A comparison of different MSPE estimators of EBLUP for the Fay-Herriot model

Shijie Chen, Research Triangle Institute
P. Lahiri, University of Maryland at College Park
Shijie Chen, STRD, Research Triangle Institute, RTP, Durham NC, 27709, USA (schen@rti.org)

Key Words: Method of moments, mixed linear models, prediction, variance components.

Abstract:

Over the last two decades, a considerable attention has been given to estimate the mean square prediction error (MSPE) of an estimated best linear predictor (EBLUP) of a mixed effect for a general mixed linear normal model. All such MSPE estimators except the naive estimator are second-order correct and incorporate all sources of variabilities in estimating the true MSPE. In this paper, we compare different MSPE estimators for the simple but very important Fay-Herriot model for small sample through an extensive Monte Carlo simulation experiment. Our study indicates that the recently proposed Chen-Lahiri jackknife MSPE estimators perform very well compared to other rival MSPE estimators in a variety of situations. We prove the second order accuracy of the Chen-Lahiri jackknife MSPE estimator and in the process obtain an useful approximation to this jackknife estimator.

1. Introduction

In order to estimate per-capita income for small areas (population less than 1,000), Fay and Herriot (1979) considered an aggregate level model and used an empirical Bayes method which combines survey data from the U.S. Current Population Survey with various administrative and census records. Their empirical Bayes estimator worked well when compared to the direct survey estimator and a synthetic estimator used earlier by the Census Bureau. The model can be written as:

$$y_i = x_i'\beta + v_i + e_i, \ i = 1, \cdots, m,$$

where v_i's and e_i's are independent with $v_i \overset{iid}{\sim} N(0, A)$ and $e_i \overset{ind}{\sim} N(0, D_i)$, D_i $(i = 1, ..., m)$ being *known*. The Fay-Herriot model is a simple mixed effect model but has been found to be very effective in solving small-area estimation problems.

The research of P. Lahiri was supported in part by a grant from the Center of Excellence in Health Statistics of the University of Michigan, Ann Arbor.

For the Fay-Herriot model, an EBLUP, say $\hat{\theta}_i(y_i; \hat{A})$, of $\theta_i = x_i'\beta + v_i$ is given by:

$$\hat{\theta}_i(y_i; \hat{A}) \ = \ \frac{D_i}{\hat{A} + D_i} x_i'\hat{\beta} \ + \ \frac{\hat{A}}{\hat{A} + D_i} y_i,$$

where $\hat{A}$ is an ANOVA estimator of A and $\hat{\beta} = (\sum_{j=1} \frac{1}{\hat{A}+D_j} x_j x_j')^{-1} \sum_{j=1} \frac{1}{\hat{A}+D_j} x_j y_j$. See Prasad and Rao (1990).

We define $MSPE[\hat{\theta}_i(y_i; \hat{A})] = E[\hat{\theta}_i(y_i; \hat{A}) - \theta_i]^2$, where the expectation is taken over the Fay-Herriot model.

Naive MSPE estimator is given by $mse_i^N = g_{1i}(\hat{A}) + g_{2i}(\hat{A})$, where $g_{1i}(\hat{A}) = \frac{\hat{A}D_i}{\hat{A}+D_i}$, $g_{2i}(\hat{A}) = \frac{D_i^2}{(\hat{A}+D_i)^2} x_i' \left(\sum_{j=1}^m \frac{1}{\hat{A}+D_j} x_j x_j' \right)^{-1} x_i$. The naive MSPE estimator usually underestimates the true MSPE. There are two reasons for this underestimation problem. First, it fails to incorporate the extra variabilities incurred due to the estimation of various model parameters and the order of this underestimation is $O(m^{-1})$, where m is the number of the small-areas. Secondly, the naive MSPE estimator even underestimates the true MSPE of the BLUP, the order of underestimation being $O(m^{-1})$.

Several attempts have been made in the literature to account for these two sources of underestimation and to produce MSPE estimators that are correct up to the order $O(m^{-1})$. These are called second order correct MSPE estimators. See Prasad and Rao (1990), Datta and Lahiri (2000), Butar and Lahiri (2003), Jiang *et al.* (2002), among others, for various approaches.

(i) The Prasad-Rao (PR) MSPE estimator: $mse_i^{PR} = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + 2g_{3i}(\hat{A})$, where $g_{3i} = \frac{2D_i^2}{m^2(\hat{A}+D_i)^3} \sum_{j=1}^m (\hat{A} + D_j)^2$.

