Ann Inst Stat Math (2011) 63:1123–1140
DOI 10.1007/s10463-010-0281-x

The best EBLUP in the Fay–Herriot model

Jiming Jiang · En-Tzu Tang

Received: 7 July 2008 / Revised: 10 September 2009 / Published online: 31 March 2010

Abstract We show that in the case of Fay–Herriot model for small area estimation, there is an estimator of the variance of the random effects so that the resulting EBLUP is the best in the sense that it minimizes the leading term in the asymptotic expansion of the mean squared error (MSE) of the EBLUP. In particular, in the balanced case, i.e., when the sampling variances are equal, this best EBLUP has the minimal MSE in the exact sense. We also propose a modified Prasad–Rao MSE estimator which is second-order unbiased and show that it is less biased than the jackknife MSE estimator in a suitable sense in the balanced case. A real data example is discussed.

Keywords Empirical best linear unbiased prediction · Mean squared error · Small area estimation · Third-order approximation

1 Introduction

The empirical best linear unbiased predictor, or EBLUP, is well known in small area estimation (e.g., Rao 2003; Jiang and Lahiri 2006). It is obtained by replacing the unknown variance components in the BLUP by their estimators. Different estimators of the variance components have been used in the literature to get the EBLUP. For example, Prasad and Rao (1990) (P–R hereafter) used the method of moments (MoM, or ANOVA) estimators, while Datta and Lahiri (2000) considered maximum likelihood (ML) and restricted maximum likelihood (REML) estimators. However, it

J. Jiang (✉)
Department of Statistics, University of California, Davis, CA 95616, USA
e-mail: jiang@wald.ucdavis.edu

E.-T. Tang
Amgen Inc, Thousand Oaks, CA 91320, USA

is not very clear how different methods of variance component estimation affects the performance of the EBLUP. Since a measure of the performance of EBLUP (or any other predictor) is its mean squared error (MSE), a natural question is whether there is an estimator of the variance components such that the resulting EBLUP has the minimum MSE, in a suitable sense. In this paper we give answer to this question in the case of Fay–Herriot model.

The Fay–Herriot model Fay and Herriot (1979) is widely used in small area estimation. It was proposed to estimate the per-capita income of small places with population size less than 1,000. The model can be expressed in terms of a mixed effects model:

$$y_i = x_i'\beta + v_i + e_i, \quad i = 1, \ldots, m, \tag{1}$$

where x_i is a vector of known covariates, β is a vector of unknown regression coefficients, v_i's are area-specific random effects, and e_i's represent sampling errors. It is assumed that the v_i's, e_i's are independent with $v_i \sim N(0, A)$ and $e_i \sim N(0, D_i)$. The variance A is unknown, but the sampling variances D_i's are assumed known. Throughout this paper we assume that $A > 0$ and the D_i's are positive and bounded.

The problem of interest is to estimate, or predict, the small area means $\theta_i = x_i'\beta + v_i$, $i = 1, \ldots, m$. The BLUP for θ_i is given by $\tilde{\theta}_i = (1 - B_i)y_i + B_i x_i'\tilde{\beta}$, where $B_i = D_i/(A + D_i)$ and $\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$ with $X = (x_i')_{1\le i\le m}$, $y = (y_i)_{1\le i\le m}$, $V = AI_m + D$ and $D = \mathrm{diag}(D_i, 1 \le i \le m)$. Here I_m represents the m-dimensional identity matrix. Because A is unknown, it is replaced by an estimator, $\hat{A}$. The result is the EBLUP given by

$$\hat{\theta}_i = (1 - \hat{B}_i)y_i + \hat{B}_i x_i'\hat{\beta}, \tag{2}$$

where $\hat{B}_i = D_i/(\hat{A} + D_i)$ and $\hat{\beta}$ is $\tilde{\beta}$ with V replaced by $\hat{V} = \hat{A}I_m + D$. Different estimators of A have been used. For example, the P–R estimator of A is given by

$$\hat{A}_I = \frac{y'P_{X^\perp}y - \mathrm{tr}(P_{X^\perp}D)}{m - p}, \tag{3}$$

where $P_{X^\perp} = I - P_X$ with $P_X = X(X'X)^{-1}X'$. The notation $\hat{A}_I$ will be explained in the sequel. Here, for simplicity, we assume that X has full (column) rank p. Other estimators of A include the ML estimator which may not have a closed-form expression unless the model is balanced in the sense that the D_i's are equal. Because of its simplicity, this latter case will be considered first.

1.1 The balanced case

In this case we may assume, without loss of generality, that $D_i = 1$, $1 \le i \le m$. Then the P–R estimator is given by $\hat{A}_I = \hat{A}^{(p)}$, where for any integer d,

$$\hat{A}^{(d)} = \frac{y'P_{X^\perp}y}{m - d} - 1. \tag{4}$$

Note that, in this case, the P–R estimator is the same as the solution to the REML equation (Jiang 2007). On the other hand, the ML estimator is given by $\hat{A}^{(0)}$, if one ignores the nonnegativity constraint. The question then is: "which d (is the best)?" As mentioned, it makes sense to use the MSE of EBLUP as a criterion for choosing d. Let $\theta = (\theta_i)_{1\le i\le m}$, and $\hat{\theta} = (\hat{\theta}_i)_{1\le i\le m}$. Then, $\mathrm{MSE}(\hat{\theta}) = \mathrm{E}(|\hat{\theta} - \theta|^2) = \sum_{i=1}^m \mathrm{E}(\hat{\theta}_i - \theta_i)^2 = \sum_{i=1}^m \mathrm{MSE}(\hat{\theta}_i)$. Direct calculation in the Appendix shows

$$\mathrm{MSE}(\hat{\theta}) = \frac{Am}{A+1} + \frac{p+2}{A+1} + \frac{(p+2-d)^2}{(A+1)(m-p-2)}. \tag{5}$$

Clearly, this expression is minimized when

$$d = p + 2, \tag{6}$$

which is the best d one is looking for. Thus, in particular, neither P–R (REML) nor ML are the best.

1.2 The unbalanced case

Unfortunately, analytic expression of $\mathrm{MSE}(\hat{\theta})$ is not available in the unbalanced case. However, it can be shown (see Sect. 2) that the two leading terms of the asymptotic expansion of $\mathrm{MSE}(\hat{\theta})$, which are $O(m)$ and $O(1)$, do not depend on the choice of the estimator of A so long as it is $\sqrt{m}$-consistent. However, the next term which is $O(m^{-1})$ does depend on the choice of $\hat{A}$. Therefore, the question is: "which A minimizes the $O(m^{-1})$ term in the asymptotic expansion of $\mathrm{MSE}(\hat{\theta})$?" In Sect. 2 we answer this question within a class of estimators of A that includes the P–R estimator as a special case.

1.3 Estimation of MSE

Given the best EBLUP, the next question is how to estimate its MSE. Two of the best-known methods that provide second-order unbiased MSE estimators are the Prasad–Rao method (Prasad and Rao 1990), and the jackknife method (Jiang et al. 2002). The P–R method is second-order unbiased only when the P–R estimator of A, that is, $\hat{A}_I$, is used. In Sect. 3 we propose a modified P–R MSE estimator so that it maintains second-order unbiasedness for the best EBLUP. On the other hand, for the jackknife MSE estimator to be second-order unbiased, it is required that the estimator of A be a component of an M-estimator of the model parameters, a condition not necessarily satisfied by the estimator of A corresponding to the best EBLUP. Furthermore, it is shown that in the balanced case, where both the modified P–R and jackknife MSE estimators are second-order unbiased for the best EBLUP, the second-order term [i.e., the $O(m^{-2})$ term] of the bias of the modified P–R MSE estimator is smaller, in absolute value, than that of the jackknife MSE estimator uniformly across the small areas (i.e., for any $1 \le i \le m$).

