Jackknife and Bootstrap Methods for Small Area Estimation

J.N.K. Rao[1]
School of Mathematics and Statistics, Carleton University, Ottawa, Canada K1S 5B6[1]

Abstract

Resampling methods have long been used in survey sampling, dating back to Mahalanobis (1946). More recently, jackknife and bootstrap resampling methods have been proposed for small area estimation; in particular for mean squared error (MSE) estimation and confidence intervals. We present a brief overview of early uses of resampling methods in survey sampling, and then provide an appraisal of recent resampling methods for small area estimation.

KEY WORDS: Bootstrap, Confidence intervals, Conditional properties, Jackknife, MSE estimation, Small area models.

1. Early Uses of Resampling

The importance of measurement errors in sample surveys was recognized as early as the 1940's. Mahalanobis (1946) developed the technique of interpenetrating subsamples (also called replicated sampling, Deming (1960)) for assessing both sampling and measurement errors, and used it extensively in large-scale sample surveys in India. The sample is drawn in the form of two or more independent subsamples according to the same sampling design such that each subsample provides a valid estimate of the finite population total or mean. By assigning the subsamples to different interviewers (or teams), a valid estimate of the total variance, that takes proper account of the correlated response variance due to interviewers, is obtained. Hall (2003) provides a scholarly historical account of Mahalanobis' seminal contributions to early development of survey sampling in India.

For the case of independent and identically distributed (IID) observations $y_1, \ldots, y_n$, Quenouille (1956) developed an ingenious method of bias reduction in a full-sample estimator, $\hat{\theta}$, of a model parameter θ. The sample of size n is first divided at random into g groups $G_1, \ldots, G_g$, each of size m, assuming that $n = gm$. The groups, G_j, are deleted in turn and the delete-group estimates $\hat{\theta}_{(j)}$, $j = 1, \ldots, g$, are computed. Quenouille (1956) showed that the estimator

$$\hat{\theta} = \frac{1}{g}\sum_{j=1}^{g}\left\{g\hat{\theta} - (g-1)\hat{\theta}_{(j)}\right\}$$

$$\equiv g\hat{\theta} - (g-1)\hat{\theta}_{(\cdot)} \equiv \frac{1}{g}\sum_{j=1}^{g}\hat{\theta}_{Q_j}$$

leads to bias reduction, in the sense that the bias of $\hat{\theta}_Q$ is of order $O(n^{-2})$ if the bias of $\hat{\theta}$ is of the form

$$B(\hat{\theta}) = \frac{a}{n} + \frac{b}{n^2} + O\Big(\frac{1}{n^3}\Big).$$

In the sample survey context, Durbin (1959) applied Quenouille's method to ratio estimation, using $g = 2$ groups. Rao (1963) and Rao and Webster (1966) studied the optimal choice of g for bias reduction in ratio estimation, and showed that $g = n$ is the optimal choice. In the latter case, we have the delete–1 jackknife.

Tukey (1958) noted that for $g = n$ and $\hat{\theta} = \overline{y}$, the sample mean, the "pseudo-values" $\hat{\theta}_{Qj}$ reduce to $\hat{\theta}_{Qj} = y_j$ and hence IID. Motivated by this result, Tukey suggested regarding the $\hat{\theta}_{Qj}$ as IID for general $\hat{\theta}$ and then using

$$v_J(\hat{\theta}_Q) = \frac{1}{n(n-1)}\sum_{j=1}^{n}\Big(\hat{\theta}_{Qj} - \hat{\theta}_Q\Big)^2$$

$$= \frac{n-1}{n}\sum_{j=1}^{n}\Big\{\hat{\theta}_{(j)} - \hat{\theta}_{(\cdot)}\Big\}^2$$

as the "jackknife" variance estimator of $\hat{\theta}_Q$ or $\hat{\theta}$. Note that $v_J(\hat{\theta}_Q)$ is computer-intensive if $\hat{\theta}$ requires iterative calculation, because n sets of iterative calculations need to be performed to calculate $\hat{\theta}_{(j)}$, $j = 1, \ldots, n$ and hence the jackknife variance estimate. In the 50's this was indeed a problem, given the state of high-speed computing in those days. Miller (1964) established the asymptotic consistency of v_J for smooth functions of means, $\hat{\theta}$ and studied the question "Is the jackknife trustworthy?" We refer the reader to Shao and Tu (1995, Chapter 2) for later work on the jackknife.