(ii) The Jiang-Lahiri-Wan (JLW) estimator:

$$mse_i^{JLW}$$

$$= g_{1i}(\hat{A}) - \frac{m-1}{m}\sum_{u=1}^{m}\Big(g_{1i}(\hat{A}_{-u}) - g_{1i}(\hat{A})\Big)$$
$$+ \frac{m-1}{m}\sum_{u=1}^{m}[\hat{\theta}_i(y_i;\hat{A}_{-u}) - \hat{\theta}_i(y_i;\hat{A})]^2 \quad (1)$$

where

$$\hat{\theta}_i(y_i;\hat{A}_{-u}) = \frac{D_i}{\hat{A}_{-u}+D_i}x_i'\hat{\beta}_{-u} + \frac{\hat{A}_{-u}}{\hat{A}_{-u}+D_i}y_i,$$

$$\hat{\beta}_{-u} = (\sum_{j\neq u}\frac{1}{\hat{A}_{-u}+D_j}x_jx_j')^{-1}\sum_{j\neq u}\frac{1}{\hat{A}_{-u}+D_j}x_jy_j.$$

(iii) The Chen-Lahiri (CL) jackknife:

$$\begin{aligned} mse_i^{CL} &= g_{1i}(\hat{A}) + g_{2i}(\hat{A}) \\ &- \sum_{u=1}^{m} w_u\Big(g_{1i}(\hat{A}_{-u}) + g_{2i}(\hat{A}_{-u}) \\ &\quad -[g_{1i}(\hat{A}) + g_{2i}(\hat{A})]\Big) \\ &+ \sum_{u=1}^{m} w_u[\hat{\theta}_i(y_i;\hat{A}_{-u}) - \hat{\theta}_i(y_i;\hat{A})]^2 \quad (2) \end{aligned}$$

Note that mse_i^{CL} could be negative for small m and thus Chen and Lahiri (2002) recommended using the following MSPE estimator in case the above formula yields a negative value: $mse_i^{CL} = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + D_i^2(\hat{A}+D_i)^{-3}v_{WJ} + \sum_{u=1}^{m} w_u[\hat{\theta}_i(y_i;\hat{A}_{-u}) - \hat{\theta}_i(y_i;\hat{A})]^2$, where $v_{WJ} = \sum_{u=1}^{m} w_u(\hat{A}_{-u} - \hat{A})^2$. We consider two choices of $w_u = \frac{m-1}{m}$ (CL-1) and $w_u = x_u'\left(\sum_{j=1}^{m} x_jx_j'\right)^{-1} x_u$ (CL-2).

(iv) A new approximation to the Chen-Lahiri jackknife (ACL): $mse_i^{ACL} = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + \frac{D_i^2}{(\hat{A}+D_i)^3}v_{WJ} + \frac{D_i^2}{(\hat{A}+D_i)^4}(y_i - x_i'\hat{\beta})^2 v_{WJ}$. This approximation is obtained by approximating $\sum_{u=1}^{m} w_u\left(g_{1i}(\hat{A}_{-u}) + g_{2i}(\hat{A}_{-u}) - [g_{1i}(\hat{A}) + g_{2i}(\hat{A})]\right)$ and $\sum_{u=1}^{m} w_u[\hat{\theta}_i(y_i;\hat{A}_{-u}) - \hat{\theta}_i(y_i;\hat{A})]^2$ by Taylor series expansion.

For small m, JLW could produce negative MSPE estimates. See Bell (2001). Thus, we correct JLW for negative MSPE estimates using a method similar to Chen and Lahiri (2002). Also, for small m, $\hat{A}$ could yield zero estimate. This is problematic for all second order correct methods. In order to achieve good small sample properties of all the second order correct MSPE estimators, we suggest using $g_{2i}(\hat{A})$ for the MSPE estimate whenever $\hat{A} = 0$.

Our jackknife estimators given in (iii) and (iv) above are also second order correct. However, the proof of this second order property does not follow from the one given in Jiang *et al.* (2002) for proving second order property of (ii). To this end, we assume the following regularity conditions:

(r.1) $0 < D_L \leq D_i \leq D_U < \infty, \quad \forall i = 1, \cdots, m$

(r.2) $\sup_{i\geq 1} h_{ii} = O(\frac{1}{m})$.

We need the following technical result to prove Theorem 2.

