A Corrected Confidence Interval for a Small Area Parameter Through the Weighted Estimator under the Basic Area Level Model

Yegnanew Shiferaw and Jacqueline Galpin

Department of Statistics, University of Johannesburg, Johannesburg, South Africa.

Abstract. Area level linear mixed models can be generally applied to produce small area indirect estimators when only aggregated data such as sample means are available. This paper tries to fill an important research gap in small area estimation literature, the problem of constructing confidence intervals (CIs) when the estimated variance of the random effect as well as the estimated mean squared error (MSE) is negative. More precisely, the coverage accuracy of the proposed CI is of the order $O(m^{-3/2})$, where m is the number of sampled areas. The performance of the proposed method is illustrated with respect to coverage probability (CP) and average length (AL) using a simulation experiment. Simulation results demonstrate the superiority of the proposed method over existing naive CIs. In addition, the proposed CI based on the weighted estimator is comparable with the existing corrected CIs based on empirical best linear unbiased predictor (EBLUP) in the literature.

Corresponding Author: Yegnanew Shiferaw(yegnanews@uj.ac.za)
Jacqueline Galpin (jacky@galpin.co.za)

1   Introduction

The demand for small area estimates has been increased tremendously in recent years in order to formulate policies and strategic plans, allocate funds, address social issues and so on (Datta and Ghosh , 2012; Datta and Lahiri , 2000; Rao and Molina , 2015). "Demand from the private sector has also been increased because business decisions, particularly those related to small businesses, rely heavily on the local socio-economic, environmental and other conditions" (Rao and Molina , 2015, page 3).

It is known that direct estimates from the survey have wide confidence intervals (CIs), because the aim of the sample design in sample surveys is to provide reliable data for large areas such as national or regional levels and pays little or no attention to the smaller areas of interest such as sub-regional area (Datta and Ghosh , 2012; Rao and Molina , 2015). It is impossible to conduct a new survey to increase the sample size at the small area level due to cost and operational considerations. In other words, we need sufficient information under these restricted resources in order to make an estimation at an adequate level of precision. This can often be done by using "indirect estimators that "borrow strength from different surveys or from a census and a survey (Chatterjee *et al.* , 2008; Kim and Rao , 2012; Rao and Molina , 2015, among others).

For a linear mixed model, CIs based on empirical best linear unbiased predictor (EBLUP) have been proposed by Datta *et al.* (2002); Basu *et al.* (2003) ; Diao *et al.* (2013); Shiferaw and Galpin (2016); among others, using the Taylor series expansion method. Moreover, Morris (1983), Prasad and Rao (1990), Datta *et al.* (2002) amongst others constructed CIs for a special case of equal sampling variances. However, in small area estimation, we are interested in small areas and hence the sampling error variance estimates ψ_i in the Fay-Herriot (FH) model can be unstable (Boonstra and Buelens , 2011). The recent study by Datta and Ghosh (2012) argued that equal sampling variances seldom arise for small area problems.

Hall and Maiti (2006) proposed interval estimation based on a double bootstrap calibrated sample which is centered at the regression estimate rather than at EBLUP estimate. Their prediction intervals have coverage accuracy $O(m^{-3})$, where m is the number of sampled areas. Chatterjee *et al.* (2008) proposed a CI which is based on a single bootstrap sample for area level means using a normal approximation. Their bootstrap histogram differs from the true EBLUP distribution by only $O(d^3 n^{-3/2})$ where d is the number of parameters and n is the number of observations within the areas. The coverage error is defined in terms of n, sample sizes within the areas, rather than m, the number of sampled areas, as under the Hall and Maiti (2006) approach. This method

necessitates repeated generation of a pivotal quantity from several bootstrap samples. Kubokawa and Nagashima (2012) have also obtained CIs based on the parametric bootstrap methods.

There are some challenges on CIs based on bootstrap methods. For example, the coverage, accuracy of CIs based on bootstrap methods has been improved through calibration. However, it is not always clear either coverage or length or some other characteristic of an interval ought to be calibrated. There are also questions on the use of pivotal statistics and calibration. In addition, the applicability of these CIs is computationally expensive and the results often lack straightforward interpretability (Chatterjee *et al.* , 2008).

None of the above authors developed a CI based on the weighted estimator with fixed weights. Having this in mind, the main objective of this paper is to develop CI for a small area mean (i) based on the weighted estimator with fixed weights using the Taylor series methods under unequal sampling variances (ii) that matches the coverage error properties of Datta *et al.* (2002), Chatterjee *et al.* (2008), Kubokawa and Nagashima (2012) and Yoshimori and Lahiri (2014) (i.e., are of the order $O(m^{-3/2})$), (iii) that has a coverage probability (CPs) larger than the naïve methods and (iv) that rely on simple computation.

This paper deals with the following problems.

i) Firstly, the construction of CIs based on EBLUP has been developed for linear mixed models in small area estimation. However, when we use different methods of estimating variance components, it is possible to get negative values for these components. For example, if the estimated variance of the random component (under the linear mixed model) is negative, Prasad and Rao (1990) and others use the truncated estimator, $max(\hat{A}, 0)$, where $\hat{A}$ is the estimated variance of the random effect. Those truncated estimates will lead the weight attached to the direct estimator to take zero values, in which case the EBLUP estimator reduces to a synthetic estimator. This also makes the contribution of the MSE estimate, assuming all parameters are known, to become zero (Rao and Molina , 2015; Shiferaw and Galpin , 2016).

ii) Secondly, the estimated MSE of EBLUP sometimes takes negative values. A disadvantage of CIs based on EBLUP is that it cannot give an interval when the estimates of MSE are negative (and are set to zero) (Kubokawa and Nagashima , 2012).

In summary, the main contributions of this paper are as follows:

- The CIs based on EBLUP estimator can be used so long as the estimated variance of the random effect remains bounded away from zero (Diao *et al.* , 2013). However, our method can be used even when the estimated variance of the random effect is negative without truncating and setting equal to zero.

- The proposed method allows us to develop a CI when it is impossible to get an interval based on the EBLUP estimator. Unlike the MSE estimates for the EBLUP estimator, the MSE estimates for the weighted estimator with fixed weights are more stable and always positive (Datta *et al.* , 2011). In this paper, we further have used the MSE estimators for the weighted estimator derived by Datta *et al.* (2011) to develop a corrected CI. "Note that the problem of constructing a good prediction interval is much more difficult than the problem of obtaining an accurate MSE" (Nandram , 1999, page 328).

- As discussed briefly in Diao *et al.* (2013) and Shiferaw and Galpin (2016), one of the major advantages of CIs based on the Taylor series methods is computational. They have simple and straightforward applicability, estimate the parameter using standard software and then plug-in the estimates of the parameters in the CIs. This is also an advantage of the proposed CI in this paper.

- The proposed CI based on the weighted estimator with fixed weights might be slightly more efficient than the CIs based on EBLUP estimator with a good choice of weights. This is because, the weighted estimator with fixed weights avoids the uncertainty due to the estimation of the variance of the random effect A (Datta *et al.* , 2011).

This paper is organized as follows. Section 2 describes the basic area level model, Section 3 discusses the estimation of A, Section 4 describes EBLUP estimator. In Section 5, we describe the measurements of the uncertainty of weighted estimator. In Section 6, we present the simulation study results. Finally, in Section 7, we summarize the main findings of the study.

2   Basic Area Level Model

In this paper, we considered a basic area level model, known as the FH model (Fay and Herriot , 1979). Diao *et al.* (2013, page 2) state that "another important advantage of the Fay-Herriot model is that it only requires summary data, and not unit-level data that might be unavailable to the analyst, because of confidentiality concerns". Since

the FH model deals with area level summary data and not unit-level data, the BLUP estimator is applicable for general sampling designs (Rao and Molina , 2015). Let $\mathbf{y} = (y_1, ..., y_m)'$ be a vector of direct survey estimates (observations) for the m small areas and $\mathbf{X} = (x_1, ..., x_m)'$ be a vector of auxiliary variables associated with small areas. Under this setup, the matrix notation of the FH model is given as follows

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} + \mathbf{e}, \tag{2.1}$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)'$ is a $k \times 1$ unknown vector of the regression coefficients, $\mathbf{v} = (v_1, ..., v_m)$ is an $m \times 1$ vector of the random effects, and $\mathbf{e} = (e_1, ..., e_m)$ is an $m \times 1$ vector of the random errors. Here, $\mathbf{v}$ and $\mathbf{e}$ are independent random variables with dispersion matrices $A\mathbf{I}_m$ and $\boldsymbol{\Psi}$, respectively, where $\mathbf{I}_m$ is the $m \times m$ identity matrix. The variance-covariance matrix of $\mathbf{y}$ is $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(A) = \mathrm{diag}(\psi_1 + A, ..., \psi_m + A)$. The aim of this paper is to derive interval estimates for the population small area means $\theta_i = \mathbf{x_i'}\boldsymbol{\beta} + v_i, \ 1, ..., m$.

3 Estimation of A

In this section, we review different methods of estimating the variance component A such as a method of moment estimator by Prasad and Rao (1990), a method of moment estimator by Fay and Herriot (1979), maximum likelihood (ML) and restricted maximum likelihood (REML) estimators. We consider the simple moment estimator given by $\hat{A}_{PR} = max(0, \tilde{A}_{PR})$ where

$$\tilde{A}_{PR} = \frac{1}{m-p}[\mathbf{r}'\mathbf{r} - \sum_{i=1}^{m} \psi_i(1 - h_{ii})], \tag{3.1}$$

which is given by Prasad and Rao (1990), where the matrix $\mathbf{X}$ is full column rank, $\mathbf{r} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ is a vector of residuals, $h_{ii} = \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$ diagonal entries of the hat matrix given by $\mathbf{H} = \mathbf{X}(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}'$ and $\hat{\boldsymbol{\beta}} = (\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}'\mathbf{y}$ is the OLS estimator of $\boldsymbol{\beta}$. Another moment estimator by Fay and Herriot (1979), is given by $\hat{A}_{FH} = max(0, A^*_{FH})$ with A^*_{FH} obtained iteratively as the solution of the following non-linear equation in A

$$\sum_{i=1}^{m} \frac{(y_i - \mathbf{x}'_i\tilde{\boldsymbol{\beta}})^2}{\psi_i + A} = m - p, \tag{3.2}$$

where m is the number of small areas, p is the dimension of the vector of auxiliary variables x_i and

$$\tilde{\beta} = \left\{ \sum_{i=1}^{m} \frac{\mathbf{x_i}\mathbf{x_i'}}{\psi_i + A} \right\}^{-1} \left\{ \sum_{i=1}^{m} \frac{\mathbf{x_i}y_i}{\psi_i + A} \right\}. \tag{3.3}$$

The ML and REML estimators can be derived using the arguments of Datta and Lahiri (2000). The ML estimator of A is given by $\hat{A}_{ML} = \max(0, A_{ML}^*)$, where A_{ML}^* is gained by solving the following non-linear equation in A iteratively

$$\sum_{i=1}^{m} (A + \psi_i)^{-1} = \sum_{i=1}^{m} (A + \psi_i)^{-2} \left\{ y_i - \mathbf{x}_i'\tilde{\beta}(A) \right\}^2. \tag{3.4}$$

The REML estimator of A is given by $\hat{A}_{REML} = \max(0, A_{REML}^*)$, where A_{REML}^* is gained by solving the following non-linear equation in A iteratively

$$\sum_{i=1}^{m} (A + \psi_i)^{-1} - \sum_{i=1}^{m} \frac{\mathbf{x}_i' \left\{ \sum_{u=1}^{m} (A + \psi_u)^{-1} \mathbf{x}_u\mathbf{x}_u' \right\}^{-1} \mathbf{x}_i}{(A + \psi_i)^2} = \sum_{i=1}^{m} (A + \psi_i)^{-2} \left\{ y_i - \mathbf{x}_i'\tilde{\beta}(A) \right\}^2. \tag{3.5}$$

For a given A, the ML or REML estimator of β is given as follows

$$\tilde{\beta}(A) = \left\{ \sum_{i=1}^{m} (A + \psi_i)^{-1}\mathbf{x}_i\mathbf{x}_i' \right\}^{-1} \sum_{i=1}^{m} (A + \psi_i)^{-1}\mathbf{x}_i\mathbf{y}_i. \tag{3.6}$$