The derivations are outlined in the Appendix with more details given in a technical report.

2 The best EBLUP (unbalanced case)

Throughout the rest of this paper, we assume that $\hat{A}$ is a $\sqrt{m}$-consistent estimator in the sense that $\hat{A} - A = O(m^{-1/2})$, where and hereafter the big O and small o are understood in a suitable sense in each occasion (e.g., in probability for random variables). Let $B_i(A) = B_i = D_i/(A + D_i)$, $\theta_i(A) = \tilde{\theta}_i$ given above (2), and $P(A) = (X'V^{-1}X)^{-1}X'V^{-1}$. Then, we have

$$\begin{aligned}\text{MSE}\ (\hat{\theta}_i) &= \text{E}(\hat{\theta}_i - \theta_i)^2 \\ &= \text{E}\ \{\hat{\theta}_i -\ \text{E}\ (\theta_i|y)\}^2 +\ \text{E}\ \{\ \text{E}\ (\theta_i|y) - \theta_i\}^2 \qquad (7)\end{aligned}$$

with $E(\theta_i|y) = (1 - B_i)y_i + B_i x_i'\beta = \theta_i(A) - B_i x_i'(\tilde{\beta} - \beta)$, where $\tilde{\beta}$ is given above (2). Simple calculation shows that

$$\text{E}\{\text{E}(\theta_i|y) - \theta_i\}^2 = \frac{AD_i}{A + D_i}. \qquad (8)$$

Furthermore, by Taylor expansion, we have

$$\begin{aligned}\hat{\theta}_i -\ \text{E}\ (\theta_i|y) &= B_i x_i'(\tilde{\beta} - \beta) + \theta_i'(A)(\hat{A} - A) + \frac{1}{2}\theta_i''(A)(\hat{A} - A)^2 \\ &\quad + \frac{1}{6}\theta_i'''(A)(\hat{A} - A)^3 + O(m^{-2}). \qquad (9)\end{aligned}$$

By 7–9, it can be shown that

$$\begin{aligned}\text{MSE}\ (\hat{\theta}) &= \sum_{i=1}^{m} \frac{AD_i}{A + D_i} + \sum_{i=1}^{m} \frac{D_i^2}{(A + D_i)^2} x_i'(X'V^{-1}X)^{-1}x_i \\ &\quad + 2\sum_{j=1}^{4} \frac{1}{j!} E\{(\hat{A} - A)^j u' G_j u\} + o(m^{-1}), \qquad (10)\end{aligned}$$

where $G_j = G_j(A) = \sum_{i=1}^{m} G_{j,i}(A)$, $j = 1, 2, 3, 4$ with $G_{1,i}(A) = B_i(A)V^{1/2}P(A)'$ $x_i H_{1,i}(A)$,

$$H_{1,i}(A) = \{B_i'(A)x_i'P(A) + B_i(A)x_i'P'(A) - B_i'(A)\delta_i'\}V^{1/2},$$

$\delta_i = (1_{(k=i)})_{1 \le k \le m}$; $G_{2,i}(A) = B_i(A)V^{1/2}P(A)'x_i H_{2,i}(A) + H_{1,i}(A)'H_{1,i}(A)$,

$$H_{2,i}(A) = \{B_i''(A)x_i'P(A) + 2B_i'(A)x_i'P'(A) + B_i(A)x_i'P''(A) - B_i''(A)\delta_i'\}V^{1/2};$$

$G_{3,i} = B_i(A)V^{1/2}P(A)'x_i H_{3,i}(A) + 3H_{1,i}(A)'H_{2,i}(A),$

$$H_{3,i}(A) = \left\{B_i'''(A)x_i'P(A) + 3B_i''(A)x_i'P'(A) + 3B_i'(A)x_i'P''(A)\right.$$
$$\left. + B_i(A)x_i'P'''(A) - B_i'''(A)\delta_i'\right\} V^{1/2};$$

$G_{4,i}(A) = 4H_{1,i}(A)'H_{3,i}(A) + 3H_{2,i}(A)'H_{2,i}(A)$; and $u = V^{-1/2}(y - X\beta) \sim N(0, I_m)$. Hereafter, the notation $P(A)'$ denotes transpose while $P'(A)$, $P''(A)$ and $P'''(A)$ denote first, second, and third derivatives with respect to A.

2.1 Class of estimators

We consider an estimator of A, $\hat{A}$, that has the following asymptotic expansion:

$$\hat{A} - A = \left(1 + \frac{c}{m}\right)\frac{u'Qu - \text{tr}(Q)}{m - p} + \frac{\hat{c}\text{tr}(Q)}{m(m - p)} + O(m^{-2}), \tag{11}$$

where $Q = V^{1/2}P_{X^\perp}V^{1/2}$, $c = c(A)$ is a continuously differentiable function of A, $\hat{c}$ is an estimator of c that satisfies $\hat{c} - c = O(m^{-1/2})$, and

$$E[\{u'Qu - \text{tr}\,(Q)\}(\hat{c} - c)] = o(1). \tag{12}$$

The motivation of (11), (12) is the following. Ideally, one would like to extend the result of Sect. 1.1 by considering an estimator of the following form:

$$\hat{A} = \left(1 + \frac{c}{m}\right)\frac{y'P_{X^\perp}y}{m - p} - \frac{\text{tr}(P_{X^\perp}D)}{m - p},$$

where c is a constant to be determined. This estimator satisfies

$$\hat{A} - A = \left(1 + \frac{c}{m}\right)\frac{u'Qu - \text{tr}(Q)}{m - p} + \frac{c\text{tr}(Q)}{m(m - p)},$$

which is (11) with $\hat{c} = c$ and $O(m^{-2}) = 0$. Unfortunately, as is shown below, in the unbalanced case the best c depends on A, i.e., $c = c(A)$ (so is no longer a constant). In this situation one would naturally consider an estimated version of the best c, for example, by estimating $c(A)$ by $c(\hat{A})$. However, the estimation results in a bias that potentially affects the optimality of the resulting EBLUP. Therefore, what one needs is a bias-corrected estimator of c and (12) turns out to be what is needed for the bias correction. Note that, typically, one has $u'Qu - \text{tr}(Q) = O(m^{1/2})$ and $\hat{c} - c = O(m^{-1/2})$; hence one would normally expect the left side of (12) to be $O(1)$. What (12) needs is a little bit further reduction of the bias. We consider some special cases of (11) and (12).

Example 1 The P–R estimator $\hat{A}_I$ of (3) corresponds to (11) and (12) with $c = \hat{c} = 0$ and $O(m^{-2}) = 0$. It clearly satisfies (12).

