが別のパラメータの推定にも波及してしまう.

実データへの応用

2001 年の京急線の公示地価 (PLP) データに BC-NERM を適用する.

y_{ij} : 公示地価 (1000 円単位)
FAR_{ij} : 容積率
DST_{ij} : 最寄り駅からの距離
TRN_i : 駅 i から東京駅までの所用時間 (分)

$$h(y_{ij}, \lambda) = \beta_0 + \beta_1 \log(\mathrm{FAR}_{ij}) + \beta_2 \log(\mathrm{TRN}_i) + \beta_3 \log(\mathrm{DST}_{ij}) + v_i + \varepsilon_{ij},$$

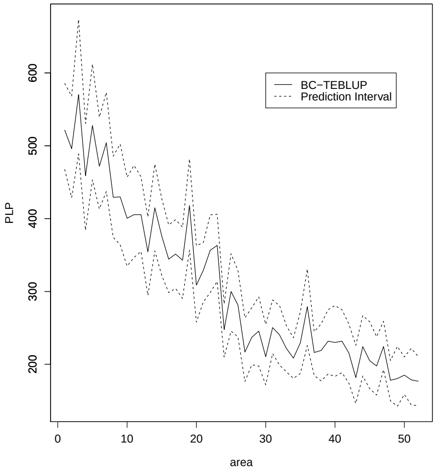
$v_i \sim N(0, \tau^2)$, $\varepsilon_{ij} \sim N(0, \sigma^2)$

パラメータ推定値は

$\hat{\lambda}$	$\hat{\tau}$	$\hat{\boldsymbol{\sigma}}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
0.43	0.67	2.21	49.0	2.05	-7.10	-1.18

実データへの応用

1,000 回のブートストラップにより予測区間を構成すると以下のようになる.



まとめ

本研究では Box-Cox 変換を取り入れた LMM(BC-LMM) を考え，その推定および予測問題について考えた.

⇒ 変換パラメータ λ についてシンプルな推定手法を提案し，その漸近的性質を示した.

⇒ BC-LMM における予測量を提案し，そのリスク評価のための予測区間を構成する手法を提案した.

⇒ BC-LMM の具体例として BC-NERM を考え，PLP データに応用した.