4 EBLUP Estimator

If the parameters A and β were known, the small area estimates of $\theta_i = \mathbf{x_i'}\beta + v_i$ would be its best predictor given by

$$\tilde{\theta}_i^B = (1 - \gamma_i)y_i + \gamma_i\mathbf{x_i'}\tilde{\beta}, \qquad i = 1, ..., m. \tag{4.1}$$

where $\tilde{\beta} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{y}$ and $\gamma_i = \frac{\psi_i}{\psi_i + A}$. Since A and β are unknown, the best predictor is not usable until we estimate these model parameters. When the unknown parameters are replaced by their estimators, then we will have EBLUP of θ_i which is given by

$$\hat{\theta}_i^{EB} = (1 - \hat{\gamma}_i)y_i + \hat{\gamma}_i\mathbf{x_i'}\hat{\beta}, \qquad i = 1, ..., m. \tag{4.2}$$

The EBLUP estimator θ_i^{EB} reduces to the synthetic estimator $\mathbf{x_i'}\hat{\beta}$ when $\hat{A} = 0$ irrespective of the sampling variance, ψ_i, of the direct estimator y_i (Rao and Molina , 2015). For

example, this difficulty was encountered in the example, dealing with poverty counts of school-age children in the United States. In this example, the estimated variance of the random effect based on ML and REML was zero. This leads the EBLUP estimator to attach zero weight to all direct estimates. In other words, the EBLUP estimator reduces to the regression of synthetic estimator regardless of the sampling variance of the direct estimator (Rao and Molina , 2015). Furthermore, the contribution to MSE or error assuming all parameters are known becomes zero when $\hat{A} = 0$. As a solution, Rao and Molina (2015) among others, suggested a weighted combination of y_i and $\mathbf{x'_i}\hat{\boldsymbol{\beta}}$ with fixed weights $1 - w_i$ and w_i

$$\hat{\theta}_i^w = (1 - w_i)y_i + w_i\mathbf{x'_i}\hat{\boldsymbol{\beta}}, \qquad i = 1, ..., m, \tag{4.3}$$

where w_i has a value $0 \leq w_i \leq 1$. It can be chosen as $w_i = 1/2$ or can be determined from past knowledge. Both $\hat{\theta}_i^w$ and $\hat{\theta}_i^{\text{EB}}$ are special cases of the general estimator of the form $\hat{\theta}_i = y_i + h_i(y)$ proposed by Rivest and Belmonte (2000).

5 Measurements for Uncertainty of Weighted Estimator

MSEs and interval estimates are generally used to measure the variability of EBLUP and weighted estimators in small area problems. These are commonly used when the estimators are used to estimate a small area mean based on real data.

5.1 MSE Estimation for Weighted Estimator

An asymptotic expression of the $MSE(\hat{\theta}_i^w)$ which is accurate to the order $O(m^{-1})$ is given by (see Datta *et al.* , 2011)

$$MSE(\hat{\theta}_i^w) = g_{1i}(A) + g_{2i}(A) + g_{3wi}(A), \tag{5.1}$$

where,

$$g_{1i}(A) = A\psi_i(A + \psi_i)^{-1},$$

$$g_{2i}(A) = \frac{\psi_i^2}{(A + \psi_i)^2}\mathbf{x'_i}\left\{\sum_{u=1}^{m}(A + \psi_u)^{-1}\mathbf{x_u}\mathbf{x'_u}\right\}^{-1}\mathbf{x_i},$$

$$g_{3wi}(A) = (\gamma_i - w_i)^2\left\{(A + \psi_i) - x'_i\left(X'\boldsymbol{\Sigma}^{-1}X\right)^{-1}x_i\right\} + o(m^{-1}),$$

and $\gamma_i = \frac{\psi_i}{\psi_i + A}$.

For example, a nearly unbiased estimator of $MSE(\hat{\theta}_i^w)$ under the REML estimators of $\hat{A}$ is given by (see Datta *et al.* , 2011)

$$\widehat{MSE}(\hat{\theta}_i^w) = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + g_{3i}(\hat{A}) + g_{3wi}(\hat{A}). \tag{5.2}$$

A nearly unbiased estimator of the $MSE(\hat{\theta}_i^w)$ under the FH moment estimator $\hat{A}$ is given by (see Datta *et al.* , 2011)

$$\widehat{MSE}(\hat{\theta}_i^w) = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + g_{3i}(\hat{A}) + g_{3wi}(\hat{A}) - \hat{b}_{FH}(\hat{A})w_i^2, \tag{5.3}$$

where $\hat{b}_{FH}(\hat{A}) = \frac{2\left[m\sum_{i=1}^{m}(\hat{A}+\psi_i)^{-2} - \{\sum_{i=1}^{m}(\hat{A}+\psi_i)^{-1}\}^2\right]}{\{\sum_{i=1}^{m}(\hat{A}+\psi_i)^{-1}\}^3}.$

A nearly unbiased estimator of the $MSE(\hat{\theta}_i^w)$ under the ML estimator of $\hat{A}$ is given by

$$\widehat{MSE}(\hat{\theta}_i^w) = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + g_{3i}(\hat{A}) + g_{3wi}(\hat{A}) - \hat{b}_{ML}(\hat{A})w_i^2. \tag{5.4}$$

For the ML estimator $\hat{A}$, the estimated bias of $\hat{A}$ derived by Datta and Lahiri (2000) is given by

$$\hat{b}_{ML}(\hat{A}) = \frac{-tr\left\{[\sum_{u=1}^{m}(\hat{A}+\psi_u)^{-1}\mathbf{x}_u\mathbf{x}_u']^{-1}[\sum_{u=1}^{m}(\hat{A}+\psi_u)^{-2}\mathbf{x}_u\mathbf{x}_u']\right\}}{\sum_{u=1}^{m}(\hat{A}+\psi_u)^{-2}}.$$

Similarly, a nearly unbiased estimator of the $MSE(\hat{\theta}_i^w)$ under the PR moment estimator of $\hat{A}$ is given by

$$\widehat{MSE}(\hat{\theta}_i^w) = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + g_{3i}(\hat{A}) + g_{3wi}(\hat{A}). \tag{5.5}$$

5.2   CIs Based on Weighted Estimator with Corrected Coverage Probability

We here provide CI for $\theta_i = \mathbf{x}_i\boldsymbol{\beta} + v_i$ based on the weighted estimator with fixed weights using the FH model. Cox (1975) proposed the conventional closed-form CI is of the form: $\hat{\theta}_i^w \pm z_{\alpha/2}\psi_i^{1/2}(1-\hat{\gamma}_i)^{1/2}$, where $z_{\alpha/2}$ denotes the $100 \times \alpha/2\%$ upper quantile of the standard normal distribution. This is the first attempt to construct empirical Bayes (EB) CI in small area estimation problems. Following Cox (1975), Prasad and Rao (1990) proposed another traditional closed-form CI based on EBLUP by incorporating covariates when the error variances are equal. The two-sided version of the Prasad and

Rao (1990) CI is given by

$$\hat{\theta}_i^w \pm z_{\alpha/2} \sqrt{\widehat{MSE}(\hat{\theta}_i^w)},$$

where $\widehat{MSE}(\hat{\theta}_i^w)$ is a second-order unbiased estimator of $MSE(\hat{\theta}_i^w)$. Both methods are easy to construct, but their CPs are open to question. These methods have coverage accuracy of order $O(m^{-1})$ (Diao *et al.* , 2013; Chatterjee *et al.* , 2008). Thus, they cannot be an appropriate choice in most small area applications (Diao *et al.* , 2013; Yoshimori and Lahiri , 2014).

As we discussed earlier, Diao *et al.* (2013), Shiferaw and Galpin (2016), Chatterjee *et al.* (2008), Yoshimori and Lahiri (2014) among others proposed CIs based on EBLUP with a higher coverage accuracy. In the case that $\hat{A} = 0$, the EBLUP $\hat{\theta}_i^{EB}$ of θ_i reduces to the regression synthetic estimator $\hat{\theta}_{synthetic, i}^{EB} = \mathbf{x_i'}\hat{\boldsymbol{\beta}}$. This results in $g_{1i}(0) = 0$ (Molina *et al.* , 2015; Rao and Molina , 2015). Moreover, we cannot calculate CI when the estimated MSE of EBLUP is negative. To fill these gaps, we propose CI with a coverage, accuracy $O(m^{-3/2})$ based on the weighted estimator with fixed weights using the Taylor series methods. Our method provides improved coverage in the sense that the margin of error of the CP is of order $O(m^{-3/2})$.

We use the following regularity conditions in proving Theorem 5.1.

R1: v_i and e_i are independent with $v_i \overset{i.i.d.}{\sim} N[0, A]$ and $e_i \overset{ind}{\sim} N[0, \psi_i]$, $\qquad i = 1, ..., m$.

R2: An estimator of A that is $\hat{A} = \hat{A}(\mathbf{y})$ satisfies

- $\hat{A}(-\mathbf{y}) = \hat{A}(\mathbf{y})$ (i.e., they are even functions of $\mathbf{y}$)
- $\hat{A}(\mathbf{y} + \mathbf{X}d) = \hat{A}(\mathbf{y})$ for any $d \in R^p$ and for all $\mathbf{y}$ (i.e., they are translation invariant functions).

R3: $\max_{1 \leq i \leq m} \mathbf{x_i'}(\mathbf{X'X})^{-1}\mathbf{x}_i = O(m^{-1})$ (i.e., the elements of X are uniformly bounded) (see Prasad and Rao , 1990; Yoshimori and Lahiri , 2014).

We use the Stein identity to derive the proposed CI. Stein (1981) showed that

$$E\left[h_i(\mathbf{y})(y_i - \mathbf{x_i'}\boldsymbol{\beta})\right] = (\psi_i + A)E\left[\frac{\partial h_i(\mathbf{y})}{\partial y_i}\right], \tag{5.6}$$

where $h_i(\mathbf{y})$ is an absolutely continuous function. This identity is called the Stein identity. The proposed CI based on the weighted estimator with fixed weights is given by the following theorem.

Theorem 5.1. *Under the above regularity conditions and for any z > 0, we have*

$$P[\theta_i \in I^{CW}(\hat{A})] = 2\Phi(z) - 1 - z\phi(z)\eta_w + O(m^{-3/2}),$$

where

$$\eta_w = (z^2 + 1)\frac{\psi_i^2}{4A^2(A + \psi_i)^2}V(\hat{A}) + \left\{\frac{\psi_i^2 + (\gamma_i - w_i)^2(A + \psi_i)^2}{A\psi_i(A + \psi_i)}x_i'\left(\sum_{i=1}^{m}\frac{x_i'x_i}{A + \psi_i}\right)^{-1}x_i\right\}.$$

Note that the proof of this Theorem is given at the end of the paper.

6   Numerical Studies

6.1   Comparison of the CIs

We conducted a simulation study to examine the finite sample performance of the proposed CI. We adopt the simulation setup of Datta *et al.* (2005). For the simulation study, we consider model (1.1) with m = 10, 15, 30, 45 and 60 areas, $\mathbf{x}_i = (1, \mathbf{x}_i)'$, where $\mathbf{x}_i; i = 1, ..., m$ were generated i.i.d. from $N(0, 1)$, $\boldsymbol{\beta} = (0, 0)'$ and $A = 1$. For the sampling variances ψ_i, we considered three different variance patterns: pattern (I) 0.7, 0.6, 0.5, 0.4, 0.3, pattern (II) 2, 0.6, 0.5, 0.4, 0.2 and pattern (III) 4, 0.6, 0.5, 0.4, 0.1. According to Rao and Molina (2015) pattern (I) is nearly balanced (max $(\psi_i)/\min(\psi_i) = 2.3$), pattern (II) has intermediate variability (max $(\psi_i)/\min(\psi_i) = 10$) and pattern (III) has the highest variability (max $(\psi_i)/\min(\psi_i) = 40$). The random effects, v_i are generated from two different distributions, namely the normal $N(0, 1)$ and the Laplace $(0, 1)$ distributions. The Laplace distribution is considered to assess the robustness of the proposed method to possible deviations from the normality assumptions. We consider $x_i\boldsymbol{\beta} = 0$, and generate 10,000 data sets from $y_i = v_i + e_i$, $(i = 1, \cdots, m)$. We computed the CP and AL of the proposed CI over simulation as

$$CP_i = \sum_{i=1}^{m}\frac{|\theta_i \in CI_i|}{m}, \qquad AL_i = \sum_{i=1}^{m}\frac{|length\ of\ CI_i|}{m}, \qquad (6.1)$$

where , i=1,...,m, CI_i is the CI of a certain method. The purpose of this section is to compare the performance of the proposed method with the two 'naive' CIs namely the Cox (1975) and Prasad and Rao (1990) methods under different sampling variances.