Example 2 Suppose that $\hat{c}$ is an estimator of c that satisfies $\hat{c} - c = O(m^{-1/2})$. Then, the estimator

$$\hat{A} = \left(1 + \frac{\hat{c}}{m}\right) \frac{y' P_{X^\perp} y}{m - p} - \frac{\mathrm{tr}(P_{X^\perp} D)}{m - p} \tag{13}$$

satisfies (11). To see this, note that it is easy to obtain the following expression:

$$\begin{aligned}\hat{A} - A &= \left(1 + \frac{\hat{c}}{m}\right) \frac{u'Qu - \mathrm{tr}(Q)}{m - p} + \frac{\hat{c}\mathrm{tr}(Q)}{m(m - p)} \\ &= \left(1 + \frac{c}{m}\right) \frac{u'Qu - \mathrm{tr}(Q)}{m - p} + \frac{\hat{c}\mathrm{tr}(Q)}{m(m - p)} + O(m^{-2}).\end{aligned}$$

2.2 Further asymptotic expansion

It can be shown that for an estimator satisfying (11) and (12), (10) can be further expressed as

$$\begin{aligned}\mathrm{MSE}(\hat{\theta}) &= A \sum_{i=1}^{m} \frac{D_i}{A + D_i} + \sum_{i=1}^{m} \frac{D_i^2}{(A + D_i)^2} x_i'(X'V^{-1}X)^{-1}x_i + \frac{\mathrm{tr}^2(Q)\mathrm{tr}(F_2)}{m^2(m - p)^2} c^2 \\ &\quad + \frac{2}{m(m - p)} \left(2\mathrm{tr}(QG_1) + \mathrm{tr}(Q)\mathrm{tr}\,(G_1) + \frac{E[\{u'Qu - \mathrm{tr}(Q)\}^2 u'G_2u]}{m - p}\right. \\ &\quad \left. + \frac{\mathrm{tr}(Q)E[\{u'Qu - \mathrm{tr}(Q)\}u'G_2u]}{m - p} + \frac{\mathrm{tr}(Q)E[\{u'Qu - \mathrm{tr}(Q)\}^2 u'G_3u]}{2(m - p)^2}\right) c \\ &\quad + \frac{4\mathrm{tr}(QG_1)}{m - p} + \frac{E[\{u'Qu - \mathrm{tr}(Q)\}^2 u'G_2u]}{(m - p)^2} + \frac{E[\{u'Qu - \mathrm{tr}(Q)\}^3 u'G_3u]}{3(m - p)^3} \\ &\quad + \frac{E[\{u'Qu - \mathrm{tr}(Q)\}^4 u'G_4u]}{12(m - p)^4} + o(m^{-1}),\end{aligned} \tag{14}$$

where $F_j = D^2 V^{-j-1}$, $j = 1, 2, \ldots$. To obtain a more explicit expression we need expressions for $E[\{u'Qu - \mathrm{tr}(Q)\}^k u'Gu]$ for $k = 1, 2, 3, 4$ and any matrix G. Lemma 1 in the Appendix provides such expressions. With these results, (14) can be further expressed as

$$\begin{aligned}\mathrm{MSE}(\hat{\theta}) &= A \sum_{i=1}^{m} \frac{D_i}{A + D_i} + \sum_{i=1}^{m} \frac{D_i^2}{(A + D_i)^2} x_i'(X'V^{-1}X)^{-1}x_i \\ &\quad + \frac{2\mathrm{tr}(U^2)\mathrm{tr}(P_{X^\perp} R)}{(m - p)^2} + \frac{\mathrm{tr}^2(U)\mathrm{tr}(P_{X^\perp} F_2)}{m(m - p)^3} c^2 \\ &\quad + \frac{4}{m} \left\{\frac{\mathrm{tr}(U)\mathrm{tr}(P_{X^\perp} F_1) + \mathrm{tr}(U^2)\mathrm{tr}(P_{X^\perp} F_2)}{(m - p)^2} - \frac{3\,\mathrm{tr}(U)\,\mathrm{tr}(U^2)\mathrm{tr}(P_{X^\perp} F_3)}{(m - p)^3}\right\} c\end{aligned}$$

$$+\frac{8\,\mathrm{tr}(U^2R)}{m(m-p)}-\frac{24\,\mathrm{tr}(U^2)\,\mathrm{tr}(P_{X^\perp}F_2)+16\,\mathrm{tr}(U^3)\,\mathrm{tr}(P_{X^\perp}F_3)}{m(m-p)^2}$$
$$+\frac{36\,\mathrm{tr}^2(U^2)\,\mathrm{tr}(P_{X^\perp}F_4)}{m(m-p)^3}+o(m^{-1}), \tag{15}$$

where $U=P_{X^\perp}V$ and $R=H'HV$ with $H=D\{V^{-1}-V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}\}^2$.

At this point we would like to take a break by considering a special case.

Example 3 (*The balanced case*) Consider the balanced case discussed in Sect. 1.1. It is easy to see that, in this case, $R=(A+1)^{-3}P_{X^\perp}$, and the right side of (15) reduces to

$$\frac{Am}{A+1}+\frac{p+2}{A+1}+\frac{(c-2)^2}{(A+1)m}+o(m^{-1}).$$

This is exactly the same as the right side of (5) with $p+2-d$ replaced by $c-2$ and

$$o(m^{-1})=\frac{(p+2)(c-2)^2}{(A+1)m(m-p-2)}.$$

The best EBLUP therefore has $c=2$, which corresponds to (6) if we consider the estimators $\hat{A}^{(d)}$ defined by (4). By the way, $\hat{A}^{(d)}$ satisfies (11) and (12) for any d in the balanced case.

There is an alternative expression of the MSE in terms of the positive eigenvalues of the matrix $S=P_{X^\perp}DP_{X^\perp}$, denoted by $\lambda_1,\ldots,\lambda_{m-p}$. Since $P_{X^\perp}S=SP_{X^\perp}$, there is an orthogonal matrix T such that $S=T\mathrm{diag}(\lambda_1,\ldots,\lambda_{m-p},0,\ldots,0)T'$ and $P_{X^\perp}=T\mathrm{diag}(\mu_1,\ldots,\mu_m)T'$ (Searle 1982), where $\mu_i=0$ or 1. Then, since $SP_{X^\perp}=S$, it follows that $\lambda_i\mu_i=\lambda_i$, $1\le i\le m-p$, which imply that $\mu_i=1$, $1\le i\le m-p$ and hence $\mu_i=0$, $i>m-p$. From this result and also the fact that $R=P_{X^\perp}R$, it follows by (15) that

$$\mathrm{MSE}\,(\hat{\theta})=a_0+b_0+2b_1c+b_2c^2+o(m^{-1}), \tag{16}$$

where

$$a_0=A\sum_{i=1}^m\frac{D_i}{A+D_i}+\sum_{i=1}^m\frac{D_i^2}{(A+D_i)^2}x_i'(X'V^{-1}X)^{-1}x_i+2\frac{\mathrm{tr}(U^2)\mathrm{tr}(P_{X^\perp}R)}{(m-p)^2},$$

$$b_0=\frac{8}{m}\cdot\frac{\mathrm{tr}(U^2P_{X^\perp}R)}{m-p}-\frac{24}{m}\left\{\frac{\sum_{i=1}^{m-p}(A+\lambda_i)^2}{m-p}\right\}\left\{\frac{\sum_{i=1}^{m-p}\lambda_i^2(A+\lambda_i)^{-3}}{m-p}\right\}$$
$$-\frac{16}{m}\left\{\frac{\sum_{i=1}^{m-p}(A+\lambda_i)^3}{m-p}\right\}\left\{\frac{\sum_{i=1}^{m-p}\lambda_i^2(A+\lambda_i)^{-4}}{m-p}\right\}$$

$$+\frac{36}{m}\left\{\frac{\sum_{i=1}^{m-p}(A+\lambda_i)^2}{m-p}\right\}^2\left\{\frac{\sum_{i=1}^{m-p}\lambda_i^2(A+\lambda_i)^{-5}}{m-p}\right\},$$

$$b_1=\frac{2}{m}\left\{\frac{\sum_{i=1}^{m-p}(A+\lambda_i)}{m-p}\right\}\left\{\frac{\sum_{i=1}^{m-p}\lambda_i^2(A+\lambda_i)^{-2}}{m-p}\right\}$$

$$+\frac{2}{m}\left\{\frac{\sum_{i=1}^{m-p}(A+\lambda_i)^2}{m-p}\right\}\left\{\frac{\sum_{i=1}^{m-p}\lambda_i^2(A+\lambda_i)^{-3}}{m-p}\right\}$$

$$-\frac{6}{m}\left\{\frac{\sum_{i=1}^{m-p}(A+\lambda_i)}{m-p}\right\}\left\{\frac{\sum_{i=1}^{m-p}(A+\lambda_i)^2}{m-p}\right\}\left\{\frac{\sum_{i=1}^{m-p}\lambda_i^2(A+\lambda_i)^{-4}}{m-p}\right\},$$

$$b_2=\frac{1}{m}\left\{\frac{\sum_{i=1}^{m-p}(A+\lambda_i)}{m-p}\right\}^2\left\{\frac{\sum_{i=1}^{m-p}\lambda_i^2(A+\lambda_i)^{-3}}{m-p}\right\}.$$