Wu (1986) studied the linear regression model $y_i = x_i'\beta + \varepsilon_i$, where the independent model errors ε_i have zero mean and possibly unequal variances σ_i^2. Let $\hat{\beta}$ be the ordinary least squares estimator of β and $\hat{\theta} = g(\hat{\beta})$ for some vector smooth function $g(\cdot)$. Under the weighted jackknife method, pairs (y_i, x_j) are deleted in turn for $j = 1, \ldots, n$ and the resulting estimates $\hat{\beta}_{(j)}$ and $\hat{\theta}_{(j)} = g(\hat{\beta}_{(j)})$ are computed. The weighted jackknife variance estimator of $\hat{\theta}$ is then given by

$$v_{Jw}(\hat{\theta}) = \sum_{j=1}^{n}(1 - w_j)(\hat{\theta}_{(j)} - \hat{\theta})(\hat{\theta}_{(j)} - \hat{\theta})',$$

where $w_j = x_j'\left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} x_j$. Wu (1986) established the asymptotic consistency of $v_{Jw}(\hat{\theta})$ under the condition $\max(w_j) \to 0$ as $n \to \infty$. He also showed that in the linear case $\hat{\theta} = \hat{\beta}$, the weighted jackknife variance estimator is exactly unbiased if the error variances σ_i^2 are equal ($\sigma_i^2 = \sigma^2$). In Section 3, we show that both Quenouille's bias reduction method and Tukey's jackknife or Wu's weighted jackknife play important roles in MSE estimation for small areas.

Bootstrap resampling was first introduced by Efron (1979). Efron's pioneering 1979 paper for the IID case and the subsequent enormous amount of research on bootstrap had a huge impact on the practice of statistics, especially after the ready availability of high-speed computing. Bootstrap offers wider flexibility than the jackknife, and in the IID case the bootstrap variance estimator for non-smooth estimators, like the median, is consistent unlike the delete–1 jackknife. Moreover, it can provide "better" confidence intervals than the normal approximation based methods. We refer the reader to the excellent books by Hall (1992) and Shao and Tu (1995) for detailed theoretical accounts of the bootstrap method.

Stratified multi-stage cluster sampling is commonly used in large-scale socio-economic surveys. Pioneering work on delete-cluster jackknife and balanced repeated replication (BRR) for variance estimation is due to McCarthy (1969) and Kish and Frankel (1974). Krewski and Rao (1981) provided theoretical justification by establishing the asymptotic consistency of delete-cluster jackknife and BRR variance estimators for surveys with a large number of strata and small numbers of sampled clusters within strata. They considered estimators $\hat{\theta}$ that can be expressed as smooth functions of estimated totals or means. We refer the reader to Shao and Tu (1995, Chapter 6) for various extensions including the consistency of BRR variance estimator for non-smooth estimators such as the median; consistency or inconsistency of the delete-cluster jackknife in the non-smooth case is not known.

Bootstrap sampling of first-stage clusters within strata was studied by Rao and Wu (1988), Rao, Wu and Yue (1992), Sitter (1992) and others. Bootstrap offers flexibility in terms of number of resamples, B, especially for surveys with a large number of first-stage sample clusters, unlike the delete-cluster jackknife. The data file reports the sample data as well as the associated full sample weights and the B bootstrap weights. The user simply computes $\hat{\theta}, \hat{\theta}_1, \ldots, \hat{\theta}_B$ from the data file, using the full sample weights and the B bootstrap weights. The bootstrap variance estimator of $\hat{\theta}$ is simply obtained as

$$v_{\text{BOOT}}(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^{B} (\hat{\theta}_b - \hat{\theta})(\hat{\theta}_b - \hat{\theta})'.$$

Bootstrap with $B = 500$ is currently used in Statistics Canada for variance estimation.

Lahiri (2003) provides a nice account of the impact of bootstrap in survey sampling.

2. Basic Area Level Model

Model-based small area estimation has received a lot of attention in recent years due to its potential in providing reliable area level estimates, even with small area-specific sample sizes, by borrowing information across areas through linking models based on auxiliary information. A basic Fay-Herriot (FH) area level model is obtained as follows. Let $\theta_i = g(Y_i)$ be a suitable function of the i-th small area total Y_i linearly related to predictor variables z_i, $i = 1, \ldots, m$. The linking model is given by

$$\theta_i = z_i'\beta + v_i; \quad v_i \sim_{\text{iid}} N(0, \sigma_v^2).$$

A matching sampling model is of the form

$$\hat{\theta}_i = g(\hat{Y}_i) = \theta_i + e_i, \quad e_i \sim_{\text{iid}} N(0, \psi_i)$$

with known sampling variance ψ_i, where $\hat{Y}_i$ is a direct estimator of Y_i (Fay and Herriot (1979)). A mismatched sampling model $\hat{Y}_i = Y_i + f_i$ with $E(f_i) = 0$ is more realistic for small area samples because $E[g(\hat{Y}_i)]$ can differ significantly from θ_i if $g(\cdot)$ is non-linear. However, we focus on the simple case $\theta_i = Y_i$ in which case the two sample models are identical.