Cox (1975) introduced the idea of developing an EB CI for θ_i given as

$$Met\ IW : \hat{\theta}_i^w \pm z_{\alpha/2}\psi_i^{1/2}(1 - \hat{\gamma}_i)^{1/2}, \tag{6.2}$$

where $\hat{\theta}_i^w$ is a weighted estimator with fixed weights.

Prasad and Rao (1990) proposed a CI for θ_i of the form

$$Met\ IIW : \hat{\theta}_i^w \pm z_{\alpha/2}\sqrt{\widehat{MSE}(\hat{\theta}_i^w)}. \tag{6.3}$$

The corrected CI based on the weighted estimator is given as follows

$$Met\ IIIW : \hat{\theta}_i^w \pm t_{\alpha w}\sqrt{\widehat{MSE}(\hat{\theta}_i^w)}, \tag{6.4}$$

where

$$t_{\alpha w} = z_{\alpha/2} + (z_{\alpha/2}^3 + z_{\alpha/2})\frac{\psi_i^2}{8\hat{A}^2(\hat{A} + \psi_i)^2}V(\hat{A}) + z_{\alpha/2}\left\{\frac{\psi_i^2 + (\hat{\gamma}_i - w_i)^2(\hat{A} + \psi_i)^2}{2\hat{A}\psi_i(\hat{A} + \psi_i)}x_i'\left(\sum\frac{x_i'x_i}{\hat{A} + \psi_i}\right)^{-1}x_i\right\}.$$

For $t_{\alpha w}$, the CI given in (6.4) satisfies that

$$P\left[\theta_i \in \{\hat{\theta}_i^w \pm t_{\alpha w}\sqrt{\widehat{MSE}(\hat{\theta}_i^w)}\}\right] = 1 - \alpha + O(m^{-3/2}). \tag{6.5}$$

In this section, we consider three different CIs given in (6.2), (6.3) and (6.4), which are referred as Method IW, Method IIW, and Method IIIW, respectively. In each table and figure, Method IW, Method IIW, and Method IIIW are the Cox (1975), Prasad and Rao (1990) and our proposed method respectively. We have carried out extensive simulations which reflect the general pattern of performance of Method IW, Method IIW and Method IIIW reported here. Based on 10000 simulations, we estimated the CPs and ALs of the three methods for all selected small areas and variance patterns. The results are reported in Tables 1 and 2. These results are dependent on the sampling variance patterns (Chatterjee *et al.* , 2008).

The main goal of this study is to develop closed form corrected CI for a small area mean based on weighted estimator with fixed weights using the Taylor series expansion under the FH model. We explored the CPs and ALs of the proposed CIs compared to Method IW and Method IIW.

For variance patterns (II) and (III), Method IW truly underestimates the CPs. For example, as shown in Table 1, Method IW has under-coverage problem (it could be as low as 69.2% compared to the nominal value of 95% for the normal random effects). Method IIIW performed uniformly well throughout. When $m = 20$, the CP of Method IIIW never differed from its nominal value by more than 1.3%, 9.2%, 2.2% and 1.3% for the ML, PR, FH and REML estimators respectively for the normal distribution. These are values slightly below the 95% nominal value.

Considering the Laplace random effects distribution helps us to investigate the robustness of our proposed methods when the data violate the normality assumptions. As we can see from Table 2, the CPs of Method IIIW are improved when compared to Method IW and Method IIW. The CPs of Method IW becomes worse when dealing with the Laplace distribution.

We also examine the performance of our proposed method over a range of a number of small areas, m, for all variance patterns as shown in Figures 2, 3 and 4. The CPs of Method IW, Method IIW, and Method IIIW are obtained through the above simulation exercise for the number of areas such as m=10, 15, 20, 30, 40, 50, 60, 80 and 100. When the number of small areas, m, increases, the CPs of Method IIIW are close to the nominal level for all patterns. When $m \geq 60$, there is no clear difference between Method IIW and Method IIIW in terms of CPs and ALs under the normal distribution, even though Method IIIW continues to be closer to the nominal level, hence it shows its consistency (Diao *et al.* , 2013). We thus conclude that Method IIIW is superior to the rest of the methods in terms of the CP.

6.2   Comparison of the Proposed Method with some of the Recent Corrected CIs

This section discusses the comparison of our method (i.e., Meth IIIW) with some of the recent corrected CIs in the literature, namely, Shiferaw and Galpin (2016)'s method (based on the Taylor series expansion), Diao *et al.* (2013)'s method (based on the Taylor series expansion) and Chatterjee *et al.* (2008)'s method (based on bootstrapping). For the sake of comparability, we adopt the simulation setup of Chatterjee *et al.* (2008) given by Pattern (IV) (4.0, 0.6, 0.5, 0.4, 0.2). The FH model with $m = 15$ and $x_i'\beta = 0$ is considered.

The simulation study is carried out under the FH model without covariates. This is in agreement with Diao *et al.* (2013) and Chatterjee *et al.* (2008). Table 3 and Table 4 show the CPs and ALs for nominal 95% CIs for Pattern (IV) under the normal and

the chi-square distributions. Note that Method IIIW represents our proposed method, while Met III, CIRao, CIJY, CIJY1 represents the area specific CIs proposed by Shiferaw and Galpin (2016), PB-ET represents the parametric bootstrap interval with equal tail probabilities proposed by Chatterjee *et al.* (2008) and Diao-EB represents the corrected CI proposed by Diao *et al.* (2013).

As shown in Tables 3 and 4, Diao-EB, Chat-PB and Shiferaw-Galpin methods are slightly more conservative than the proposed CI. However, our proposed method (Method IIIW) has slightly lower CPs, especially for smaller ψ_i. Datta *et al.* (2011) justified that the weighted estimator with $w_i = \frac{1}{2}$ does not perform very well in comparison with the EBLUP estimator, especially for areas with smaller ψ_i. In general, the CPs and ALs of the proposed methods are comparable to those recent CI developments in the literature.

6.3   Application

Here we apply the proposed CI to the 2010/11 Ethiopian household consumption expenditure survey (HCES) data. These data were used by Shiferaw and Galpin (2016) to estimate the zonal level percentage of food expenditure for their proposed area specific CIs based on EBLUP under the FH model. As auxiliary variables, we have considered the indicators of age, the indicators of sex, the indicators of the different levels of the variable education, the indicators of the categories of the variable employment and indicators of marital status. The main model fitted in this paper is the FH model given by (2.1).

The process of choosing the auxiliary variables is crucial since the small area estimation methods are mainly based on the development of a regression model. Good auxiliary information related to the variables of interest plays a vital role in determining suitable linking models (Rao and Molina , 2015). Akaike information criterion was used to select the auxiliary variables that can specify the best regression model. The area level covariates are the area proportions of individuals in each category of the auxiliary variables (since in our case all auxiliary variables are categorical). For example, if we consider sex as one of the auxiliary variables, clearly sex has two categories (i.e., male and female). In order to include the proportion of males or the proportion of females among other covariates in the fitted regression model, we considered the smallest AIC. According to Rao and Molina (2015), the usual goodness-of-fit measures such as AIC, BIC and the value of the log-likelihood at the estimated parameters are computed.

Moreover, we fitted the country level regression model instead of regional or zonal

level models to benefit from the larger sample size at the country level. For instance, the use of detailed age groups may exacerbate the limitation for modeling from the small sample sizes (Pratesi , 2015). We fitted the basic area level model (2.1) using the FH method to derive zonal level weighted estimators of the percentage of food expenditure. Thus, the estimates of the parameters are given by $\hat{A}_{FH} = 28.79$ and $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6\right) = (14.375, -5.8656, 8.788, -69.198, 29.589, -0.384)$. Rao and Molina (2015) suggest $w_i = \frac{A_0}{\psi_i + A_0}$, where A_0 is a prior guess of A from past studies (Rao and Molina , 2015). The widths of the naive CIs Method IW, Method IIW and the corrected CI Method IIIW based on the FH estimators are computed by (6.2), (6.3) and (6.4) are plotted in Figure 1 for $i = 1, \cdots, 86$. The upper and lower bounds of the CIs are also plotted in Figure 2 for $i = 1, \cdots, 86$. From Figures 1 and 2, we can easily observe that Method IIIW is more stable and shorter width than Method IW and Method IIW.