The best EBLUP therefore corresponds to the c that minimizes the quadratic function of c on the right side of (16), that is, $c=-b_1/b_2$, or

$$\begin{aligned}c(A)=&\,6\frac{\left\{\sum_{i=1}^{m-p}(A+\lambda_i)^2\right\}\left\{\sum_{i=1}^{m-p}\lambda_i^2(A+\lambda_i)^{-4}\right\}}{\left\{\sum_{i=1}^{m-p}(A+\lambda_i)\right\}\left\{\sum_{i=1}^{m-p}\lambda_i^2(A+\lambda_i)^{-3}\right\}}\\&-2\frac{(m-p)\sum_{i=1}^{m-p}\lambda_i^2(A+\lambda_i)^{-2}}{\left\{\sum_{i=1}^{m-p}(A+\lambda_i)\right\}\left\{\sum_{i=1}^{m-p}\lambda_i^2(A+\lambda_i)^{-3}\right\}}\\&-2\frac{(m-p)\sum_{i=1}^{m-p}(A+\lambda_i)^2}{\left\{\sum_{i=1}^{m-p}(A+\lambda_i)\right\}^2}.\end{aligned}\tag{17}$$

2.3 Estimation of c

The above arguments showed that, if $\hat{A}$ is given by (13) with c given by (17) and $\hat{c}$ satisfying $\hat{c}-c=O(m^{-1/2})$ and (12), then the corresponding EBLUP is the best in the sense that the leading term in its MSE expansion is minimized. The question that remains to be answered is: "is there such an estimator $\hat{c}$ in the world?" The answer is yes, and the estimator is given below.

Consider the following class of unbiased estimators of A:

$$\hat{A}_W=\frac{y'P_{X^\perp}WP_{X^\perp}y-\ \mathrm{tr}(P_{X^\perp}WP_{X^\perp}D)}{\mathrm{tr}(P_{X^\perp}W)},\tag{18}$$

where W is a known weighting matrix. The P–R estimator of (3) is a special case with $W=I$, the (m-dimensional) identity matrix (and this explains the notation therein).

Then consider the following estimator of c

$$\hat{c} = c(\hat{A}_I) + d(\hat{A}_I)(\hat{A}_I - \hat{A}_D), \tag{19}$$

where $d(A) = 0$ if all the λ_i, $1 \le i \le m - p$ [given above (16)] are equal; otherwise,

$$d(A) = \frac{2\left(\sum_{i=1}^{m-p} \lambda_i\right)\left\{\sum_{i=1}^{m-p}(A + \lambda_i)^2\right\}}{\sum_{1\le i\ne j\le m-p}(2A + \lambda_i + \lambda_j)(\lambda_i - \lambda_j)^2} c'(A). \tag{20}$$

To see that the estimator $\hat{c}$ given above satisfies the requirements, first note that, if the λ_i's are all equal, then by (17) we have $c = c(A) = 2$; hence by (19) $\hat{c} = 2$, which satisfies the requirements. Now suppose that the λ_i's are not all equal. It is shown in the Appendix that $d(A)$ is bounded and continuous in A. Therefore, $d(\hat{A}_I) - d(A) = o(1)$. It follows that

$$\begin{aligned}\hat{c} - c &= c(\hat{A}_I) - c(A) + d(A)(\hat{A}_I - \hat{A}_D) + \{d(\hat{A}_I) - d(A)\}(\hat{A}_I - \hat{A}_D) \\ &= c'(A)(\hat{A}_I - A) + d(A)(\hat{A}_I - \hat{A}_D) + o(m^{-1/2}).\end{aligned}$$

Therefore, $E[\{u'Qu - \mathrm{tr}(Q)\}(\hat{c} - c)] = (m - p)\{c'(A)E(\hat{A}_I - A)^2 + d(A)E(\hat{A}_I - A)(\hat{A}_I - \hat{A}_D)\} + o(1) = o(1)$, because $c'(A)E(\hat{A}_I - A)^2 + d(A)E(\hat{A}_I - A)(\hat{A}_I - \hat{A}_D) = 0$.

2.4 Conclusion

The best EBLUP is given by (2) with $\hat{A}$ given by (13), where $c = c(A)$ is given by (17) and $\hat{c}$ is given by (19), where $d(A) = 0$ if all the λ_i's are all equal, and $d(A)$ is given by (20) otherwise.

3 Estimating MSE of the best EBLUP

As mentioned earlier in Sect. 1, the original P–R MSE estimator is second-order unbiased only when $\hat{A} = \hat{A}_I$ is used to estimate A. If the best EBLUP is used with $\hat{A}$ given by Sect. 2.4, we need to modify the P–R MSE estimator so that it remains second-order unbiased. The modified P–R MSE estimator is given by the following:

$$\begin{aligned}\widehat{\mathrm{MSE}}_{\mathrm{MPR},i} = \frac{\hat{A}D_i}{\hat{A} + D_i} + \frac{D_i^2}{(\hat{A} + D_i)^2} x_i'(X'\hat{V}^{-1}X)^{-1}x_i \\ + \frac{4D_i^2 \mathrm{tr}\{(P_{X^\perp}\hat{V})^2\}}{(\hat{A} + D_i)^3 m(m - p)} - \frac{D_i^2 \mathrm{tr}(P_{X^\perp}\hat{V})}{(\hat{A} + D_i)^2 m(m - p)} c(\hat{A}),\end{aligned} \tag{21}$$

where $\hat{V} = \hat{A}I_m + D$ and $c(A)$ is given by (17). It is shown in Appendix that Bias $(\widehat{\mathrm{MSE}}_{\mathrm{MPR},i}) = E(\widehat{\mathrm{MSE}}_{\mathrm{MPR},i}) - \mathrm{MSE}(\hat{\theta}_i) = o(m^{-1})$.

Example 3 (*continued*) It is easy to show that, in this case, we have

$$\widehat{\text{MSE}}_{\text{MPR},i} = \frac{\hat{A}}{\hat{A}+1} + \frac{x_i'(X'X)^{-1}x_i}{\hat{A}+1} + \frac{2}{(\hat{A}+1)m}. \tag{22}$$

As a comparison, the original P–R MSE estimator under the same situation is given by

$$\widehat{\text{MSE}}_{\text{PR},i} = \frac{\hat{A}_I}{\hat{A}_I+1} + \frac{x_i'(X'X)^{-1}x_i}{\hat{A}_I+1} + \frac{4}{(\hat{A}_I+1)m}. \tag{23}$$

Note the difference in estimators of A in (22) and (23), in addition to the constants 2 and 4. Furthermore, by using the relation $\hat{A} = \hat{A}_I + 2(\hat{A}_I+1)/m$, it can be shown that $\widehat{\text{MSE}}_{\text{MPR},i} = \widehat{\text{MSE}}_{\text{PR},i} + O_{\text{P}}(m^{-2})$.

