COMSTA: 6485 Model 3G pp. 1–12 (col. fig: NIL)

ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Non-area-specific adjustment factor for second-order efficient empirical Bayes confidence interval

Masayo Yoshimori Hirose

Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo, Japan

A R T I C L E I N F O

Article history:
Received 4 January 2017
Received in revised form 23 May 2017
Accepted 1 July 2017
Available online xxxx

Keywords:
Adjusted maximum-likelihood method
Confidence interval
Empirical best linear unbiased prediction
Fay–Herriot model
Linear mixed model
Small area estimation

A B S T R A C T

An empirical Bayes confidence interval has high user demand in many applications. In particular, the second-order empirical Bayes confidence interval, the coverage error of which is of the third order for a large number of areas, m, is widely used in small area estimation when the sample size within each area is not large enough to make reliable direct estimates according to a design-based approach. Yoshimori and Lahiri (2014a) proposed a new type of confidence interval, called the *second-order efficient empirical Bayes confidence interval*, with a length less than that of the direct confidence estimated according to the design-based approach. However, this interval still has some disadvantages: (i) it is hard to use when at least one leverage value is high; (ii) many iterations tend to be required to obtain the estimators of one global model variance parameter as the number of areas, m, increases, due to the area-specific adjustment factor. To prevent such issues, this study proposes a more efficient confidence interval to allow for high leverage and reduce the number of iterations for large m. To achieve this purpose, we theoretically obtained a non-area-specific adjustment factor and the measure of uncertainty of the empirical Bayes estimator, which consist of empirical Bayes confidence interval, maintaining the existing desired properties. Moreover, we present three simulation results and real data analysis to show overall superiority of our confidence interval method over the other methods, including the one proposed in Yoshimori and Lahiri (2014a).

Contents

E-mail address: masayo@ism.ac.jp.

http://dx.doi.org/10.1016/j.csda.2017.07.002

Please cite this article in press as: Hirose, M.Y., Non-area-specific adjustment factor for second-order efficient empirical Bayes confidence interval. Computational Statistics and Data Analysis (2017), http://dx.doi.org/10.1016/j.csda.2017.07.002.

1. Introduction

There has been increasing demand for reliable statistics of government fund allocations, social services planning, etc., in smaller geographic areas and sub-populations, where large samples are not available. Because of the limited number of observations within each area or domain, a direct estimator constructed according to the design-based approach only from information within each area or domain, is not reliable. The empirical Bayes estimator and empirical best linear unbiased predictor (EBLUP) help make efficient inferences by borrowing information from other areas via model-based approaches to small area estimation. For a comprehensive overview of small area estimation, refer to Rao and Molina (2015). Fay and Herriot (1979) first applied this model-based approach to Census data through a specific Bayesian model. The model, called the Fay–Herriot model, has been widely used in practice. For $i=1,\ldots,m$,

$$\begin{aligned} &\text{Level 1 }: y_i \mid \theta_i \overset{ind.}{\sim} N(\theta_i, D_i);\\ &\text{Level 2 }: \theta_i \overset{ind.}{\sim} N(x_i'\beta, A). \end{aligned} \tag{1}$$

In the above model, level 1 is used to take into account the sampling distribution of the direct estimator y_i for small area i. A true mean for small area i, θ_i, is linked to provide the auxiliary variables $x_i=(x_{i1},\ldots,x_{ip})'$ in a level-2 linking model. In practice, the coefficient p-vector β and the model variance parameter A in the linking model are unknown, and we need to estimate them from the observed data. The assumption of a known D_i often follows from the asymptotic variances of the transformed direct estimates (Efron and Morris, 1975) or from empirical variance modeling (Fay and Herriot, 1979). This model can be viewed as the following linear mixed model:

$$y_i=\theta_i+e_i=x_i'\beta+u_i+e_i, \quad i=1,\ldots,m,$$

where u_i and e_i are mutually independent with the normality assumption $u_i \overset{i.i.d.}{\sim} N(0,A)$ and $e_i \overset{ind.}{\sim} N(0,D_i)$.

Let M_i define the mean squared error (MSE) $E[(\hat{\theta}_i-\theta_i)^2]$ of the predictor $\hat{\theta}_i$ of a small area mean θ_i, where the expectation is on the joint distribution of y and θ under the Fay–Herriot model (1) with $y=(y_1,\ldots,y_m)'$ and $\theta=(\theta_1,\ldots,\theta_m)'$.

The Bayes estimator of θ_i is consistent with the best predictor (BP) in this model, with the minimum MSE among all $\hat{\theta}_i$. It is given by

$$\hat{\theta}_i^{BP}=(1-B_i)y_i+B_ix_i'\beta,$$

where $B_i=\frac{D_i}{A+D_i}$ is called the shrinkage factor toward $x_i'\beta$ from the direct estimate y_i.

If β is unknown, the best linear unbiased predictor (BLUP), in which β of $\hat{\theta}_i^{BP}$ is replaced by $\tilde{\beta}$, minimizes the MSE among all linear unbiased predictors of θ_i, as follows:

$$\hat{\theta}_i^{BLUP}=(1-B_i)y_i+B_ix_i'\tilde{\beta},$$

where the weighted least-square estimator of β, $\tilde{\beta}=\tilde{\beta}(A)=(X'V^{-1}X)^{-1}X'V^{-1}y$, $X=(x_1,\ldots,x_m)'$ and $V=diag(A+D_1,\ldots,A+D_m)$.

From the fact that both β and A are practically unknown, the empirical best linear unbiased predictor (EBLUP), which is consistent with the empirical Bayes estimator in the case, $\hat{\theta}_i^{EB}$ is widely used for small area inference, where the unknown model variance parameter A in $\hat{\theta}_i^{BLUP}$ is replaced by a consistent estimator $\hat{A}$:

$$\hat{\theta}_i^{EB}=(1-\hat{B}_i)y_i+\hat{B}_ix_i'\hat{\beta},$$

where $\hat{B}_i=\frac{D_i}{\hat{A}+D_i}$ and $\hat{\beta}(\hat{A})=\hat{\beta}=\tilde{\beta}(\hat{A})$ and a consistent estimator $\hat{A}$ for large m is even translation invariant for all y_i and β. To estimate the model variance parameter A, the methods of moments estimator (see Fay and Herriot, 1979; Prasad and Rao, 1990) and standard maximum likelihood estimators, such as profile maximum likelihood (ML) estimator and residual maximum likelihood (REML) estimator are utilized. In particular, the REML estimator of A is widely used in terms of higher-order asymptotic properties for large m under some mild regularity conditions. Hereafter, we indicate the REML estimator as $\hat{A}_{RE}$, obtained from

$$\hat{A}_{RE}=\underset{0\le A<\infty}{\arg\max}\, L_{RE}(A|y),$$

where the residual likelihood function $L_{RE}(A|y)=|X'V^{-1}X|^{-1/2}|V|^{-1/2}\exp\{-y'Py/2\}$ and $P=V^{-1}-V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$.

This study focuses on the confidence interval for θ_i, used widely in small area estimation as well as point estimation. Let I_i denote the general form of the confidence interval as follows:

$$I_i: \xi_i \pm q_is_i, \tag{2}$$

Please cite this article in press as: Hirose, M.Y., Non-area-specific adjustment factor for second-order efficient empirical Bayes confidence interval. Computational Statistics and Data Analysis (2017), http://dx.doi.org/10.1016/j.csda.2017.07.002.