Theorem 1: Under the regularity conditions (r.1) and (r.2),

(i) $\sum_{u=1}^{m} w_u(\tilde{A}_{-u} - \tilde{A}) = 0$;

(ii) For any random variable η satisfying $E(\eta^2) = O(1)$, we have

$$E\left[\eta\sum_{u=1}^{m} w_u(\hat{A}_{-u} - \hat{A})\right] = O(m^{-s}). \quad \forall s \geq 0.$$

The following theorem establishes the second order properties of our jackknife estimators given in (iii) and (iv).

Theorem 2: Under (r.1) and (r.2), we have

$$\begin{aligned} E(mse_i^{WJ}) &= MSE[\hat{\theta}_i(y;\hat{A})] + o(m^{-1}); \\ E\left[mse_i^{AWJ} - mse_i^{WJ}\right] &= o(m^{-1}). \end{aligned}$$

2. Monte Carlo Simulations

In this section, we investigate the performances of different MSPE estimators given in section 1 for small m through Monte Carlo simulations. We consider $m = 12$, $p = 1$, $\beta = 1$, and $A = 10$. For the first eleven areas, we consider the following combinations of the sampling variance and covariate values: (D, x) : $(10, 1); (9, 1.5); (14, 2); (14, 2.5); (11, 3); (10, 3.5); (10, 4); (13, 5); (4, , 6); (3, 7); (14, 8)$. To study the effect of the covariate on the accuracies of different MSPE estimators, we change $x = x_{12}$ for the last area, keeping $D = D_{12}$ fixed. Similarly, to study the effect of the sampling variance on the accuracies of different MSPE estimators, we change the sampling variance D for the last area keeping $x = x_{12}$ fixed.

For a specific simulation, $R = 10,000$ independent samples of $(v_i,\ e_i)$, $i = 1, \cdots, 12$, are generated from the Fay-Herriot model. We then calculate the relative bias (RB) of each of these MSPE estimators for all the 12 areas using the Monte Carlo method as follows:

$$RB_i = 100\frac{E(mspe_i) - MSPE_i}{MSPE_i},\ i = 1, \cdots, 12,$$

where $mspe_i$ denotes an estimator of $MSPE_i$, the MSPE of EBLUP $\hat{\theta}_i(y_i;\hat{A})$ of the true small-area mean θ_i $(i = 1, \cdots, 12)$, and the expectation "E" is approximated by the Monte Carlo method. We report RB's for the last area and summary statistics (mean and the standard deviation) of the RB's for the rest of 11 areas. In all these tables, the numbers in the second row represent the RB of the MSPE estimators corrected for the situation when $\hat{A} = 0$. We notice that in general the naive estimator underestimates, the severity of the underestimation depends on the values of h and D/A.

We illustrate the effect of x through leverage value defined as $h = h_{12} = \frac{x^2}{\sum_{j=1}^{12} x_j^2}$. Note that h is an increasing function of $x > 0$. We increase h from 0 to 1 by increasing x. Table 1 displays RB's of different MSPE estimators for different x. Table 2 presents the means and the standard deviations of the RB's for the rest of the 11 components. The naive estimator underestimates MSPE in general, the magnitude of the underestimation being severe when h approaches 0. For the last area, the Prasad-Rao estimator performs well for small h but it tends to overestimate when h approaches its maximum value 1. On the other hand, the average performance of the Prasad-Rao estimator for the remaining 11 areas is good when h is small but not so good when h is large. Our jackknife estimators are very robust for different x's - they perform extremely well in protecting against outlying x. The performance of the JLW even after correction for negative estimates is not good, espectially when h is close to 1. For small h it is similar to our jackknife with $w_u = \frac{m-1}{m}$.

Table 4 displays RB's of different MSPE estimators for different D/A for the last small-area. Here again, the naive estimator generally underestimates the true MSPE, the underestimation being severe for large D/A. The RB's for the Prasad-Rao and our jackknife methods exhibit an interesting pattern for varying D/A. The absolute RB's generally increases with increasing $|D/A - 1|$. Both the methods usually overestimate when $D/A < 1$ and underestimate when $D/A > 1$. Small values of D/A cause severe overestimation for the Prasad-Rao method. In comparison, our jackknife MSPE estimators are quite robust. Here again the performance of JLW is not good, especially when D/A for the last area is large.

In Table 5 we display the means and the standard deviations of the RB's for the rest of the 11 components for different values of D/A. In this table also, the robustness of our jackknife methods in comparison with the naive, the Prasad-Rao and JLW methods is clearly demonstrated.