7 Conclusions

In this paper, we have obtained the asymptotically corrected CIs based on weighted estimator with fixed weights under the basic area level model. CIs based on weighted estimator with fixed weights have the advantage over CIs based on EBLUP estimator due to the possibility of producing CIs when the variance of the random effect A as well as the estimated MSE are negative. Moreover, the simulation results show that the corrected CI (i.e., Method IIIW) has been numerically superior to the naive CIs (i.e., Method IW and Method IIW) when evaluated based on the coverage probabilities. Method IIIW satisfies the nominal confidence level by extending the widths of Method IW and Method IIW. Finally, the methodology is illustrated with an application to Ethiopian HCES data. The direct survey estimates of the percentage of food expenditure based on the small sample can be considerably improved by using the auxiliary variables from the census.
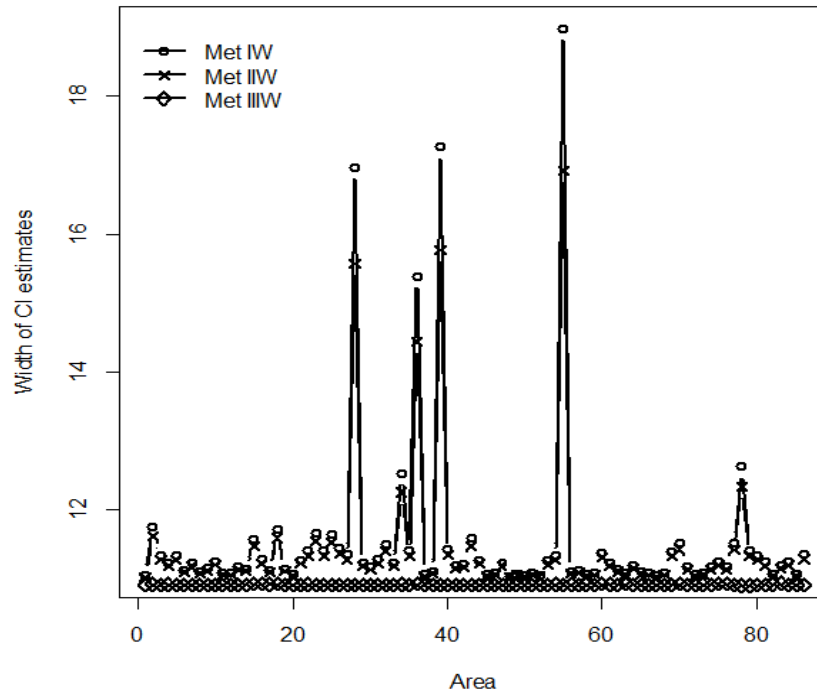
Figure 1: Plots of widths for the CIs Method IW, Method IIW and Method IIIW for $i = 1, \ldots, 86$.
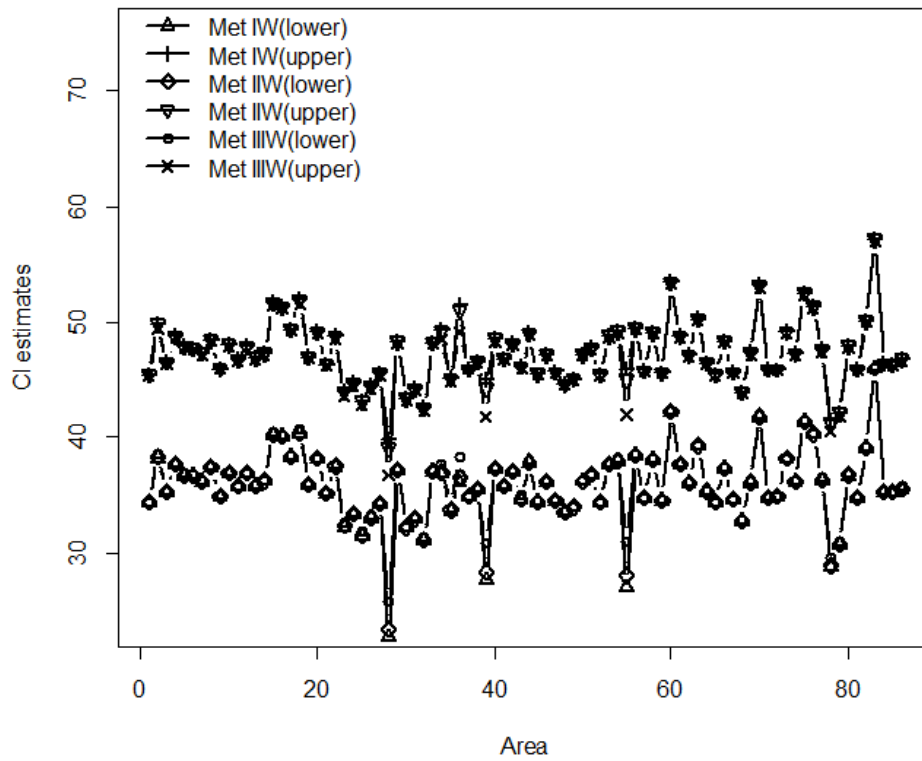
Figure 2: Plots of CIs for the mean based on Method IW, Method IIW and Method IIIW for $i = 1, \ldots, 86$.
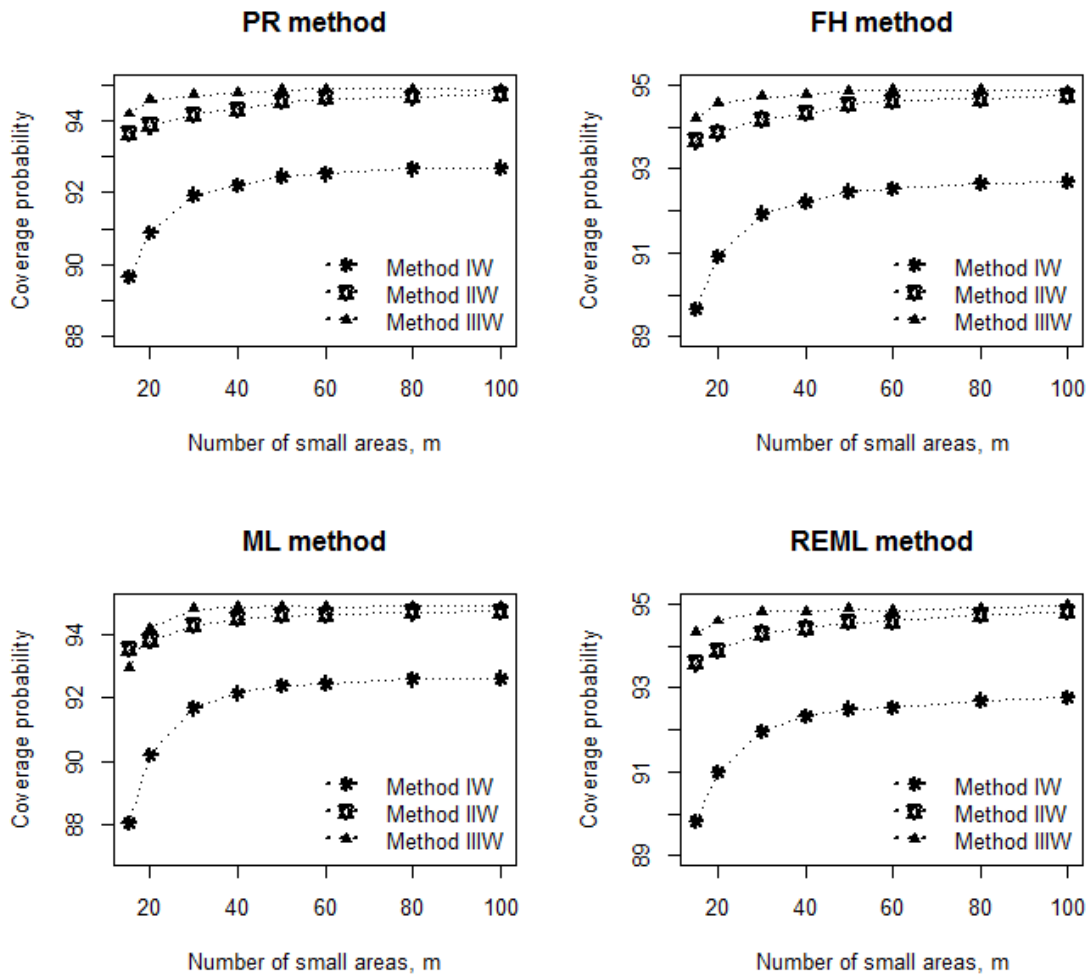
Figure 3: Simulated values of the CPs of Method IW, Method IIW and Method IIIW for nominal 95% CIs for ψ_i pattern (I). The points are connected with lines for visibility purposes.
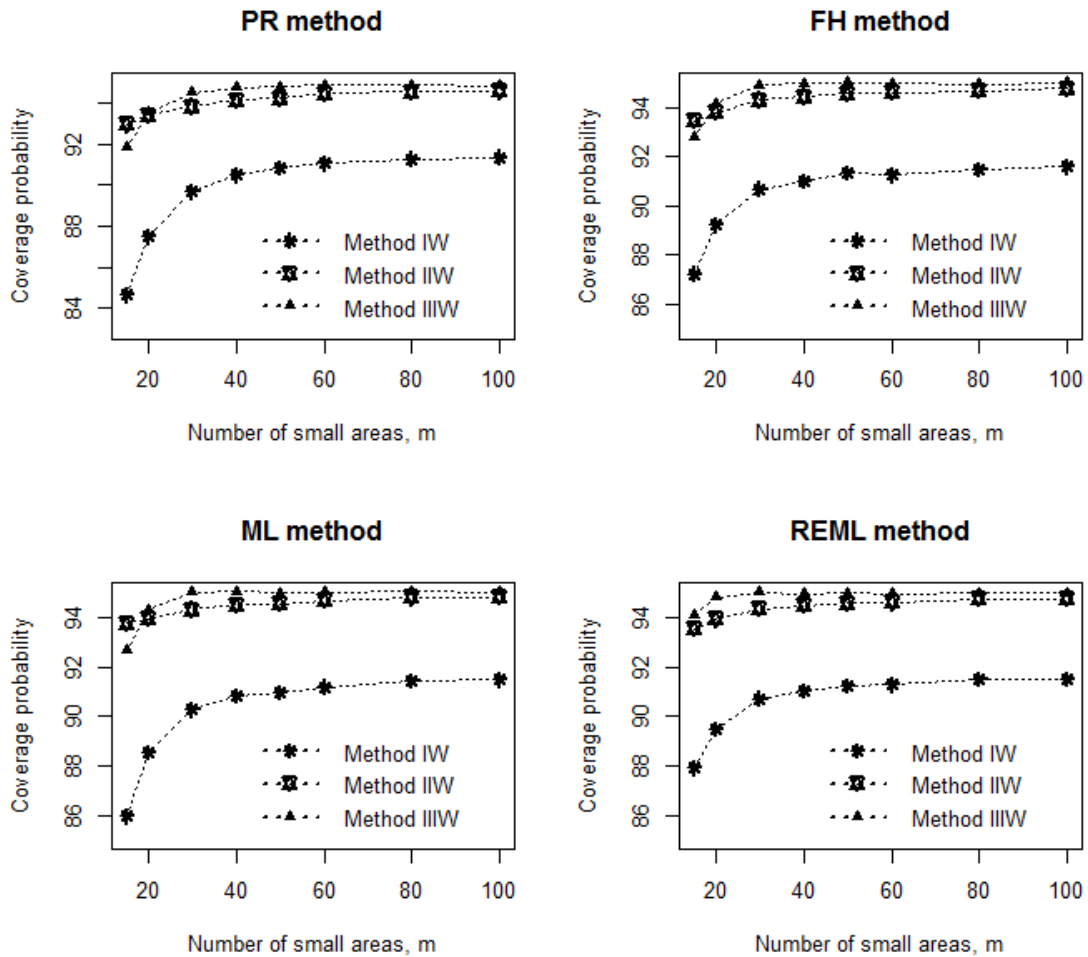
Figure 4: Simulated values of the CPs of Method IW, Method IIW and Method IIIW for nominal 95% CIs for ψ_i pattern (II). The points are connected with lines for visibility purposes.

Figure 5: Simulated values of the CPs of Method IW, Method IIW and Method IIIW for nominal 95% CIs for ψ_i pattern (III). The points are connected with lines for visibility purposes.

Table 1: CPs and ALs of a small area mean for nominal 95% CIs for ψ_i pattern (I), (II) and (III). Results obtained for the normal random effects distributions, $m = 20$ and $w_i = \frac{1}{2}$.

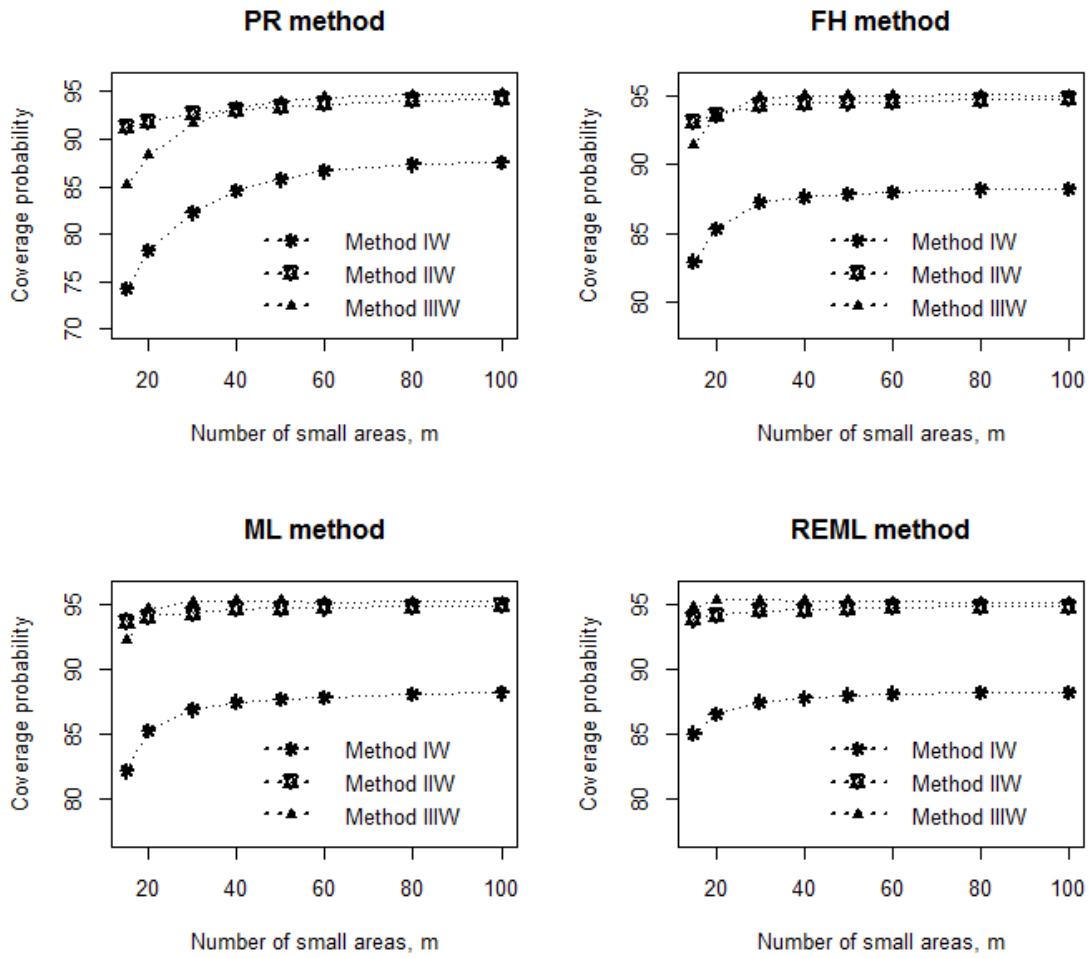| Groups	CP											
	PR			FH			ML			REML		
	MetIW	MetIIW	MetIIIW	MetIW	MetIIW	MetIIIW	MetIW	MetIIW	MetIIIW	MetIW	MetIIW	MetIIIW
						Pattern I						
G1	92.9	94.3	95.0	92.9	94.2	94.9	92.5	94.3	95.0	93.0	94.2	95.0
G2	92.5	94.0	94.7	92.7	94.3	94.8	92.2	94.2	94.8	92.7	94.2	94.9
G3	92.2	94.3	94.8	92.2	94.2	94.7	91.9	94.2	94.8	92.3	94.2	94.8
G4	91.2	94.0	94.6	91.1	94.0	94.5	91.0	94.1	94.7	91.3	94.0	94.7
G5	89.0	93.9	94.4	89.3	94.1	94.5	88.9	94.0	94.5	89.2	93.9	94.5
						Pattern II						
G1	89.0	94.6	96.0	90.9	94.9	96.1	89.5	94.7	95.8	90.7	94.6	96.1
G2	91.3	94.0	94.5	92.4	94.1	94.6	91.6	93.9	94.6	92.7	94.3	94.9
G3	90.8	93.7	94.1	91.8	94.0	94.4	91.3	94.1	94.6	92.2	94.2	94.8
G4	89.8	93.5	93.9	91.0	94.0	94.3	90.6	94.2	94.6	91.2	94.0	94.6
G5	83.9	93.1	93.4	85.0	93.8	94.3	84.8	94.1	94.5	85.0	93.9	94.5
						Pattern III						
G1	78.3	95.1	93.8	84.8	95.0	96.2	83.6	94.9	96.4	85.1	94.9	96.7
G2	85.9	92.6	90.7	92.0	94.0	94.0	91.8	94.0	94.6	92.5	94.0	94.8
G3	85.2	92.1	90.3	91.5	93.8	93.8	91.3	93.9	94.4	92.0	94.0	94.6
G4	84.8	91.9	89.9	90.3	93.6	93.5	90.8	94.2	94.6	91.2	94.3	94.8
G5	69.2	89.4	87.9	73.2	93.4	94.1	73.7	94.2	95.2	74.4	94.4	95.7
						AL						
						Pattern I						
G1	2.44	2.57	2.42	2.44	2.57	3.15	2.41	2.56	3.12	2.44	2.57	3.00
G2	2.33	2.48	2.32	2.33	2.48	3.06	2.30	2.48	3.04	2.33	2.48	2.92
G3	2.20	2.40	2.21	2.20	2.40	2.98	2.18	2.39	2.96	2.20	2.39	2.83
G4	2.04	2.30	2.10	2.04	2.30	2.91	2.02	2.30	2.88	2.04	2.30	2.75
G5	1.84	2.21	1.98	1.84	2.21	2.84	1.83	2.20	2.81	1.84	2.20	2.67
						Pattern II						
G1	3.03	3.44	4.04	3.10	3.45	3.66	3.01	3.44	3.55	3.09	3.44	3.74
G2	2.28	2.48	2.88	2.33	2.49	2.52	2.28	2.48	2.40	2.33	2.48	2.61
G3	2.16	2.39	2.79	2.20	2.40	2.42	2.16	2.39	2.30	2.20	2.39	2.52
G4	2.00	2.29	2.70	2.04	2.31	2.32	2.01	2.30	2.19	2.04	2.30	2.42
G5	1.54	2.09	2.55	1.57	2.10	2.12	1.55	2.09	1.97	1.57	2.10	2.24
						Pattern III						
G1	3.15	4.43	5.96	3.39	4.44	4.95	3.29	4.43	5.16	3.40	4.44	4.98
G2	2.14	2.45	3.08	2.32	2.48	2.60	2.29	2.48	2.85	2.33	2.49	2.67
G3	2.02	2.36	2.99	2.19	2.39	2.50	2.17	2.39	2.76	2.20	2.40	2.57
G4	1.88	2.26	2.90	2.03	2.30	2.40	2.01	2.30	2.68	2.04	2.31	2.48
G5	1.09	1.92	2.80	1.16	1.97	2.17	1.16	1.98	2.53	1.17	1.98	2.27