The best estimator (under squared loss) of θ_i is given by $\hat{\theta}^B = E(\theta_i|\hat{\theta}_i, \beta, \sigma_v^2) \equiv h(\hat{\theta}_i, \beta, \sigma_v^2)$. We estimate the model parameters β and σ^2 by a suitable method, such as maximum likelihood (ML), residual maximum likelihood (REML) or the FH method of moments. Substituting the estimators $\hat{\beta}$ and $\hat{\sigma}_v^2$ in $\hat{\theta}_i^B$, we get the empirical best (EB) estimator: $\hat{\theta}_i^{\text{EB}} \equiv h(\hat{\theta}_i, \hat{\beta}, \hat{\sigma}_v^2) = \hat{\gamma}_i\hat{\theta}_i + (1 - \hat{\gamma}_i)z_i'\hat{\beta}$ under the FH area level model, where $\gamma_i = \sigma_v^2/(\sigma_v^2 + \psi_i)$. This estimator is also the empirical best linear unbiased prediction (EBLUP) estimator without normality assumption.

Mean squared error of $\hat{\theta}_i^{\text{EB}}$ may be written as

$$\begin{aligned} \text{MSE}(\hat{\theta}_i^{\text{EB}}) &= E(\hat{\theta}_i^{\text{EB}} - \theta_i)^2 \\ &= E(\hat{\theta}_i^B - \theta_i)^2 + E(\hat{\theta}_i^{\text{EB}} - \theta_i)^2 \\ &\equiv M_{1i}(\sigma_v^2) + M_{2i}(\sigma_v^2). \end{aligned} \tag{1}$$

For the FH model, the leading term in (1) is $M_{1i}(\sigma_v^2) = g_{1i}(\sigma_v^2) = \gamma_i\psi_i$ which shows efficiency gain over the direct estimator $\hat{\theta}_i$ with $\text{MSE}(\hat{\theta}_i) = E(\hat{\theta}_i - \theta_i)^2 = \psi_i$. No closed-form expression for $M_{2i}(\sigma_v^2)$ exists. Prasad and Rao (1990), Datta and Lahiri (2000) and Datta, Rao and Smith (2005) obtained a Taylor linearization approximation to $M_{2i}(\sigma_v^2)$ for large m as $M_{2i}(\sigma_v^2) \approx g_{2i}(\sigma_v^2) + g_{3i}(\sigma_v^2)$, where the neglected terms are of order $O(m^{-2})$, and $g_{2i}(\sigma_v^2)$ and $g_{3i}(\sigma_v^2)$, depend on the asymptotic variance of $\hat{\beta}$ and $\hat{\sigma}_v^2$, respectively. Note that the neglected terms in the second order approximation, $g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) + g_{3i}(\sigma_v^2)$, to $\text{MSE}(\hat{\theta}_i^{\text{EB}})$ are of order $O(m^{-2})$.

Turning to MSE estimation, a "nearly" unbiased estimator under REML is given by (Datta and Lahiri (2000))

$$\text{mse}(\hat{\theta}_i^{\text{EB}}) = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{3i}(\hat{\sigma}_v^2). \tag{2}$$

Under ML and FH moment method, an extra bias correction term is needed (Datta and Lahiri (2000), Datta, Rao and Smith (2005), Rao (2003, Chapter 7). Note that the MSE estimator (2) is not area-specific in the sense that it does not depend on $\hat{\theta}_i$. Alternatives to $2g_{3i}(\hat{\sigma}_v^2)$ that depend on $\hat{\theta}_i$ have been proposed (Rao (2003, Chapter 7)).

If $\theta_i = g(Y_i)$, then the best estimator of Y_i, $E(Y_i|\hat{Y}_i, \beta, \sigma_v^2) = h(\hat{Y}_i, \beta, \sigma_v^2)$ has no closed form expression. As a result, MSE estimation using Taylor linearization becomes complex or difficult. In Sections 3 and 4, we show that the jackknife and bootstrap can be used to handle such general cases including generalized linear mixed models.