Table 3 and 6 report percentages of times $\hat{A}$ is zero for different values of D/A and h.

References

[1] Bell, W. (2001), Discussion with "Jackknife in the Fay-Herriott Model with An Example," *Proceeding of the Seminar on Funding Opportunity in Survey Research* , 98-104.

[2] Butar, F., and Lahiri, P. (2003), "On measures of uncertainty of empirical Bayes small-area estimators," *Journal of Statistical Planning and Inference,* 112, 63-76.

[3] Chen, S. (2001), "Empirical Best Prediction and Hierarchical Bayes Methods in Small-Area Estimation," Ph.D. Dissertation, Department of Mathematics and Statistics, University of Nebraska-Lincoln.

[4] Chen, S. and Lahiri, P. (2002), "A weighted jackknife MSPE estimator in small-area estimation," *Proceedings of the Survey Research Methods Section of the ASA.*

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

[6] Fay, R. E., and Herriot, R. A. (1979), "Estimates of Income for Small Places: an Application of James-Stein Procedure to Census Data, " *Journal of American Statistical Association* 74, 269-277.

[7] Ghosh, M., and Rao, J.N.K. (1994), "Small Area Estimation: An Appraisal," *Statistical Science* 9, No. 1, 55-93, (with discussion).

[8] Jiang, J., Lahiri, P., and Wan, S.(2002), "A Unified Jackknife Method", *Annals of Statistics* , 30, 1782-1810.

[9] Prasad, N.G.N., and Rao, J.N.K. (1990), "The Estimation of Mean Squared Errors of Small Area Estimators," *Journal of American Statistical Association* 85, 163-171.

[10] Wan, S. (1999), "Jackknife Method in Small-Area Estimation and Related Problems," Ph.D. Dissertation, Department of Mathematics and Statistics, University of Nebraska-Lincoln.

3. APPENDIX

We need the following lemma to prove our main theorems.

Lemma A.1 Under (r.1) and (r.2), we have, for all $s \geq 0$,

$$(a) E|\tilde{A}_{-u} - A|^{2s} = O(m^{-s}); \quad (b) P(\tilde{A}_{-u} \leq 0) = O(m^{-s})$$
$$(c) E|\hat{A}_{-u} - A|^{2s} = O(m^{-s}); \quad (d) E|\hat{A} - \tilde{A}|^{s} = O(m^{-s})$$
$$(e) E|\tilde{A}_{-u} - \tilde{A}|^{s} = O(m^{-s}); \quad (f) E|\hat{A}_{-u} - \hat{A}|^{s} = O(m^{-s})$$

Proof: See Chen (2001).

Proof of Theorem 1: (i) It follows directly from the results $\sum_{i=1}^{m} w_u(\delta_{-u} - \delta) = 0$ (Wan (1999)) and $\sum_{i=1}^{m} w_u(c_{-u} - c) = 0$ (Chen (2001)).

(ii) Because $\hat{A}_{-u} - \hat{A} = \hat{A}_{-u} - \tilde{A}_{-u} + \tilde{A}_{-u} - \tilde{A} + \tilde{A} - \hat{A}$, $\sum w_u = m - p$ and (i), we get

$$\begin{aligned} & \eta \sum w_u(\hat{A}_{-u} - \hat{A}) \\ = & \sum w_u(\hat{A}_{-u} - \tilde{A}_{-u})\eta + (m-p)(\tilde{A} - \hat{A})\eta \\ = & \sum w_u(A_0 - \tilde{A}_{-u})\eta I_{(\tilde{A}_{-u} \leq 0)} \\ & + (m-p)(\tilde{A} - A_0)\eta I_{(\tilde{A} \leq 0)} \end{aligned}$$

The result follows from Lemma A.1 (b), the facts $P(\tilde{A} \leq 0) = O(m^{-s})$ (Lahiri and Rao (1995) Lemma C.1), $E|A_0 - \tilde{A}_{-u}|^t \leq 2^{t-1}[|A_0 - A|^t + E|\tilde{A}_{-u} - A|^t] = O(1)\ \forall t > 0$ by Lemma A.1 (a), and $E|A_0 - \tilde{A}|^2 = O(1)\ \forall t > 0$ by Lemma A.1 (c) and Cauchy-Schwarz inequality.

We now state and prove the following important Lemma used in the proof of Theorem 2.