Table 2: CPs and ALs of a small area mean for nominal 95% CIs for ψ_i pattern (I), (II) and (III). Results obtained for the Laplace (double exponential) random effects distributions, $m = 20$ and $w_i = \frac{1}{2}$.

Groups	CP											
	PR			FH			ML			REML		
	MetIW	MetIIW	MetIIIW	MetIW	MetIIW	MetIIIW	MetIW	MetIIW	MetIIIW	MetIW	MetIIW	MetIIIW
Pattern I												
G1	87.0	92.1	94.2	86.7	90.5	94.2	84.8	89.8	94.3	87.7	92.8	94.4
G2	87.4	92.0	94.1	87.1	90.4	94.2	85.4	89.8	94.3	87.8	92.6	94.0
G3	88.0	92.0	94.1	87.5	90.1	94.1	86.0	89.7	94.3	88.1	92.3	94.0
G4	87.8	91.5	93.6	87.3	90.0	93.9	85.6	89.2	93.6	88.6	92.3	93.8
G5	87.8	91.4	93.5	87.1	89.6	93.5	85.7	89.2	93.6	88.3	91.9	93.6
Pattern II												
G1	73.8	87.2	95.0	77.4	87.7	94.9	76.8	90.2	94.9	79.8	93.8	94.9
G2	80.5	86.0	93.6	84.0	86.7	93.8	84.7	89.4	94.1	87.8	92.6	94.1
G3	80.9	85.8	93.6	84.4	86.6	93.8	85.0	89.1	93.9	88.0	92.3	93.7
G4	81.0	85.5	93.1	84.3	86.3	93.4	85.1	88.8	93.5	88.4	92.2	93.7
G5	79.9	84.7	91.9	82.9	85.8	92.6	84.1	88.5	93.3	87.1	91.6	93.1
Pattern III												
G1	56.7	76.8	95.5	65.5	82.1	95.0	66.7	89.9	95.0	70.5	94.6	94.8
G2	70.0	75.4	93.1	79.1	81.0	93.7	84.1	88.7	94.0	88.4	93.4	94.2
G3	70.2	75.3	92.6	79.1	80.6	93.2	84.5	88.6	93.7	88.9	93.3	94.0
G4	70.3	74.8	92.1	79.1	80.5	93.0	84.7	88.3	93.4	88.8	92.8	93.6
G5	65.9	73.3	88.4	73.2	79.8	90.4	79.4	87.8	92.7	83.9	92.1	92.8
AL												
Pattern I												
G1	1.92	2.17	2.34	1.93	2.11	2.18	1.85	2.17	2.30	1.93	2.05	2.18
G2	1.86	2.08	2.24	1.86	2.00	2.08	1.79	2.08	2.19	1.87	1.93	2.08
G3	1.78	1.98	2.13	1.78	1.87	1.98	1.72	1.97	2.08	1.79	1.80	1.98
G4	1.68	1.87	2.01	1.68	1.74	1.87	1.62	1.86	1.96	1.69	1.66	1.87
G5	1.55	1.75	1.89	1.55	1.60	1.75	1.50	1.75	1.84	1.56	1.51	1.75
Pattern II												
G1	2.14	3.14	3.61	2.22	3.14	3.24	2.16	3.14	3.32	2.27	3.15	3.35
G2	1.73	2.07	2.39	1.81	1.97	2.07	1.78	2.08	2.11	1.86	2.08	2.16
G3	1.65	1.97	2.28	1.73	1.84	1.96	1.71	1.98	1.99	1.78	1.98	2.04
G4	1.56	1.86	2.16	1.63	1.70	1.85	1.61	1.87	1.87	1.68	1.87	1.92
G5	1.25	1.60	1.93	1.31	1.38	1.60	1.31	1.59	1.61	1.36	1.61	1.66
Pattern III												
G1	2.14	4.22	5.04	2.27	4.19	4.80	2.26	4.19	4.50	2.41	4.20	4.43
G2	1.56	2.09	2.13	1.72	2.05	2.45	1.77	2.07	2.14	1.87	2.02	2.07
G3	1.48	1.98	1.98	1.65	1.94	2.34	1.69	1.96	2.03	1.79	1.90	1.97
G4	1.39	1.82	1.87	1.55	1.83	2.24	1.60	1.85	1.91	1.69	1.77	1.86
G5	0.85	1.24	1.44	0.95	1.40	1.95	1.01	1.44	1.50	1.06	1.31	1.45

Table 3: CPs and ALs of a small area mean for nominal 95% CIs for ψ_i pattern (IV). Results obtained for the normal random effects distributions and $m = 15$. Note that the first value in the cell is the proposed method and the value in the parenthesis is the naive method.

ψ_i		Shiferaw and Galpin method (Shiferaw-Galpin)								Method from this paper		Diao *et al.* method		Chatterjee *et al.* method
		Met III		CIRao		CIJY		CIJY1		Met IIIW		Diao-EB		PB-ET
		FH	REML	FH	REML	FH	REML	FH	REML	FH	REML	FH	REML	FH
							Normal							
4.0	CP	93.7	93.4	93.8	93.7	93.6	93.6	93.9	94.1	96.1	97.1	96.0	95.3	95.7
		(91.2)	(91.7)	(90.9)	(92.3)	(91.1)	(91.9)	(91.4)	(92.6)	(95.0)	(94.2)	(91.9)	(90.7)	
	AL	4.00	3.91	4.02	4.01	3.95	3.99	4.01	4.04	4.80	5.16	4.26	4.10	4.73
		(3.57)	(3.54)	(3.59)	(3.64)	(3.53)	(3.62)	(3.58)	(3.66)	(4.45)	(4.45)	(3.79)	(3.62)	
0.6	CP	94.4	93.7	94.2	95.1	93.8	95.2	94.5	95.2	94.3	94.6	96.4	96.1	95.6
		(93.9)	(93.2)	(93.7)	(94.8)	(93.3)	(94.8)	(94.0)	(94.8)	(93.6)	(93.8)	(94.5)	(93.7)	
	AL	2.54	2.47	2.54	2.61	2.51	2.58	2.54	2.60	2.48	2.71	2.73	2.69	2.95
		(2.49)	(2.42)	(2.49)	(2.56)	(2.45)	(2.53)	(2.49)	(2.55)	(2.33)	(2.47)	(2.57)	(2.52)	
0.5	CP	94.4	93.9	94.5	95.4	94.3	95.0	94.4	95.4	94.0	94.5	96.3	96.1	95.5
		(93.9)	(93.6)	(94.0)	(95.0)	(93.9)	(94.7)	(94.0)	(95.1)	(93.3)	(93.8)	(94.8)	(94.2)	
	AL	2.39	2.32	2.39	2.45	2.36	2.42	2.39	2.44	2.39	2.61	2.56	2.53	2.70
		(2.34)	(2.28)	(2.35)	(2.41)	(2.32)	(2.39)	(2.34)	(2.41)	(2.21)	(2.38)	(2.42)	(2.37)	
0.4	CP	94.6	94.1	94.4	95.5	94.3	95.0	94.5	95.4	93.8	94.4	96.4	96.2	95.6
		(94.3)	(93.8)	(94.1)	(95.3)	(94.0)	(94.8)	(94.2)	(95.1)	(93.6)	(93.8)	(95.1)	(94.6)	
	AL	2.20	2.14	2.20	2.26	2.18	2.23	2.20	2.25	2.29	2.51	2.36	2.34	2.49
		(2.17)	(2.11)	(2.18)	(2.23)	(2.15)	(2.21)	(2.17)	(2.22)	(2.09)	(2.29)	(2.24)	(2.20)	
0.2	CP	95.2	94.7	95.0	95.5	95.1	95.5	95.1	95.6	93.8	94.6	95.9	95.8	95.9
		(95.1)	(94.6)	(94.9)	(95.5)	(95.0)	(95.4)	(95.0)	(95.6)	(93.3)	(93.9)	(95.4)	(95.1)	
	AL	1.67	1.62	1.68	1.70	1.66	1.69	1.67	1.69	2.07	2.33	1.81	1.79	1.83
		(1.66)	(1.61)	(1.67)	(1.70	(1.66)	(1.68)	(1.66)	(1.69)	(1.84)	(2.07)	(1.70)	(1.68)	

Table 4: CPs and ALs of a small area mean for nominal 95% CIs for ψ_i pattern (IV). Results obtained for the chi-square random effects distributions and $m = 15$. Note that the first value in the cell is the proposed method and the value in the parenthesis is the naive method.