An alternative MSE estimator is the jackknife estimator (Jiang et al. 2002), given by

$$\begin{aligned}\widehat{\text{MSE}}_{\text{J},i} &= \frac{\hat{A}D_i}{\hat{A}+D_i} - \frac{m-1}{m}\sum_{j=1}^{m}\left(\frac{\hat{A}_{-j}}{\hat{A}_{-j}+D_i} - \frac{\hat{A}}{\hat{A}+D_i}\right)D_i \\ &\quad + \frac{m-1}{m}\sum_{j=1}^{m}(\hat{\theta}_{i,-j} - \hat{\theta}_i)^2,\end{aligned}$$

where $\hat{\theta}_i$ is the EBLUP given by (2),

$$\begin{aligned}\hat{\theta}_{i,-j} &= \frac{\hat{A}_{-j}}{\hat{A}_{-j}+D_i}y_i + \frac{D_i}{\hat{A}_{-j}+D_i}x_i'\hat{\beta}_{-j}, \\ \hat{\beta}_{-j} &= \left(\sum_{k\neq j}\frac{D_k}{\hat{A}_{-j}+D_k}x_k x_k'\right)^{-1}\sum_{k\neq j}\frac{D_k}{\hat{A}_{-j}+D_k}x_k y_k,\end{aligned}$$

and $\hat{A}_{-j}$ is the estimator of A obtained in the same way as $\hat{A}$ but with data from the jth small area deleted. For example, if $\hat{A} = \hat{A}_I$ given by (3), then

$$\hat{A}_{-j} = \frac{y_{-j}'P_{X_{-j}^{\perp}}y_{-j} - \text{tr}(P_{X_{-j}^{\perp}}D_{-j})}{m-1-p},$$

where $y_{-j} = (y_1,\ldots,y_{j-1},y_{j+1},\ldots,y_m)'$, $X_{-j} = (x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_m)'$ and $D_{-j} = \text{diag}(D_1,\ldots,D_{j-1},D_{j+1},\ldots,D_m)$.

For the jackknife MSE estimator to be second-order unbiased, it is required that $\hat{A}$ is a component of an M-estimator (Jiang et al. 2002). Here, by M-estimator it means an estimator of $\phi = (\beta', A)'$ which is a solution to a system of equations of the following type: $\sum_{i=1}^{m} f_i(\phi, y_i) + a(\phi) = 0$, where $f_i(\phi, y_i)$ satisfies $E\{f_i(\phi, y_i)\} = 0$, if ϕ is the true parameter vector, and $a(\phi)$ is a (vector-valued) function which may depend on the joint distribution of y. For example, it is easy to verify that any estimator $\hat{\phi} = (\hat{\beta}, \hat{A})$ is an M-estimator, where $\hat{\beta}$ is given below (2) and $\hat{A} = \hat{A}_W$ given by (18), provided that W is diagonal. However, the estimator $\hat{A}$ of Sect. 2.4 corresponding to the best EBLUP does not necessarily satisfy this requirement.

One special case in which both the modified P–R and the jackknife MSE estimators are second-order unbiased is the balanced case of Sect. 1.1, because in this case it is straightforward to show that the estimator $\hat{A}$ of Sect. 2.4 is a component of an M-estimator. Although both estimators are second-order unbiased, it can be shown that the modified P–R MSE estimator is less biased than the jackknife MSE estimator, and this is true uniformly across all the small areas. More specifically, we have the following result. For the best EBLUP, which corresponds to (4) with d given by (6), we have, for any $1 \le i \le m$,

$$\begin{aligned} &\text{Bias}\left(\widehat{\text{MSE}}_{\text{MPR},i}\right) \\ &\quad = \frac{1}{A+1}\left\{\frac{2}{m}x_i'(X'X)^{-1}x_i - \frac{2p}{m^2}\right\} + o(m^{-2}), \end{aligned} \tag{24}$$

$$\begin{aligned} &\text{Bias}\left(\widehat{\text{MSE}}_{\text{J},i}\right) \\ &\quad = \frac{1}{A+1}\left\{\frac{1}{m}x_i'(X'X)^{-1}x_i + x_i'Q_X x_i + \frac{2(p+4)}{m^2}\right\} + o(m^{-2}), \end{aligned} \tag{25}$$

where $Q_X = (X'X)^{-1}\{\sum_{j=1}^{m} x_j x_j'(X'X)^{-1}x_j x_j'\}(X'X)^{-1}$. Furthermore, it can be shown that $x_i'Q_X x_i \ge (1/m)x_i'(X'X)^{-1}x_i$. It follows that the leading term of Bias $(\widehat{\text{MSE}}_{\text{MPR},i})$ is smaller in absolute value than that of Bias$(\widehat{\text{MSE}}_{\text{J},i})$ for any i.

For these reasons we recommend to use the modified P–R estimator (21) for estimating the MSE of the best EBLUP.

4 A real data example

Morris and Christiansen (1996) reported data from a medical survey involving 23 hospitals (out of a total of 219 hospitals) that had at least 50 kidney transplants during a 27-month period (see Table 1). Here the y_i's are graft failure rates for kidney transplant operations, that is, y_i = number of graft failures$/n_i$, where n_i is the number of kidney transplants at hospital i during the period of interest. The variance for graft failure rate, D_i, is approximated by $(0.2)(0.8)/n_i$, where 0.2 is the observed failure rate for all hospitals. Thus, D_i is known for all i. In addition, a severity index x_i is available for each hospital, which is the average fraction of females, blacks, children, and extremely ill kidney recipients at hospital i.

Ganesh (2009) proposed a Fay–Herriot model for the graft failure rates, using x_i as the only covariate. However, inspections of the raw data suggest some nonlinear trends (see Fig. 1). This suggests that transformations may be needed before applying the

Table 1 Data from Morris and Christiansen (1996) and results of analysis

Area	logit(y_i)	x_i	D_i^*	$\hat{\theta}_{\text{BEB}}$ (MSE in %)	$\hat{\theta}_{\text{PREB}}$ (MSE in %)
1	−0.838	0.112	0.118	−1.155 (3.488)	−1.202 (3.158)
2	−1.815	0.206	0.110	−1.572 (3.296)	−1.541 (3.004)
3	−1.368	0.104	0.106	−1.314 (3.348)	−1.316 (3.062)
4	−0.695	0.168	0.106	−1.140 (3.105)	−1.203 (2.807)
5	−0.632	0.337	0.086	−0.623 (8.808)	−0.620 (9.492)
6	−1.289	0.169	0.083	−1.287 (2.998)	−1.293 (2.809)
7	−1.688	0.211	0.083	−1.572 (3.281)	−1.553 (3.134)
8	−1.791	0.195	0.083	−1.536 (2.986)	−1.498 (2.805)
9	−1.266	0.221	0.076	−1.481 (3.505)	−1.514 (3.444)
10	−1.355	0.077	0.076	−1.544 (3.381)	−1.576 (3.246)
11	−1.331	0.195	0.069	−1.399 (2.887)	−1.413 (2.798)
12	−1.015	0.185	0.066	−1.252 (2.805)	−1.300 (2.730)
13	−1.153	0.202	0.066	−1.366 (2.943)	−1.406 (2.895)
14	−1.036	0.108	0.051	−1.176 (2.860)	−1.216 (2.981)
15	−1.782	0.204	0.051	−1.604 (2.760)	−1.563 (2.865)
16	−2.031	0.072	0.048	−1.847 (3.146)	−1.802 (3.338)
17	−1.380	0.142	0.043	−1.276 (2.703)	−1.258 (2.928)
18	−1.313	0.136	0.040	−1.245 (2.659)	−1.236 (2.926)
19	−1.457	0.172	0.038	−1.374 (2.372)	−1.356 (2.601)
20	−1.313	0.202	0.033	−1.393 (2.286)	−1.418 (2.606)
21	−1.614	0.087	0.033	−1.545 (2.327)	−1.527 (2.646)
22	−1.565	0.177	0.028	−1.456 (2.060)	−1.423 (2.402)
23	−1.621	0.072	0.024	−1.662 (2.177)	−1.676 (2.690)