Lemma A.2. Under regularity conditions (r.1) and (r.2), we have

(i) $E\left[h_{1i}^{WJ} - H_{1i}(A)\right] = -g_{3i}(A) + E\left[\frac{D_i^2}{(\hat{A}+D_i)^3} v_{WJ}(\hat{A})\right] + o(m^{-1})$;

(ii) $E\left[\frac{D_i^2}{(\hat{A}+D_i)^3} v_{WJ}(\hat{A})\right] = g_{3i}(A) + o(m^{-1})$;

(iii) $E\left[h_{2i}^{WJ}\right] = E\left[\frac{D_i^2}{(\hat{A}+D_i)^4}(y_i - x_i'\hat{\beta})^2 v_{WJ}(\hat{A})\right] + o(m^{-1})$;

(iv) $E\left[\frac{D_i^2}{(\hat{A}+D_i)^4}(y_i - x_i'\hat{\beta})^2 v_{WJ}(\hat{A})\right] = g_{3i}(A) + o(m^{-1})$.

Proof of Lemma A.2: To prove (i), let A_u^* and $\tilde{A}_u^*$ lie between $\hat{A}_{-u}$ and $\hat{A}$, $\forall u = 1, \cdots, m$. Then we have

$$\begin{aligned} & \sum_{u=1}^{m} w_u[H_{1i}(\hat{A}_{-u}) - H_{1i}(\hat{A})] \\ = & \sum w_u[g_{1i}(\hat{A}_{-u}) - g_{1i}(\hat{A})] \\ & + \sum w_u[g_{2i}(\hat{A}_{-u}) - g_{2i}(\hat{A})] \\ = & I + II, \text{ say,} \end{aligned}$$

where $I = \sum w_u[g_{1i}(\hat{A}_{-u}) - g_{1i}(\hat{A})]$ and $II = \sum w_u[g_{2i}(\hat{A}_{-u}) - g_{2i}(\hat{A})]$.

$$\begin{aligned} I = & \sum w_u[(\hat{A}_{-u} - \hat{A})g_{1i}'(\hat{A}) + \frac{1}{2}(\hat{A}_{-u} - \hat{A})^2 g_{1i}''(\hat{A}) \\ & + \frac{1}{6}(\hat{A}_{-u} - \hat{A})^3 g_{1i}'''(A_u^*)], \\ II = & \sum w_u[(\hat{A}_{-u} - \hat{A})g_{2i}'(\hat{A}) + \frac{1}{2}(\hat{A}_{-u} - \hat{A})^2 g_{2i}''(\tilde{A}_u^*)]. \end{aligned}$$

where $f'(x), f''(x)$ and $f'''(x)$ are the first, second and third derivative of the function $f(x)$ with respective to x.

Since $|g_{1i}'(\hat{A})| = \frac{D_i^2}{(\hat{A}+D_i)^2} \leq \frac{D_U^2}{D_L^2}$, $E\left[\sum w_u(\hat{A}_{-u} - \hat{A})g_{1i}'(\hat{A})\right] = o(m^{-1})$ by Theorem 3.1 (ii). Noting $g_{1i}'''(A_u^*) = 6D_i^2(A_u^* + D_i)^{-4} \leq 6D_L^{-2}$, $E|\sum w_u((\hat{A}_{-u} - \hat{A})^3 g_{1i}'''(A_u^*)| = o(m^{-1})$ by Lemma A.1 (f). Therefore, $E(I) = E\left[\frac{D_i^2}{(\hat{A}+D_i)^3} v_{WJ}(\hat{A})\right] + o(m^{-1})$.

From direct calculation and using the regularity conditions, we can easily get $|g_{2i}'(\hat{A})| \leq c_1(\sup_i h_{ii})|\hat{A} + D_U|$ and $|g_{2i}''(\tilde{A}_u^*)| \leq c_2(\sup_i h_{ii})|\tilde{A}_u^* + D_U|$, where c_1 and c_2 are independent constants, therefore $E(II) = o(m^{-1})$ by Theorem 1 (ii) and Lemma A.1 (f).