ψ_i		Shiferaw and Galpin method (Shiferaw-Galpin)								Method from this paper		Diao *et al.* method		Chatterjee *et al.* method
		Met III		CIRao		CIJY		CIJY1		Met IIIW		Diao-EB		PB-ET
		FH	REML	FH	REML	FH	REML	FH	REML	FH	REML	FH	REML	FH
						Centered chi-square								
4	CP	93.2	93.5	93.4	93.6	93.3	93.5	93.0	93.7	95.6	97.2	91.6	92.3	91.7
		(91.6)	(92.1)	(91.8)	(92.4)	(91.6)	(92.0)	(91.5)	(92.3)	(94.7)	(95.1)	(85.3)	(87.7)	
	AL	5.17	4.24	5.21	4.27	5.14	4.21	5.20	4.24	4.82	4.89	3.60	3.51	4.35
		(4.24)	(3.98)	(4.26)	(4.00)	(4.21)	(4.00)	(4.26)	(3.99)	(4.61)	(4.45)	(2.99)	(2.87)	
0.6	CP	94.5	94.6	95.1	95.5	94.6	95.1	94.4	94.7	93.9	94.8	95.4	94.5	92.3
		(94.0)	(94.1)	(94.8)	(95.1)	(94.1)	(94.7)	(93.9)	(94.3)	(93.1)	(94.0)	(91.6)	(90.9)	
	AL	2.65	2.59	2.68	2.62	2.64	2.58	2.65	2.59	3.03	2.48	2.94	2.75	2.71
		(2.59)	(2.43)	(2.62)	(2.43)	(2.58)	(2.44)	(2.59)	(2.43)	(2.22)	(2.47)	(2.18)	(2.10)	
0.5	CP	94.7	94.8	95.4	95.5	94.9	95.1	94.6	94.8	93.7	94.5	95.8	95.1	92.4
		(94.4)	(94.4)	(95.1)	(95.2)	(94.6)	(94.7)	(94.3)	(94.4)	(92.9)	(93.8)	(92.1)	(91.5)	
	AL	2.47	2.42	2.50	2.45	2.46	2.42	2.47	2.42	2.95	2.39	2.93	2.72	2.45
		(2.43)	(2.28)	(2.45)	(2.28)	(2.42)	(2.28)	(2.43)	(2.28)	(2.11)	(2.36)	(2.10)	(2.02)	
0.4	CP	94.8	94.8	95.6	96.0	95.0	95.3	94.7	95.1	93.7	94.7	96.3	96.0	93.3
		(94.5)	(94.5)	(95.3)	(95.8)	(94.7)	(95.0)	(94.4)	(94.9)	(92.9)	(94.1)	(93.1)	(92.4)	
	AL	2.26	2.22	2.28	2.25	2.26	2.23	2.26	2.22	2.87	2.29	2.94	2.69	2.34
		(2.23)	(2.09)	(2.25)	(2.10)	(2.23)	(2.10)	(2.23)	(2.10)	(1.99)	(2.25)	(2.01)	(1.94)	
0.2	CP	95.2	95.3	96.1	96.0	95.6	95.6	95.2	95.4	93.6	94.4	97.5	97.6	94.4
		(95.1)	(95.3)	(96.0)	(95.9)	(95.5)	(95.5)	(95.1)	(95.3)	(93.0)	(93.7)	(96.3)	(96.6)	
	AL	1.69	1.67	1.72	1.70	1.71	1.68	1.69	1.67	2.69	2.08	3.36	2.91	1.73
		(1.68)	(1.58)	(1.71)	(1.58)	(1.70)	(1.58)	(1.68)	(1.58)	(1.73)	(2.03)	(1.78)	(1.71)	

References

Basu, R., Ghosh, J. K. and Mukerjee, R. (2003), Empirical Bayes prediction intervals in a normal regression model: higher order asymptotics. *Statistics and Probability Letters*, **63**, 197–203.

Boonstra, J. H. and Buelens, B. (2011), *Model based estimation*. Statistics Netherlands, The Hague/Heerlen.

Chatterjee, S., Lahiri, P. and Li, H. (2008), Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models. *The Annals of Statistics*, **36**, 1221–1245.

Cox, D. R. (1975), Prediction intervals and empirical Bayes confidence intervals. In *Perspectives in probability and statistics*, Papers in honor of M.S. Bartlett (ed. J. Gani), 47-55, Academic Press, London.

Datta, G. S. and Ghosh, M. (2012), Small area shrinkage estimation. *Statistical Science*, **27**, 95–114.

Datta, G. S., Ghosh, M., Smith, D.D. and Lahiri, P. (2002), On an asymptotic theory of conditional and unconditional coverage probabilities of Empirical Bayes confidence intervals. *Scandinavian Journal of Statistics*, **29**, 139–152.

Datta, G. S., Kubokawa, T., Molina, I. and Rao, J. N. K. (2011), Estimation of mean squared error of model-based small area estimators. *Test*, **20**, 367–388.

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.

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.

Diao, L., Smith, D. D., Datta, G. S., Maiti, T. and Opsomer, J. D. (2013), Accurate confidence interval estimation of small area parameters under the Fay-Herriot model. *Scandinavian Journal of Statistics*, **41**, 497–515.

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 the American Statistical Association*, **74**, 269–277.

Hall, P. and Maiti, T. (2006). On parametric bootstrap methods for small area prediction. *Journal of the Royal Statistical Society B*, **68**, Part 2, 221–238.

Kim, J. K. and Rao, J. N. K. (2012), Combining data from two independent surveys: A model assisted approach. *Biometrika*, **99**, 85–100.

Kubokawa, T. and Nagashima, B. (2012), Parametric bootstrap methods for bias correction in linear mixed models. *Journal of Multivariate Analysis*, **106**, 1–16.

Molina, I., Rao, J. N. K. and Datta, G. S. (2015), Small area estimation under a Fay-Herriot model with premilinary testing for the presence of random area effects. *Survey Methodology*, **41**, 1–19.

Morris, C. N. (1983), *Parametric Empirical Bayes Confidence Intervals, in Scientific Inference, Data Analysis and Robustness*. Academic Press, New York, 25–50.

Nandram, B. (1999). An Empirical Bayes prediction interval for the finite population mean of small area. *Statistica Sinica*, **9**, 325-343.

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

Pratesi, M. (2015), *Analysis of Poverty Data by Small Area Estimation*. John Wiley & Sons, Incorporated.

Rao, J. N. K. and Molina, I. (2015), *Small Area Estimation*. John Wiley and Sons, Inc., New York.

Rivest, L. P. and Belmonte, E. (2000), A Conditional Mean Squared Error of Small Area Estimators. *Survey Methodology*, **26**, 79–90.

Shiferaw, Y. A. and Galpin, J. S. (2016), Area specific confidence intervals for a small area mean under the Fay-Herriot model. *Journal of the Iranian Statistical Society*, **15**, 1–44.

Stein, C. (1981), Estimation of the mean of a multivariate normal distribution, *Annals of Statistics*, **9**, 1135–1151.

Yoshimori, M. and Lahiri, P. (2014), A second-order efficient empirical Bayes confidence interval. *The Annals of Statistics*, **42**, 1233–1261.

Appendix

The Proof of Theorem 5.1

Proof. This Theorem is proved similarly to Datta et al. (2002), Diao et al. (2013) and Shiferaw and Jacquline (2016). The conditional distribution of θ_i given y_i is given by

$$\theta_i \mid y_i \sim N(\hat{\theta}_i^w, g_{1i}(A)),$$

where $\hat{\theta}_i^w = (1 - w_i)y_i + w_i\mathbf{x}_i'\hat{\boldsymbol{\beta}}$, $\tilde{\theta}_i^B = (1 - \gamma_i)y_i + \gamma_i\mathbf{x}_i'\tilde{\boldsymbol{\beta}}$, $g_{1i}(A) = \frac{A\psi_i}{A+\psi_i}$. Consider first that

$$P\left[\theta_i \leq \hat{\theta}_i^w + z \times \sqrt{\widehat{MSE}(\hat{\theta}_i^w)}\right]$$

$$= P\left[\frac{\theta_i - \hat{\theta}_i^B}{\sqrt{E(\theta_i - \hat{\theta}_i^B)^2}} \leq \frac{\hat{\theta}_i^w - \hat{\theta}_i^B + z \times \sqrt{\widehat{MSE}(\hat{\theta}_i^w)}}{\sqrt{E(\theta_i - \hat{\theta}_i^B)^2}}\right]$$

$$= P\left[\frac{\theta_i - \hat{\theta}_i^B}{\sqrt{g_{1i}(A)}} \leq \frac{\hat{\theta}_i^w - \hat{\theta}_i^B + z \times \sqrt{\widehat{MSE}(\hat{\theta}_i^w)}}{\sqrt{g_{1i}(A)}}\right]$$

$$= P\left[\frac{\theta_i - \hat{\theta}_i^B}{\sqrt{g_{1i}(A)}} \leq \frac{z \times \sqrt{g_{1i}(A)}}{\sqrt{g_{1i}(A)}} + \frac{z \times \sqrt{\widehat{MSE}(\hat{\theta}_i^w)} - z \times \sqrt{g_{1i}(A)} + \hat{\theta}_i^w - \hat{\theta}_i^B}{\sqrt{g_{1i}(A)}}\right]$$

$$= P\left[\frac{\theta_i - \hat{\theta}_i^B}{\sqrt{g_{1i}(A)}} \leq \frac{z \times \sqrt{g_{1i}(A)}}{\sqrt{g_{1i}(A)}} + \frac{z \times \left(\sqrt{\widehat{MSE}(\hat{\theta}_i^w)} - \sqrt{g_{1i}(A)}\right) + \hat{\theta}_i^w - \hat{\theta}_i^B}{\sqrt{g_{1i}(A)}}\right]$$

$$= P\left[\frac{\theta_i - \hat{\theta}_i^B}{\sqrt{g_{1i}(A)}} \leq z + \frac{z \times \left(\sqrt{\widehat{MSE}(\hat{\theta}_i^w)} - \sqrt{g_{1i}(A)}\right) + \hat{\theta}_i^w - \hat{\theta}_i^B}{\sqrt{g_{1i}(A)}}\right]$$

$$= P\left[\frac{\theta_i - \hat{\theta}_i^B}{\sqrt{g_{1i}(A)}} \leq z + F(z)\right] = E[\Phi(z + F(z))],$$

where $F(z) = \frac{z \times \left(\sqrt{\widehat{MSE}(\hat{\theta}_i^w)} - \sqrt{g_{1i}(A)}\right) + \hat{\theta}_i^w - \tilde{\theta}_i^B}{\sqrt{g_{1i}(A)}}$.

Thus, the CP of $I^{CW}(\hat{A})$ is written as

$$P[\theta_i \in I^{CW}(\hat{A})] = P[-z + F(-z) < \frac{\theta_i - \hat{\theta}_i^w}{\sqrt{g_{1i}(A)}} < z + F(z)]$$

$$= E[\Phi(z + F(z)) - \Phi(-z + F(-z))].$$

Using the Taylor series expansion with an integral remainder term, $\Phi(z + F(z))$ is evaluated as

$$\Phi(z + F(z)) = \Phi(z) + F(z)\phi(z) + \frac{1}{2}F^2(z)\phi'(z) + \frac{1}{2}\int_z^{z+F(z)} (z + F(z) - x)^2 \phi''(x)dx$$

$$= \Phi(z) + [F(z) - \frac{z}{2}F^2(z)]\phi(z) + \frac{1}{2}\int_z^{z+F(z)} (z + F(z) - x)^2(x^2 - 1)\phi(x)dx.$$

Taking expectations on both sides reveals that

$$E(\Phi(z + F(z))) = \Phi(z) + E[F(z)\phi(z) - \frac{z}{2}F^2(z)\phi(z)$$

$$+ \frac{1}{2}\int_z^{z+F(z)} (z + F(z) - x)^2(x^2 - 1)\phi(x)dx].$$

But

$$\hat{\theta}_i^w - \tilde{\theta}_i^B = (1 - w_i)y_i + w_i\mathbf{x}_i'\hat{\beta} - \left((1 - \gamma_i)y_i + \gamma_i\mathbf{x}_i'\tilde{\beta}\right)$$

$$= y_i - w_i(y_i - \mathbf{x}_i'\hat{\beta}) - y_i + \gamma_i(y_i - \mathbf{x}_i'\tilde{\beta}(A))$$

$$= \gamma_i(y_i - \mathbf{x}_i'\tilde{\beta}(A)) - w_i(y_i - \mathbf{x}_i'\hat{\beta}).$$

Using the Taylor series expansion we have

$$\hat{\beta} = \tilde{\beta}(A) + \tilde{\beta}^{(1)}(A)(\hat{A} - A),$$

and

$$\hat{\theta}_i^w - \tilde{\theta}_i^B = \gamma_i(y_i - \mathbf{x}_i'\tilde{\beta}(A)) - w_i\left(y_i - \mathbf{x}_i'(\tilde{\beta}(A) + \tilde{\beta}^{(1)}(A)(\hat{A} - A))\right)$$

$$= \gamma_i(y_i - \mathbf{x}_i'\tilde{\beta}(A)) - w_i(y_i - \mathbf{x}_i'\tilde{\beta}(A)) - w_i\mathbf{x}_i'\tilde{\beta}^{(1)}(A)(\hat{A} - A))$$

$$= (\gamma_i - w_i)(y_i - \mathbf{x}_i'\tilde{\beta}(A)) - w_i\mathbf{x}_i'\tilde{\beta}^{(1)}(A)(\hat{A} - A)).$$