Fay–Herriot model. Since the data are proportions, a logit transformation is appropriate. Note that the D_i's after the logit transformation can be approximated by the Delta method, namely, if $E(y) = p$ and $\text{var}(y) = D$, then $\text{var}\{\text{logit}(y)\} \approx D/\{p(1-p)\}^2$. The transformed D_i is denoted by D_i^*. After the logit transformation some nonlinear trends still persist. However, a closer look at the plot of the transformed data suggests that a cubic regression may be more appropriate for the mean function than linear (or quadratic) regression. Thus, the following Fay–Herriot model is proposed:

$$\text{logit}(y_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + v_i + e_i, \tag{26}$$

$i = 1, \dots, 23$, where β_j, $j = 0, 1, 2, 3$ are unknown coefficients and everything else is the same as in the Fay–Herriot model (1). It is clear that (26) is a special case of (1).

We apply the methods of Sects. 2 and 3 to the transformed data. The best EBLUPs (BEB) are computed for all the small areas, along with their modified P–R MSE estimates. As a comparison, the P–R EBLUPs, that is, EBLUPs with A estimated by $\hat{A}_I$ (PREB) are also computed with their (original) P–R MSE estimates. The original P–R MSE estimator (or P–R MSE estimator without modification) is appropriate for

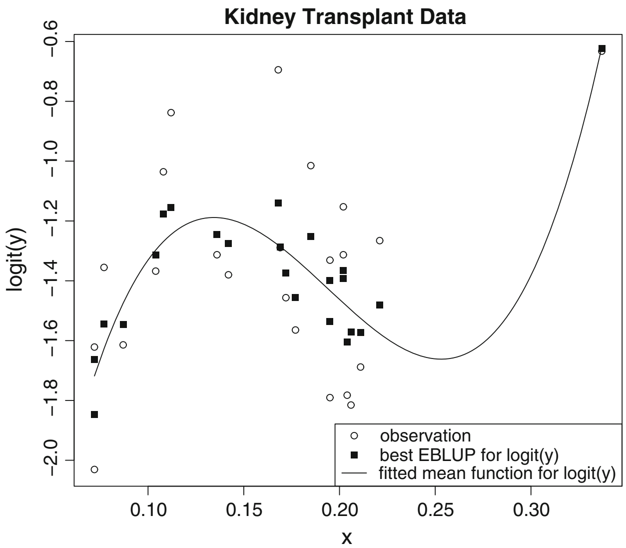

Fig. 1 Data from Morris and Christiansen (1996) after logit transformation, the best EBLUP, and the fitted mean function

the PREB, because $\hat{A}_I$ is the Prasad–Rao estimator of A. The results are reported in Table 1. The estimated overall MSE, in %, is 72.18 for BEB and 73.37 for PREB. Note that these are the estimated MSEs rather than the true MSEs. Also note that, although the BEB has the smallest overall MSE, it does not imply that its MSE is the smallest for all individual small areas. This can be seen from Table 1. Again, these are the estimated MSEs rather than the true MSEs.

The best EBLUPs are plotted in Fig. 1. In addition, the estimated β coefficients in (26) are $\hat{\beta}_0 = -4.38$, $\hat{\beta}_1 = 57.75$, $\hat{\beta}_2 = -328.94$ and $\hat{\beta}_3 = 565.90$. A fitted mean function curve, that is, $\hat{f}(x) = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \hat{\beta}_3 x^3$ is also plotted in Fig. 1.

5 Concluding remark and open problem

The $\hat{A}$ for best EBLUP is within a class of estimators defined by (11). Although this class includes the P–R estimator, it does not include, for example, the ML and REML estimators; neither does it include the estimator proposed by Fay and Herriot in their original paper (1979; also see Pfeffermann and Nathan 1981; FH–PN hereafter), with the exception of some special cases. The reason is that, typically, asymptotic expansion of an estimator such as the ML/REML estimator (or any estimator that is a solution to an equation that does not have a closed-form expression) includes an $O(m^{-3/2})$ term, while (11) assumes that this term is absent. Datta et al. (2005) showed by simulation comparisons that, to the $O(m^{-2})$ order of approximation, the smallest MSE of the EBLUP results from the ML/REML estimators, next in line is the FH–PN estimator, and the last is the P–R estimator. An interesting open problem therefore is to compare, both theoretically and empirically, these other estimators with the estimator proposed in this paper in terms of the MSE of the EBLUP.

Acknowledgments Jiming Jiang's research is partially supported by NSF grant DMS-0402824. The authors are grateful to the referees for their valuable comments.

Appendix

In this appendix, we outline the derivations of the main results. Interested readers are referred to a technical report (Tang 2008) for the details. The derivations, which are rather tedious, are double-checked by both authors to ensure their correctness.

Section 1.1 We have

$$\begin{aligned}\hat{\theta}_i - \theta_i &= \hat{\theta}_i - E(\theta_i|y) + E(\theta_i|y) - \theta_i \\ &= (B - \hat{B})(y_i - x_i'\hat{\beta}) + Bx_i'(\hat{\beta} - \beta) + E(\theta_i|y) - \theta_i.\end{aligned}$$

Note that $(B - \hat{B})(y_i - x_i'\hat{\beta})$ is a function of $P_{X^\perp}y$ which is independent of $\tilde{\beta}$ (note that, in the balanced case, $\tilde{\beta}(A) = (X'X)^{-1}X'y$, which does not depend on A). Using this fact it can be shown that $E\{(B - \hat{B})(y_i - x_i'\hat{\beta})Bx_i'(\hat{\beta} - \beta)\} = 0$. Thus, we have

$$\begin{aligned}\text{MSE}(\hat{\theta}_i) &= E\{(B - \hat{B})(y_i - x_i'\hat{\beta}) + Bx_i'(\hat{\beta} - \beta)\}^2 + E\{E(\theta_i|y) - \theta_i\}^2 \\ &= E\{(B - \hat{B})^2(y_i - x_i'\hat{\beta})^2\} + B^2E[\{x_i'(\hat{\beta} - \beta)\}^2] + 1 - B.\end{aligned}$$

Thus, we have $\text{MSE}(\hat{\theta}) = E\{(\hat{B}-B)^2|y-X\hat{\beta}|^2\}+B^2E\{|X(\hat{\beta}-\beta)|^2\}+m(1-B)$. (5) is obtained by using the expression (4) and moment properties of the χ^2 distribution.