From Prasad and Rao (1990) Theorem $A.3$, we have

$$E[H_{1i}(\hat{A}) - H_{1i}(A)] = -g_{3i}(\hat{A}) + o(m^{-1}) \qquad (3)$$

Therefore,

$$\begin{aligned} & E[h_{1i}^{WJ} - H_{1i}(A)] \\ = & E[H_{1i}(\hat{A}) - H_{1i}(A) - I - II] \\ = & -g_{3i}(A) + E[\frac{D_i^2}{(\hat{A}+D_i)^3} v_{WJ}(\hat{A})] + o(m^{-1}). \end{aligned}$$

To prove (ii), let us look at $E(v_{WJ}(\tilde{A}))$ first. From direct calculation (Chen (2001)), we have

$$\begin{aligned} & m^2 E(v_{WJ}(\tilde{A})) \\ = & m^2 E\left[\sum(\delta_{-u} - \delta)^2\right] - 2E\left[\sum(\delta_{-u} - \delta)(c_{-u} - c)\right] \\ & + \sum(c_{-u} - c)^2 \\ = & 2\sum(A + D_u)^2 + \sum D_i^2 - \frac{1}{m}(\sum(D_u))^2 \\ & + \frac{2}{m}\sum(A + D_u)\sum D_i \end{aligned}$$

$$-2\sum(A+D_i)D_i+\sum D_i^2-\frac{1}{m}(\sum D_u)^2+O(1)$$
$$= 2\sum(A+D_u)^2+O(1).$$

Noting that $E|v_{WJ}(\hat{A}) - v_{WJ}(\tilde{A})| \leq 2\sum w_u\{(E|\hat{A}_{-u}-\hat{A}|^2+E|\tilde{A}_{-u}-\tilde{A}|^2)(E|(\hat{A}_{-u}-\tilde{A}_{-u})-(\hat{A}-\tilde{A})|^2\}^{1/2}=O(m^{-s})\ \forall s>0$ by Lemma A.1 (b), (e), (f) and Cauchy-Schwarz inequality and that $\frac{D_i^2}{(\hat{A}+D_i)^3}\leq\frac{D_U^2}{D_L^3}$, we have $E[\frac{D_i^2}{(\hat{A}+D_i)^3}v_{WJ}(\hat{A})]=E[\frac{D_i^2}{(\hat{A}+D_i)^3}\cdot v_{WJ}(\tilde{A})]+o(m^{-1})$. Also, because $E\left|(\frac{D_i^2}{(\hat{A}+D_i)^3}-\frac{D_i^2}{(A+D_i)^3})(\tilde{A}_{-u}-\tilde{A})^2\right|=O(m^{-2.5})$ by Lemma A.1 (c), (e) and Cauchy-Schwarz inequality, we get

$$\begin{aligned}
& E\left[\frac{D_i^2}{(\hat{A}+D_i)^3}v_{WJ}(\tilde{A})\right] \\
= \ & E\left[\frac{D_i^2}{(A+D_i)^3}v_{WJ}(\tilde{A})\right]+o(m^{-1}). \\
= \ & g_{3i}(A)+o(m^{-1}) \\
& E[h_{1i}^{WJ}-H_{1i}(A)]=o(m^{-1}).
\end{aligned}$$

(ii) is proved.

To Proof (iii) and (iv) we first apply Taylor series expansion of $\hat{\theta}_i(y_i;A)$ about the point $\hat{A}$ to get

$$\begin{aligned}
& \hat{\theta}_i(y_i;\hat{A}_{-u})-\hat{\theta}_i(y_i;\hat{A}) \\
= \ & \hat{\theta}_i'(y_i;\hat{A})(\hat{A}_{-u}-\hat{A})+\frac{1}{2}\hat{\theta}_i''(y_i;A_u^*)(\hat{A}_{-u}-\hat{A})^2.
\end{aligned}$$

where A_u^* is between $\hat{A}_{-u}$ and $\hat{A}$. Following the same algebra in Lahiri and Rao (1995) Theorem C.1, we can get $E|\hat{\theta}_i'(y_i;\hat{A})\hat{\theta}_i''(y_i;A_u^*)(\hat{A}_{-u}-\hat{A})^3|\leq O(m^{-3})$ and $E[|\hat{\theta}_i''(y_i;A_u^*)|^2(\hat{A}_{-u}-\hat{A})^4]\leq O(m^{-4})$. Hence,

$$\begin{aligned}
& E[\hat{\theta}_i(y_i;\hat{A}_{-u})-\hat{\theta}_i(y_i;\hat{A})]^2 \\
= \ & E[\hat{\theta}_i'(y_i;\hat{A})(\hat{A}_{-u}-\hat{A})]^2+O(m^{-3}). \quad (4)
\end{aligned}$$