Substituting reveals that

$$
F(z) = \frac{z \times \left(\sqrt{\widehat{\mathrm{MSE}}(\hat{\theta}_i^w)} - \sqrt{g_{1i}(A)} \right) + \hat{\theta}_i^w - \tilde{\theta}_i^B}{\sqrt{g_{1i}(A)}}
$$

$$
= \frac{z \times \left(\sqrt{\widehat{\mathrm{MSE}}(\hat{\theta}_i^w)} - \sqrt{g_{1i}(A)} \right)}{\sqrt{g_{1i}(A)}} + \frac{(\gamma_i - w_i)(y_i - \mathbf{x}_i'\tilde{\boldsymbol{\beta}}(A))}{\sqrt{g_{1i}(A)}} - \frac{w_i \mathbf{x}_i'\tilde{\boldsymbol{\beta}}^{(1)}(A)(\hat{A} - A))}{\sqrt{g_{1i}(A)}}
$$

$$
= S_1 + R_2 + R_3.
$$

Using the Taylor series expansion it can be shown that

$$
ES_1 = z\frac{1}{2}\left(g_{1i}(A)\right)^{-1} E\left(\widehat{\mathrm{MSE}}(\hat{\theta}_i^w) - g_{1i}(A)\right)
$$

$$
- z\frac{1}{8}\left(g_{1i}(A)\right)^{-2} E\left(\widehat{\mathrm{MSE}}(\hat{\theta}_i^w) - g_{1i}(A)\right)^2
$$

$$
+ z\frac{3}{16}\left(g_{1i}(A)\right)^{-1/2} E\int_{g_{1i}(A)}^{\widehat{\mathrm{MSE}}(\hat{\theta}_i^w)} x^{-5/2}\left(\widehat{\mathrm{MSE}}(\hat{\theta}_i^w) - g_{1i}(A)\right)^2 dx.
$$

Let us compute the expectations of each term in turn

$$
E\left(\widehat{\mathrm{MSE}}(\hat{\theta}_i^w) - g_{1i}(A)\right) = E(\widehat{\mathrm{MSE}}(\hat{\theta}_i^w)) - E(g_{1i}(A)).
$$

$$
E(\widehat{\mathrm{MSE}}(\hat{\theta}_i^w)) = \mathrm{MSE}(\hat{\theta}_i^w),
$$

since $E(\widehat{\mathrm{MSE}}(\hat{\theta}_i^w))$ is nearly a second order unbiased estimator of $\mathrm{MSE}(\hat{\theta}_i^w)$.

Using the Taylor series expansion, the second term can be expressed as

$$
\widehat{\mathrm{MSE}}(\hat{\theta}_i^w) = \mathrm{MSE}(\hat{\theta}_i^w(A)) + \mathrm{MSE}(\hat{\theta}_i^w(A^*))^{(1)}(\hat{A} - A) + O(m^{-3/2}).
$$

This can be rewritten as

$$
\widehat{\mathrm{MSE}}(\hat{\theta}_i^w) - g_{1i}(A) = g_{1i}^{(1)}(A^*)(\hat{A} - A).
$$

Squaring and taking expectation on both sides:

$$
E\left(\widehat{\mathrm{MSE}}(\hat{\theta}_i^w) - g_{1i}(A)\right)^2 = (g_{1i}^{(1)}(A^*))^2 E(\hat{A} - A)^2,
$$

where $g_{1i}(A) = \frac{A\psi_i}{A+\psi_i}$, $g_{1i}^{(1)}(A) = \frac{\psi_i}{A+\psi_i} - \frac{A\psi_i}{(A+\psi_i)^2} = \frac{\psi_i^2}{(A+\psi_i)^2}$, $(g_{1i}^{(1)}(A))^2 = \frac{\psi_i^4}{(A+\psi_i)^4}$.

Substituting reveals that

$$E\left(\widehat{\text{MSE}}(\hat{\theta}_i^w) - g_{1i}(A)\right)^2 = \frac{\psi_i^4}{(A+\psi_i)^4}E(\hat{A} - A)^2 + O(m^{-3/2}),$$

where, $E(\hat{A} - A)^2 = V(\hat{A}) + O_p(m^{-1})$, $V(\hat{A})$ can be the asymptotic variance of the FH, PR, ML and REML estimators of A.

Let us compute the integral term using the same approach by Datta et al. (2002); Chatterjee and Lahiri (2002), Diao et al. (2013) and Shiferaw and Jacquline (2016) as follows.

$$E \int_{g_{1i}(A)}^{\widehat{\text{MSE}}(\hat{\theta}_i^w)} x^{-5/2} \left(\widehat{\text{MSE}}(\hat{\theta}_i^w) - g_{1i}(A)\right)^2 dx = EI_1 + EI_2.$$

Also,

$$EI_1 = \left[E \int_{g_{1i}(A)}^{\widehat{\text{MSE}}(\hat{\theta}_i^w)} x^{-5/2} \left(\widehat{\text{MSE}}(\hat{\theta}_i^w) - g_{1i}(A)\right)^2 dx\right] I_{\{\widehat{\text{MSE}}(\hat{\theta}_i^w) \geq g_{1i}(A)\}},$$

and

$$EI_2 = \left[E \int_{g_{1i}(A)}^{\widehat{\text{MSE}}(\hat{\theta}_i^w)} x^{-5/2} \left(\widehat{\text{MSE}}(\hat{\theta}_i^w) - g_{1i}(A)\right)^2 dx\right] I_{\{\widehat{\text{MSE}}(\hat{\theta}_i^w) \leq g_{1i}(A)\}}.$$

But

$$
\begin{aligned}
EI_1 &= \left[E \int_{g_{1i}(A)}^{\widehat{\text{MSE}}(\hat{\theta}_i^w)} x^{-5/2} \left(\widehat{\text{MSE}}(\hat{\theta}_i^w) - g_{1i}(A)\right)^2 dx\right] I_{\{\widehat{\text{MSE}}(\hat{\theta}_i^w) \geq g_{1i}(A)\}} \\
&\leq \left[E \int_{g_{1i}(A)}^{\widehat{\text{MSE}}(\hat{\theta}_i^w)} (g_{1i}(A))^{-5/2} \left(\widehat{\text{MSE}}(\hat{\theta}_i^w) - g_{1i}(A)\right)^2 dx\right] I_{\{\widehat{\text{MSE}}(\hat{\theta}_i^w) \geq g_{1i}(A)\}} \\
&\leq \left[\frac{1}{3}(g_{1i}(A))^{-5/2} E\left(\widehat{\text{MSE}}(\hat{\theta}_i^w) - g_{1i}(A)\right)^3\right] \\
&= O(m^{\frac{-3}{2}}),
\end{aligned}
$$

using the Taylor series expansion $E\left(\widehat{\text{MSE}}(\hat{\theta}_i^w) - g_{1i}(A)\right)^3 = O(m^{\frac{-3}{2}})$.

In order to evaluate EI_2, we choose $\epsilon_m = m^{-\alpha}$ ($0 < \alpha < 1/2$) (see Datta et al., 2002). According to Diao et al. (2013) we can rewrite $I_{\{\widehat{\text{MSE}}(\hat{\theta}_i^w) \leq g_{1i}(A)\}}$ as follows

$$I_{\{\widehat{\text{MSE}}(\hat{\theta}_i^w) \leq g_{1i}(A)\}} = I_{\{\widehat{\text{MSE}}(\hat{\theta}_i^w) \leq g_{1i}(A) - \epsilon_m\}} + I_{\{g_{1i}(A) - \epsilon_m \leq \widehat{\text{MSE}}(\hat{\theta}_i^w) \leq g_{1i}(A)\}}.$$

Then,

$$\left[E \int_0^{g_{1i}(A) - \widehat{\text{MSE}}(\hat{\theta}_i^w)} x^2 \left(\widehat{\text{MSE}}(\hat{\theta}_i^w) + x \right)^{-5/2} dx \right] I_{\{\widehat{\text{MSE}}(\hat{\theta}_i^w) \leq g_{1i}(A) - \epsilon_m\}}$$

$$\leq \left[E \int_0^{g_{1i}(A) - \widehat{\text{MSE}}(\hat{\theta}_i^w)} x^{4/2 - 5/2} dx \right] I_{\{\widehat{\text{MSE}}(\hat{\theta}_i^w) \leq g_{1i}(A) - \epsilon_m\}}$$

$$= E \left[2 \left\{ g_{1i}(A) - \widehat{\text{MSE}}(\hat{\theta}_i^w) \right\}^{1/2} I_{\{\widehat{\text{MSE}}(\hat{\theta}_i^w) \leq g_{1i}(A) - \epsilon_m\}} \right]$$

$$\leq 2 \sqrt{E|g_{1i}(A) - \widehat{\text{MSE}}(\hat{\theta}_i^w)|} \sqrt{P \left(g_{1i}(A) - \widehat{\text{MSE}}(\hat{\theta}_i^w) \geq \epsilon_m \right)}$$

$$\leq O(m^{-1/2}) E \sqrt{\frac{\left(g_{1i}(A) - \widehat{\text{MSE}}(\hat{\theta}_i^w) \right)^3}{\epsilon_m^3}}$$

$$= O(m^{\frac{-3}{2}}).$$

Also,

$$E \left[\int_{\widehat{\text{MSE}}(\hat{\theta}_i^w)}^{g_{1i}(A)} x^{-5/2} \left(x - \widehat{\text{MSE}}(\hat{\theta}_i^w) \right)^2 dx I_{\{g_{1i}(A) - \epsilon_m \leq \widehat{\text{MSE}}(\hat{\theta}_i^w) \leq g_{1i}(A)\}} \right]$$

$$\leq (g_{1i}(A) - \epsilon_m)^{-5/2} E \left[\int_{\widehat{\text{MSE}}(\hat{\theta}_i^w)}^{g_{1i}(A)} \left(x - \widehat{\text{MSE}}(\hat{\theta}_i^w) \right)^2 dx I_{\{g_{1i}(A) - \epsilon_m \leq \widehat{\text{MSE}}(\hat{\theta}_i^w) \leq g_{1i}(A)\}} \right]$$

$$\leq \frac{1}{3} (g_{1i}(A) - \epsilon_m)^{-5/2} E \left(g_{1i}(A) - \widehat{\text{MSE}}(\hat{\theta}_i^w) \right)^3$$

$$= O(m^{\frac{-3}{2}}),$$

Thus, $EI_1 + EI_2 = O(m^{\frac{-3}{2}})$.

Combination of all the above expressions gives

$$
ES_1 = \frac{z}{2} \left(g_{1i}(A) \right)^{-1} \left(g_{2i}(A) + g_{3wi}(A) \right) - \frac{z}{8} \left(g_{1i}(A) \right)^{-2} \frac{\psi_i^4}{(A + \psi_i)^4} V(\hat{A}) + O(m^{-3/2})
$$

$$
= \frac{z}{2} \left(\frac{A\psi_i}{A + \psi_i} \right)^{-1} \left[\frac{\psi_i^2}{(A + \psi_i)^2} \mathbf{x}_i'(\sum_{i=1}^m \frac{\mathbf{x}_i \mathbf{x}_i'}{A + \psi_i})^{-1} \mathbf{x}_i \right.
$$

$$
+ (\gamma_i - w_i)^2 \left[(A + \psi_i) - \mathbf{x}_i' \left(\sum \frac{\mathbf{x}_i' \mathbf{x}_i}{A + \psi_i} \right)^{-1} \mathbf{x}_i \right] \right]
$$

$$
- \frac{z}{8} \left(\frac{A\psi_i}{A + \psi_i} \right)^{-2} \frac{\psi_i^4}{(A + \psi_i)^4} V(\hat{A}) + O(m^{-3/2}).
$$

R_2 can be written as

$$
R_2 = g_{1i}^{-1/2}(A)(\gamma_i - w_i)(y_i - \mathbf{x}_i'\beta) + o(m^{-1}).
$$

Using the Stein identity given by equation (5.6),

$$
ER_2 = g_{1i}^{-1/2}(A)(A + \psi_i) \frac{\partial(\gamma_i - w_i)}{\partial y_i} = 0.
$$