Section 2.2 First show that $u'G_1u = O(1)$ and $u'G_ju = O(m)$, $j = 2, 3, 4$. To show these facts we need the following results:

(i) For any matrix G (not necessarily symmetric), $E(u'Gu)^2 \leq 2\text{tr}(G'G)+\{\text{tr}(G)\}^2$.
(ii) For any matrix A, B and C such that $B \geq 0$ (nonnegative definite) we have $|\text{tr}(ABC)| \leq \|B\| \cdot \|A\|_2\|C\|_2$, where for any matrix A, $\|A\| = \lambda_{\max}^{1/2}(A'A)$ and $\|A\|_2 = \text{tr}^{1/2}(A'A)$.
(iii) Let $P^{(k)}(A)$ denote the kth derivative of $P(A)$ with respect to A, then

$$P^{(k)}(A) = k!P(A)[V^{-1}\{XP(A) - I_m\}]^k, \quad k = 0, 1, 2, \ldots$$

To simplify the expression (14) we need the following lemma:

Lemma 1 *Let $u \sim N(0, I)$. Then, for any symmetrix matrix Q and matrix G, we have*

$$\begin{aligned}E[\{u'Qu - tr(Q)\}u'Gu] &= 2tr(QG); \\ E[\{u'Qu - tr(Q)\}^2u'Gu] &= 2tr(Q^2)tr(G) + 8tr(Q^2G); \\ E[\{u'Qu - tr(Q)\}^3u'Gu] &= 8tr(Q^3)tr(G) + 12tr(Q^2)tr(QG) + 48tr(Q^3G); \\ E[\{u'Qu - tr(Q)\}^4u'Gu] &= 12tr^2(Q^2)tr(G) + 48tr(Q^4)tr(G) + 64tr(Q^3)tr(QG) \\ &\quad +96tr(Q^2)tr(Q^2G) + 384tr(Q^4G).\end{aligned}$$

These expressions can be derived from the results of Srivastava and Tiwari (1976).

The following facts are also used to simplify the expression (14): $\mathrm{tr}(G_1) = 0$, $\mathrm{tr}(QG_1) = 0$, $P_{X^\perp} V P(A)' = 0$ and $\{XP(A) - I_m\}VP(A)' = 0$.

Section 2.3 To show that $d(A)$ is bounded we need the following lemma:

Lemma 2 *For any functions $f(x)$, $g(x)$ and $h(x)$, we have*

$$\left\{\sum_i f(\lambda_i)g(\lambda_i)h(\lambda_i)\right\}\left\{\sum_i h(\lambda_i)\right\} - \left\{\sum_i f(\lambda_i)h(\lambda_i)\right\}\left\{\sum_i g(\lambda_i)h(\lambda_i)\right\}$$
$$= \frac{1}{2}\sum_{i\neq j} h(\lambda_i)h(\lambda_j)\{f(\lambda_i) - f(\lambda_j)\}\{g(\lambda_i) - g(\lambda_j)\}.$$

Write $c(A)$ given by (17) as $6I_1 - 2I_2 - 2I_3$, and apply Lemma 2 to $I_1 - 1$, $I_2 - 1$ and $I_3 - 1$. It can be shown that

$$I_1 = 1 - \frac{1}{2}\frac{\sum_{i\neq j} c_{1,i,j}(A)(\lambda_i - \lambda_j)^2}{\{\sum_i (A+\lambda_i)\}\{\sum_i \lambda_i^2 (A+\lambda_i)^{-3}\}},$$
$$I_2 = 1 + \frac{1}{2}\frac{\sum_{i\neq j} c_{2,i,j}(A)(\lambda_i - \lambda_j)^2}{\{\sum_i (A+\lambda_i)\}\{\sum_i \lambda_i^2 (A+\lambda_i)^{-3}\}},$$
$$I_3 = 1 + \frac{1}{2}\frac{\sum_{i\neq j}(\lambda_i - \lambda_j)^2}{\{\sum_i (A+\lambda_i)\}^2},$$

where $c_{r,i,j}(A)$, $r = 1, 2$ are some bounded continuously differentiable functions of A. From these expressions it can be shown that $|c'(A)| \leq c_1 m^{-2}\sum_{i\neq j}(\lambda_i - \lambda_j)^2$, where c_1 is a constant. On the other hand, the denominator of (20) is greater than $2A\sum_{i\neq j}(\lambda_i - \lambda_j)^2$ (recall that $A > 0$). It follows that $|d(A)| \leq c_2$ for some constant c_2.

Expressions for $E(\hat{A}_I - A)^2$ and $E(\hat{A}_I - A)(\hat{A}_I - \hat{A}_D) = E(\hat{A}_I - A)^2 - E(\hat{A}_I - A)(\hat{A}_D - A)$ can be obtained from covariance of quadratic forms of normal random variables (Searle et al. 1992).

Section 3 Derivation of the modified P–R MSE estimator. First we obtain an expansion of $\mathrm{MSE}(\hat{\theta}_i)$ to the second order. It can be shown that $\mathrm{MSE}(\hat{\theta}_i) = AD_i/(A + D_i) + E\{\hat{\theta}_i - E(\theta_i|y)\}^2$, $E\{\hat{\theta}_i - E(\theta_i|y)\}^2 = B_i^2 x_i'(X'V^{-1}X)^{-1}x_i + 2B_i E\{x_i'(\tilde{\beta} - \beta)\theta_i'(A)(\hat{A} - A)\} + E\{\theta_i'(A)(\hat{A} - A)\}^2$, where $\tilde{\beta}$ and $\theta_i(A) = \tilde{\theta}_i$ are given above (2). Furthermore, it can be shown that $E\{x_i'(\tilde{\beta} - \beta)\theta_i'(A)(\hat{A} - A)\} = o(m^{-1})$ and

$$E\{\theta_i'(A)(\hat{A} - A)\}^2 = \frac{2D_i^2 \mathrm{tr}\left\{(P_{X^\perp}V)^2\right\}}{(A + D_i)^3 m(m-p)} + o(m^{-1}).$$

Thus, we have

$$\mathrm{MSE}(\hat{\theta}_i) = \frac{AD_i}{A + D_i} + \frac{D_i^2}{(A + D_i)^2} x_i'(X'V^{-1}X)^{-1}x_i + \frac{2D_i^2 \mathrm{tr}\left\{(P_{X^\perp}V)^2\right\}}{(A+D_i)^3 m(m-p)} + o(m^{-1}).$$

Now consider the following MSE estimator

$$\widehat{\mathrm{MSE}}_{\mathrm{MPR},i} = \frac{\hat{A}D_i}{\hat{A}+D_i} + \frac{D_i^2}{(\hat{A}+D_i)^2}x_i'(X'\hat{V}^{-1}X)^{-1}x_i + \frac{2D_i^2\mathrm{tr}\left\{(P_{X^\perp}\hat{V})^2\right\}}{(\hat{A}+D_i)^3 m(m-p)} + \frac{\gamma_i(\hat{A})}{m},$$

where $\hat{V} = \hat{A}I_m + D$, and $\gamma_i(A)$ is a function of A to be determined. It can be shown that

$$E(\widehat{\mathrm{MSE}}_{\mathrm{MPR},i}) = \frac{AD_i}{A+D_i} + \frac{D_i^2}{(A+D_i)^2}x_i'(X'V^{-1}X)^{-1}x_i + \frac{D_i^2\mathrm{tr}(P_{X^\perp}V)}{(A+D_i)^2 m(m-p)}c(A) + \frac{\gamma_i(A)}{m} + o(m^{-1}).$$

Compare this expression with the expression for $\mathrm{MSE}(\hat{\theta}_i)$ above we obtain an expression for $\gamma_i(A)$.