Noting that $\hat{\theta}_i'(y_i;\hat{A})=\frac{D_i}{(\hat{A}+D_i)^2}(y_i-x_i'\hat{\beta})+\frac{D_i}{(\hat{A}+D_i)^2}\frac{\partial x_i'\tilde{\beta}}{\partial A}|_{A=\hat{A}}$, we have

$$\begin{aligned}
& [\hat{\theta}_i'(y_i;\hat{A})]^2 \\
= \ & \frac{D_i^2}{(\hat{A}+D_i)^4}\left[(y_i-x_i'\hat{\beta})^2+2(y_i-x_i'\hat{\beta})\frac{\partial x_i'\tilde{\beta}}{\partial A}|_{A=\hat{A}}\right. \\
& \left.+(\frac{\partial x_i'\tilde{\beta}}{\partial A}|_{A=\hat{A}})^2\right]. \quad (5)
\end{aligned}$$

Using the same algebra in (ii), we can show

$$\begin{aligned}
& E\left\{\frac{D_i^2}{(\hat{A}+D_i)^4}\left[2(y_i-x_i'\hat{\beta})\frac{\partial x_i'\tilde{\beta}}{\partial A}|_{A=\hat{A}}\right.\right. \\
& \left.\left.+(\frac{\partial x_i'\tilde{\beta}}{\partial A}|_{A=\hat{A}})^2\right](\hat{A}_{-u}-\hat{A})^2\right\} \\
= \ & E\left\{\frac{D_i^2}{(A+D_i)^4}\left[2(y_i-x_i'\tilde{\beta})\frac{\partial x_i'\tilde{\beta}}{\partial A}+(\frac{\partial x_i'\tilde{\beta}}{\partial A})^2\right]\right. \\
& \left.(\hat{A}_{-u}-\hat{A})^2\right\}=o(m^{-2}). \quad (6)
\end{aligned}$$

Now $y_i-x_i'\tilde{\beta}$ and $\frac{\partial x_i'\tilde{\beta}}{\partial A}$ may be written as $y_i-x_i'\tilde{\beta}=\varepsilon_i+l_i'\varepsilon$ and $\frac{\partial x_i'\tilde{\beta}}{\partial A}=s_i'\varepsilon$ with $l_i=(l_{i1},\cdots,l_{im})'$, $\sup_{i,j}|l_{ij}|=O(m^{-1})$, $s_i=(s_{i1},\cdots,s_{im})'$, $\sup_{i,j}|s_{ij}|=O(m^{-1})$, $\varepsilon=(\varepsilon_1,\cdots,\varepsilon_m)$ and $\varepsilon_i=v_i+e_i$. Therefore, from direct calculation, we have

$$\begin{aligned}
& E\left\{\frac{D_i^2}{(A+D_i)^4}\left[2(y_i-x_i'\tilde{\beta})\frac{\partial x_i'\tilde{\beta}}{\partial A}+(\frac{\partial x_i'\tilde{\beta}}{\partial A})^2\right]\right. \\
& \left.(\hat{A}_{-u}-\hat{A})^2\right\}\leq O(m^{-3}), \quad (7)
\end{aligned}$$

by Cauchy-Schwarz inequality and Lemma A.1 (f). Therefore, (iii) follows from (4), (5), (6), (7) and the fact that $\sum w_u=m-p$.

Finally, using the same algebra in (ii), (iv) follows from direct calculation. We refer to Chen (2001)for details.

Table 1: Relative Biases (%) of MSPE Estimators for area=12

h	Naive	PR	JLW	CL-1	CL-2	ACL-1	ACL-2
0	-27.03	13.44	18.34	18.34	19.00	-4.35	-5.45
		1.61	5.90	5.90	6.47	-14.31	-15.30
0.3	-22.42	8.33	25.79	13.30	12.29	-5.64	-6.50
		-0.04	15.73	4.50	3.59	-12.49	-13.27
0.5	-18.66	7.36	41.10	9.84	8.54	-4.88	-5.63
		0.94	31.23	3.24	2.06	-10.01	-10.68
0.8	-11.03	9.35	173.94	4.05	3.16	-1.05	-1.66
		5.60	152.79	0.83	0.04	-3.72	-4.27
0.98	-2.81	14.87	2747.64	-0.87	-0.99	5.06	4.50
		12.90	2470.11	-1.25	-1.35	4.10	3.60