$ER_3 = 0$ since $\hat{A} - A = O_p(m^{-1/2})$ (see Datta et al., 2011). $F(z)^2$ can be written as:

$$
F(z)^2 = (S_1 + R_2 + R_3)^2
$$
$$
= S_1^2 + R_2^2 + R_3^2 + S_1 R_2 + S_1 R_3 + R_2 R_3.
$$

Evaluating the expected value of each term turn by turn reveals that

$$
ES_1^2 = \frac{z^2}{4} E \left(g_{1i}^{-1}(A)(g_{1i}^1(A)(\hat{A} - A)) \right)^2 + O(m^{-3/2})
$$

$$
= \frac{z^2}{4} g_{1i}^{-2}(A)(g_{1i}^1(A))^2 E(\hat{A} - A)^2 + O(m^{-3/2})
$$

$$
= \frac{z^2 \psi_i^2}{4A^2(A + \psi_i)^2} V(\hat{A}) + O(m^{-3/2}).
$$

and

$$R_2^2 = \left(\frac{(\gamma_i - w_i)(y_i - \mathbf{x}_i' \tilde{\boldsymbol{\beta}}(A))}{\sqrt{g_{1i}(A)}} \right)^2$$
$$= g_{1i}^{-1}(A)(\gamma_i - w_i)^2 (y_i - \mathbf{x}_i' \boldsymbol{\beta})^2 + o(m^{-1}).$$

Using the Stein identity (equation (5.6)) we have,

$$ER_2^2 = g_{1i}^{-1}(A)(\gamma_i - w_i)^2 (A + \psi_i) \frac{\partial}{\partial y_i}(y_i - \mathbf{x}_i' \boldsymbol{\beta}) + o(m^{-1})$$
$$= g_{1i}^{-1}(A)(\gamma_i - w_i)^2 (A + \psi_i) + o(m^{-1}).$$

R_3^2 can be simplified as

$$R_3^2 = \left(-\frac{w_i \mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A)(\hat{A} - A))}{\sqrt{g_{1i}(A)}} \right)^2$$
$$= g_{1i}^{-1}(A)(w_i \mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A))^2 (\hat{A} - A))^2.$$

Taking expectations on both sides

$$ER_3^2 = g_{1i}^{-1}(A)(w_i \mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A))^2 E(\hat{A} - A))^2$$
$$= g_{1i}^{-1}(A)(w_i \mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A))^2 V(\hat{A}),$$

since $\tilde{\boldsymbol{\beta}}^{(1)}(A) = O_P(m^{-1/2})$ (Datta et al., 2011).

$$S_1 R_2 = \frac{z \times \left(\sqrt{\widehat{MSE}(\hat{\theta}_i^w)} - \sqrt{g_{1i}(A)} \right)}{\sqrt{g_{1i}(A)}} \frac{(\gamma_i - w_i)(y_i - \mathbf{x}_i' \tilde{\boldsymbol{\beta}}(A))}{\sqrt{g_{1i}(A)}}$$
$$= g_{1i}^{-1} z(\gamma_i - w_i) \left(\sqrt{\widehat{MSE}(\hat{\theta}_i^w)} - \sqrt{g_{1i}(A)} \right)(y_i - \mathbf{x}_i' \tilde{\boldsymbol{\beta}}(A)).$$

Using the Taylor series expansion it follows that

$$S_1 R_2 = g_{1i}^{-1} z(\gamma_i - w_i) \left(g_{1i}^{(1)}(A)(\hat{A} - A) \right)(y_i - \mathbf{x}_i' \tilde{\boldsymbol{\beta}}(A))$$
$$= g_{1i}^{-1} z(\gamma_i - w_i) \left(g_{1i}^{(1)}(A)(\hat{A} - A) \right)(y_i - \mathbf{x}_i' \boldsymbol{\beta}) + o(m^{-1}).$$

Applying the Stein identity which is given by equation (5.6)

$$ES_1R_2 = g_{1i}^{-1} z (\gamma_i - w_i) g_{1i}^{(1)}(A)(A + \psi_i) \frac{\partial \hat{A}}{\partial y_i} + o(m^{-1}),$$

since $\frac{\partial \hat{A}}{\partial y_i} = O_p(m^{-1})$. Similarly $R_2 R_3$ can be simplified as

$$R_2 R_3 = \frac{(\gamma_i - w_i)(y_i - \mathbf{x}_i' \tilde{\boldsymbol{\beta}}(A))}{\sqrt{g_{1i}(A)}} \frac{w_i \mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A)(\hat{A} - A))}{\sqrt{g_{1i}(A)}}$$

$$= g_{1i}^{-1}(A)(\gamma_i - w_i)(y_i - \mathbf{x}_i' \boldsymbol{\beta}) w_i \mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A)(\hat{A} - A)) + o(m^{-1}).$$

Using the Stein identity (see equation (5.6))

$$ER_2R_3 = g_{1i}^{-1}(A) w_i (\gamma_i - w_i)(A + \psi_i) E \frac{\partial}{\partial y_i} \left\{ \mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A)(\hat{A} - A)) \right\}.$$

Note that

$$\frac{\partial}{\partial y_i} \left\{ \mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A)(\hat{A} - A) \right\} = \frac{\partial}{\partial y_i} \left\{ \mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A) \right\} (\hat{A} - A) + \mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A) \frac{\partial \hat{A}}{\partial y_i},$$

where

$$\frac{\partial}{\partial y_i} \left\{ \mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A) \right\} = \frac{\partial}{\partial A} \mathbf{x}_i' \left\{ \frac{\partial}{\partial y_i} \tilde{\boldsymbol{\beta}}(A) \right\}$$

$$= \frac{\partial}{\partial A} \left\{ \frac{\tilde{h}_{ii}}{A + \psi_i} \right\} + O(m^{-1}),$$

where $\tilde{h}_{ii} = \mathbf{x}_i' \left\{ \sum_{j=1}^{m} (A + \psi_j)^{-1} \mathbf{x}_j \mathbf{x}_j' \right\}^{-1}$ is $O(m^{-1})$, $\frac{\partial \hat{A}}{\partial y_i} = O_P(m^{-1})$ and $\frac{\partial}{\partial y_i}(\hat{\boldsymbol{\beta}}(\hat{A})) = O_P(m^{-1})$ (see Kubokawa, 2010; Datta et al., 2011). Hence it follows that

$$\frac{\partial}{\partial y_i} \left\{ \mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A)(\hat{A} - A)) \right\} = O(m^{-1}).$$

Similarly $S_1 R_3$ can be simplified as

$$S_1 R_3 = -\frac{z \times \left(\sqrt{\widehat{\text{MSE}}(\hat{\theta}_i^w)} - \sqrt{g_{1i}(A)} \right) w_i \mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A)(\hat{A} - A))}{\sqrt{g_{1i}(A)}} \frac{}{\sqrt{g_{1i}(A)}}$$

$$= -g_{1i}^{-1}(A) z w_i \left(\sqrt{\widehat{\text{MSE}}(\hat{\theta}_i^w)} - \sqrt{g_{1i}(A)} \right) \mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A)(\hat{A} - A).$$

Using the Taylor series expansion we have,

$$S_1 R_3 = -g_{1i}^{-1}(A)zw_i g_{1i}^{(1)}(A)(\hat{A} - A)\mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A)(\hat{A} - A)$$
$$= -g_{1i}^{-1}(A)zw_i g_{1i}^{(1)}(A)\mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A)(\hat{A} - A)^2.$$

Taking expectations on both sides

$$ES_1 R_3 = -g_{1i}^{-1}(A)zw_i g_{1i}^{(1)}(A)\mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A)E(\hat{A} - A)^2$$
$$= -g_{1i}^{-1}(A)zw_i g_{1i}^{(1)}(A)\mathbf{x}_i' \tilde{\boldsymbol{\beta}}^{(1)}(A)V(\hat{A}),$$

since $\tilde{\boldsymbol{\beta}}^{(1)}(A) = O_P(m^{-1/2})$ (Datta et al., 2011). Combination of all the above expressions gives

$$EF(z) = E(S_1) + E(R_2) + E(R_3)$$
$$= \frac{z}{2}\left(\frac{A\psi_i}{A + \psi_i}\right)^{-1}\left[\frac{\psi_i^2}{(A + \psi_i)^2}\mathbf{x}_i'(\sum_{i=1}^{m}\frac{\mathbf{x}_i\mathbf{x}_i'}{A + \psi_i})^{-1}\mathbf{x}_i\right.$$
$$+ \left. (\gamma_i - w_i)^2\left[(A + \psi_i) - \mathbf{x}_i'\left(\sum\frac{\mathbf{x}_i'\mathbf{x}_i}{A + \psi_i}\right)^{-1}\mathbf{x}_i\right]\right]$$
$$- \frac{z}{8}\left(\frac{A\psi_i}{A + \psi_i}\right)^{-2}\frac{\psi_i^4}{(A + \psi_i)^4}V(\hat{A}) + O(m^{-3/2}).$$

Also,

$$E(F(z) - F(-z)) = z\frac{\psi_i}{A(A + \psi_i)}\mathbf{x}_i'(\sum_{i=1}^{m}\frac{\mathbf{x}_i\mathbf{x}_i'}{A + \psi_i})^{-1}\mathbf{x}_i$$
$$+ z\left(\frac{A\psi_i}{A + \psi_i}\right)^{-1}(\gamma_i - w_i)^2(A + \psi_i))$$
$$- z\left(\frac{A\psi_i}{A + \psi_i}\right)^{-1}(\gamma_i - w_i)^2\mathbf{x}_i'\left(\sum\frac{\mathbf{x}_i'\mathbf{x}_i}{A + \psi_i}\right)^{-1}\mathbf{x}_i$$
$$- \frac{z}{4}\frac{\psi_i^2}{A^2(A + \psi_i)^2}V(\hat{A}) + O(m^{-3/2}).$$

Also,

$$EF(z)^2 = ES_1^2 + ER_2^2 + ER_3^2 + ES_1R_2 + ES_1R_3 + ER_2R_3$$

$$= \frac{z^2\psi_i^2}{4A^2(A+\psi_i)^2}V(\hat{A}) + g_{1i}^{-1}(A)(\gamma_i - w_i)^2(A+\psi_i) + O(m^{-3/2}).$$

and,

$$E(F(z)^2 + F(-z)^2) = \frac{z^2\psi_i^2}{2A^2(A+\psi_i)^2}V(\hat{A}) + 2\left(\frac{A\psi_i}{A+\psi_i}\right)^{-1}(\gamma_i - w_i)^2(A+\psi_i) + O(m^{-3/2}).$$

So,

$$E\left[F(z) - F(-z) - \frac{z}{2}(F(z)^2 + F(-z)^2)\right]$$

$$= -z\phi(z)\left\{(z^2 + 1)\frac{\psi_i^2}{4A^2(A+\psi_i)^2}V(\hat{A}) - \frac{\psi_i^2 + (\gamma_i - w_i)^2(A+\psi_i)^2}{A\psi_i(A+\psi_i)}\mathbf{x}_i'\left(\sum\frac{\mathbf{x}_i'\mathbf{x}_i}{A+\psi_i}\right)^{-1}\mathbf{x}_i\right\}.$$

The remainder term $\int_z^{z+F(z)}(z + F(z) - x)^2(x^2 - 1)\phi(x)dx$ can be simplified using the approach by Datta et al. (2002), Chatterjee and lahiri (2002) and Diao et al. (2013) as follows. Since within the limits of the integral $0 < |z + F(z) - x| < |F(z)|$ and $|(x^2 - 1)\phi(x)| < 2\phi(\sqrt{3})$, it follows that

$$E|\int_z^{z+F(z)}(x^2 - 1)(z + F(z) - x)^2\phi(x)dx| \leq E|F(z)|^2|\int_z^{z+F(z)}2\phi(\sqrt{3})dx|$$

$$= 2E|F(z)|^3\phi(\sqrt{3})$$

$$= O(m^{-3/2}).$$

Combination and simplifications of the above results give the CP of $I^{CW}(\hat{A})$ as

$$P[\theta_i \in I^{CW}(\hat{A})] = E[\Phi(z + F(z)) - \Phi(-z + F(-z))]$$

$$= 2\Phi(z) - 1 - z\phi(z)\eta_w + O(m^{-3/2}),$$

where

$$\eta_w = (z^2 + 1)\frac{\psi_i^2}{4A^2(A+\psi_i)^2}V(\hat{A}) + \frac{\psi_i^2 + (\gamma_i - w_i)^2(A+\psi_i)^2}{A\psi_i(A+\psi_i)}\mathbf{x}_i'\left(\sum\frac{\mathbf{x}_i'\mathbf{x}_i}{A+\psi_i}\right)^{-1}\mathbf{x}_i.$$

□