Section 3 Derivations of (24), (25). For any random variable $\xi > 0$, we have

$$\begin{aligned}
\frac{1}{\xi} &= \frac{1}{E(\xi)} - \frac{\xi - E(\xi)}{E(\xi)} \cdot \frac{1}{\xi} \\
&= \frac{1}{E(\xi)} - \frac{\xi - E(\xi)}{\{E(\xi)\}^2} + \frac{\{\xi - E(\xi)\}^2}{\{E(\xi)\}^2} \cdot \frac{1}{\xi} \\
&= \cdots \\
&= \sum_{k=1}^{l}(-1)^{k-1}\frac{\{\xi - E(\xi)\}^{k-1}}{\{E(\xi)\}^k} + (-1)^l\frac{\{\xi - E(\xi)\}^l}{\{E(\xi)\}^l} \cdot \frac{1}{\xi}, \quad l = 1, 2, \ldots
\end{aligned}$$

This leads to the following lemma:

Lemma 3 *For any random variables U, V such that $U > 0$ (with probability one), we have*

$$E\left(\frac{V}{U}\right) = \sum_{k=1}^{l}(-1)^{k-1}\frac{E[\{U - E(U)\}^{k-1}V]}{\{E(U)\}^k} + \frac{(-1)^l}{\{E(U)\}^l}E\left[\frac{\{U - E(U)\}^l V}{U}\right], \quad l = 1, 2, \ldots$$

The most difficult part of (25) is a fourth-order approximation to $E\{(\hat{\theta}_{i,-j} - \hat{\theta}_i)^2\}$ when $j \neq i$. By using a similar simultaneous diagonalization as the one above (16), it can be shown that

$$\hat{\theta}_{i,-j} - \hat{\theta}_i = \frac{1}{\sqrt{A+1}}\left\{(m-d)\frac{\sum_{k\in S}\rho_{ik}\eta_k}{\sum_{k\in S}\eta_k^2} - (m-1-d)\frac{\sum_{k\in S_{-j}}\rho_{ik}\eta_k}{\sum_{k\in S_{-j}}\eta_k^2}\right\}.$$

By a similar expansion to the one above Lemma 3, we have

$$\frac{1}{\sum_{k\in S}\eta_k^2}=\frac{1}{\sum_{k\in S_{-j}}\eta_k^2}-\frac{\eta_{k_j}^2}{(\sum_{k\in S_{-j}}\eta_k^2)^2}+\frac{\eta_{k_j}^4}{(\sum_{k\in S_{-j}}\eta_k^2)^3}+O(m^{-4}).$$

These lead to $(\hat{\theta}_{i,-j}-\hat{\theta}_i)^2=(A+1)^{-1}\{I_1+(m-d)(I_2-I_3)\}^2+O(m^{-4})$, where

$$\begin{aligned}
I_1&=\frac{\sum_{k\in S_{-j}}\rho_{ik}\eta_k}{\sum_{k\in S_{-j}}\eta_k^2},\\
I_2&=\frac{\rho_{ik_j}\eta_{k_j}}{\sum_{k\in S_{-j}}\eta_k^2}\left\{1-\frac{\eta_{k_j}^2}{\sum_{k\in S_{-j}}\eta_k^2}+\frac{\eta_{k_j}^4}{(\sum_{k\in S_{-j}}\eta_k^2)^2}\right\},\\
I_3&=\frac{(\sum_{k\in S_{-j}}\rho_{ik}\eta_k)\eta_{k_j}^2}{(\sum_{k\in S_{-j}}\eta_k^2)^2}\left(1-\frac{\eta_{k_j}^2}{\sum_{k\in S_{-j}}\eta_k^2}\right).
\end{aligned}$$

It follows that $E\{(\hat{\theta}_{i,-j}-\hat{\theta}_i)^2\}=(A+1)^{-1}[E(I_1^2)+2(m-d)\{E(I_1I_2)-E(I_1I_3)\}+(m-d)^2\{E(I_2^2)-2E(I_2I_3)+E(I_3^2)\}]+O(m^{-4})=(A+1)^{-1}[E(I_1^2)-2(m-d)E(I_1I_3)+(m-d)^2\{E(I_2^2)+E(I_3^2)\}]+O(m^{-4})$, because $E(I_1I_2)=E(I_2I_3)=0$. We then evaluate the remaining terms by Lemma 3 and note the fact that $\rho_{ik_j}^2=E\{x_i'(\hat{\beta}_{-j}-\hat{\beta})^2\}/(A+1)$. After some tedious derivations we obtain the following approximation for the case $j\neq i$:

$$\begin{aligned}
E\{(\hat{\theta}_{i,-j}-\hat{\theta}_i)^2\}=\frac{1}{A+1}\Bigg[&\left\{1+\frac{2(p-d+1)}{m}\right\}x_i'\{(X_{-j}'X_{-j})^{-1}-(X'X)^{-1}\}x_i\\
&+\frac{2}{m^2}\left\{1+\frac{2(2p-d)+5}{m}\right\}\{1-x_i'(X_{-j}'X_{-j})^{-1}x_i\}\Bigg]+o(m^{-3}).
\end{aligned}$$

Section 3 Proof of an inequality. For symmetric matrices Q_j, $j=1,\ldots,m$ we have $\sum_j Q_j^2\geq m^{-1}(\sum_j Q_j)^2$ ($A_1\geq A_2$ means that A_1-A_2 is nonnegative definite). This inequality follows directly from the fact that $2\{m\sum_j Q_j^2-\left(\sum_j Q_j\right)^2\}=\sum_{j,k}(Q_j-Q_k)^2\geq 0$. Now note that $Q_X=(X'X)^{-1/2}\left(\sum_j Q_j^2\right)(X'X)^{-1/2}$ with $Q_j=(X'X)^{-1/2}x_jx_j'(X'X)^{-1/2}$. It then follows that $Q_X\geq m^{-1}(X'X)^{-1/2}(\sum_j Q_j)^2(X'X)^{-1/2}=m^{-1}(X'X)^{-1}$, because $\sum_j Q_j=I_p$.

References

Datta, G. S., Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica, 10*, 613–627.

Datta, G. S., Rao, J. N. K., Smith, D. D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika, 92*, 183–196.

Fay, R. E., Herriot, R. A. (1979). Estimates of income for small places: An application of James–Stein procedures to census data. *Journal of the American Statistical Association, 74*, 269–277.

Ganesh, N. (2009). Simultaneous credible intervals for small area estimation problems. *Journal of Multivariate Analysis, 100*, 1610–1621.

Jiang, J. (2007). *Linear and generalized linear mixed models and their applications* (16 pp.). New York: Springer.

Jiang, J., Lahiri, P. (2006). Mixed model prediction and small area estimation (with discussion). *TEST, 15*, 1–96.

Jiang, J., Lahiri, P., Wan, S. (2002). A unified jackknife theory for empirical best prediction with M-estimation. *Annals of Statistics, 30*, 1782–1810.

Morris, C.N., Christiansen, C.L. (1996). Hierarchical models for ranking and for identifying extremes with applications. In J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith (Eds.), *Bayes statistics* (Vol. 5, pp. 277–297). Oxford: Oxford University Press.

Pfeffermann, D., Nathan, G. (1981). Regression analysis of data from a cluster sample. *Journal of the American Statistical Association, 76*, 681–689.

Prasad, N. G. N., Rao, J. N. K. (1990). The estimation of mean squared errors of small area estimators. *Journal of the American Statistical Association, 85*, 163–171.

Rao, J. N. K. (2003). *Small area estimation*. New York: Wiley.

Searle, S. R. (1982). *Matrix algebra useful for statistics* (312 pp.). New York: Wiley.

Searle, S. R., Casella, G., McCulloch, C. E. (1992). *Variance components* (467 pp.). New York: Wiley.

Srivastava, V. K., Tiwari, R. (1976). Evaluation of expectation of products of stochastic matrices. *Scandinavian Journal of Statistics, 3*, 135–138

Tang, E.-T. (2008). On the estimation of the mean squared error in small area estimation and related topics. *Ph.D. dissertation*. Davis, CA: Department of Statistics, University of California.