Table 2: Summary statistics of RB's for the first 11 areas

	h	Naive	PR	JLW	CL-1	CL-2	ACL-1	ACL-2
MEAN	0	-24.89	14.91	26.47	18.65	18.62	-3.86	-4.91
			4.10	14.66	7.60	7.55	-12.68	-13.62
	0.3	-25.49	14.82	24.34	18.63	18.85	-4.28	-5.37
			3.90	12.59	7.52	7.74	-13.13	-14.10
	0.5	-26.04	14.69	23.59	18.79	19.19	-4.71	-5.79
			3.76	11.91	7.64	8.01	-13.58	-14.55
	0.8	-27.15	14.17	24.22	19.75	20.31	-5.42	-6.48
			2.99	12.17	8.17	8.69	-14.57	-15.51
	0.98	-28.11	13.42	26.08	20.99	21.56	-6.19	-7.22
			2.16	13.73	9.16	9.68	-15.45	-16.36
STD	0	2.94	6.43	13.39	1.76	2.07	2.47	2.44
			6.93	13.55	1.98	1.84	3.10	3.07
	0.3	2.28	5.43	10.91	1.99	2.10	2.02	2.02
			5.74	10.61	2.21	2.10	2.22	2.22
	0.5	2.06	4.58	9.26	2.52	2.51	2.01	2.03
			4.78	8.96	2.67	2.55	1.75	1.76
	0.8	2.86	2.31	7.92	4.16	4.03	3.06	3.09
			2.35	7.39	3.91	3.75	2.47	2.49
	0.98	4.21	1.07	8.77	6.01	5.80	4.49	4.50
			0.95	7.90	5.37	5.16	4.03	4.04

Table 3: Negative ANOVA Estimates of A (%)

H	0	0.3	0.5	0.7	0.98
$\hat{A} = 0$	10.43	10.38	10.22	10.14	9.99

Table 4: Relative Biases (%) of MSPE Estimators for area=12

D/A	Naive	PR	JLW	CL-1	CL-2	ACL-1	ACL-2
0.1	-7.46	75.81	229.81	2.82	2.40	12.54	11.38
		67.09	203.49	0.98	0.60	9.75	8.70
0.5	-15.76	18.96	70.08	8.39	7.38	-0.56	-1.43
		12.91	58.72	3.45	2.53	-4.60	-5.39
1	-18.41	8.41	43.72	9.73	8.45	-4.46	-5.23
		2.07	33.79	3.32	2.15	-9.45	-10.14
2	-19.56	2.76	29.07	11.02	9.31	-6.52	-7.20
		-4.55	18.73	2.86	1.32	-12.79	-13.40
10	-21.96	4.86	24.92	21.04	13.73	-5.22	-6.03
		-10.58	5.53	2.55	-3.34	-18.76	-19.43

Table 5: Summary statistics of RB's for the first 11 areas

	D/A	Naive	PR	JLW	CL-1	CL-2	ACL-1	ACL-2
MEAN	0.1	-27.44	10.69	23.54	20.87	21.20	-6.74	-7.62
			0.25	11.99	9.60	9.90	-15.47	-16.26
	0.5	-26.33	12.68	23.08	19.14	19.57	-5.41	-6.37
			2.12	11.63	8.12	8.52	-14.13	-14.99
	1	-26.04	14.34	23.46	18.76	19.18	-4.78	-5.84
			3.49	11.84	7.67	8.05	-13.62	-14.57
	2	-26.34	18.36	24.95	19.47	19.68	-4.04	-5.33
			6.32	12.37	7.53	7.73	-13.52	-14.68
	10	-33.83	116.25	37.57	29.80	27.17	0.63	-3.83
			77.16	13.74	7.41	5.30	-16.33	-19.96
STD	0.1	3.80	1.77	8.81	6.82	6.40	4.24	4.25
			1.50	7.89	6.07	5.66	3.71	3.71
	0.5	2.34	2.77	8.49	3.70	3.56	2.53	2.55
			2.86	8.04	3.64	3.46	1.94	1.96
	1	2.06	4.29	9.13	2.61	2.60	2.03	2.05
			4.52	8.83	2.76	2.64	1.72	1.73
	2	2.25	6.34	10.10	2.04	2.13	2.06	2.07
			6.56	9.90	2.21	2.12	2.13	2.12
	10	3.19	46.79	12.91	2.60	2.54	3.37	3.08
			39.75	12.58	1.67	1.71	4.17	3.60

Table 6: Negative ANOVA Estimates of A (%)

D/A	0.1	0.5	1	2	10
$\hat{A} = 0$	96	98.6	10.16	11.01	18.84