3. Jackknife MSE Estimation

Jiang, Lahiri and Wan (2002) proposed a jackknife estimator of MSE($\hat{\theta}_i^{\text{EB}}$) that avoids the explicit evaluation of $g_{2i}(\cdot)$ and $g_{3i}(\cdot)$ terms in (2), but it still requires the derivation of $g_{1i}(\cdot)$ term which is simple for the EB estimator $\hat{\theta}_i^{\text{EB}}$ above They applied Tukey's jackknife idea to get a delete-area jackknife estimator of $M_{2i}(\sigma_v^2)$. Let $\phi = (\beta, \sigma_v^2)$ and $\hat{\phi}_{(u)}$ denote the delete u-th area estimator of ϕ; $u = 1, \ldots, m$. Then, the Jiang, Lahiri and Wan (JLW) unweighted jackknife estimators of $M_{2i}(\sigma_v^2)$ is given by

$$M_{2i,J} = \frac{m-1}{m} \sum_{u=1}^{m} \left\{ h\big(\hat{\theta}_i, \hat{\phi}_{(u)}\big) - h\big(\hat{\theta}_i, \hat{\phi}\big) \right\}^2. \qquad (3)$$

Quenouille's bias reduction method is applied to $M_{1i}(\hat{\sigma}_v^2)$ in (1) to get

$$\hat{M}_{1i,J} = g_{1i}(\hat{\sigma}_v^2) - \frac{m-1}{m} \sum_{u=1}^{m} \left\{ g_{1i}(\hat{\sigma}_{v(u)}^2) - g_{1i}(\hat{\sigma}_v^2) \right\}. \qquad (4)$$

JLW proved that $\hat{M}_{i,J} = \hat{M}_{1i,J} + \hat{M}_{2i,J}$ is a nearly unbiased estimator of MSE($\hat{\theta}_i^{\text{EB}}$) in the sense that its bias is of lower order than m^{-1}. A weighted version is obtained by applying Wu's weighted jackknife method (Chen and Lahiri (2002)) with weights $w_u = 1 - (z_u'/\psi_u)(\Sigma z_i z_i'/\psi_i)^{-1} z_u$: Replace $(m-1)/m$ in (3) and (4) by w_u $(u = 1, \ldots, m)$ and take it inside the summation terms. Note that $\hat{M}_{2i,J}$ and its weighted version are area-specific in the sense of depending on $\hat{\theta}_i$. The weighted jackknife version performed better in small samples $(m = 12)$ than $\hat{M}_{i,J}$ (Chen and Lahiri (2002)).

As noted by Bell (2001) in the context of FH model, the jackknife estimator $\hat{M}_{i,J}$, due to bias correction in (4), can take negative values under certain scenarios. Chen and Lahiri (2005) used jackknife linearization, under the REML estimator $\hat{\sigma}_v^2$, to get

$$\hat{M}_{i,JL} = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + \frac{\psi_i^2}{(\hat{\sigma}_v^2 + \psi_i)^2} v_{wJ}(\hat{\sigma}_v^2) + \frac{\psi_i^2}{(\hat{\sigma}_v^2 + \psi_i)^4} (\hat{\theta} - z_i'\hat{\beta})^2 v_{wJ}(\hat{\sigma}_v^2), \qquad (5)$$

where $v_{wJ}(\hat{\sigma}_v^2) = \sum_{u=1}^{m} w_u \big(\hat{\sigma}_{v(u)}^2 - \hat{\sigma}_v^2\big)^2$ is a weighted jackknife variance estimator of $\hat{\sigma}_v^2$. The estimator (5) is always non-negative, unlike $\hat{M}_{i,J}$ or its weighted version, but requires additional analytical work as in the case of (2). An extra term involving the bias of $\hat{\sigma}_v^2$ needs to be subtracted in the case of ML and FH estimators of $\hat{\sigma}_v^2$, and this could lead to negative MSE estimates in rare cases. A simulation study indicated superior performance of the proposed jackknife linearization MSE estimator in (5).

The JLW jackknife method is applicable to general small area models, including mismatched models and non-normal cases (binary or count unit level responses). We simply start with the best estimator of the small area parameter of interest, given the model parameters ϕ. But it may not have a closed form expression and hence may require numerical integration for specified ϕ. Moreover, the leading M_{1i} (or g_{1i}) term of the MSE can involve complex numerical computations, and it is required for bias correction as in the FH model. Lohr and Rao (2007) proposed an alternative jackknife MSE estimator that avoids the extra integration or summation with respect to marginal distribution, and as a result it is computationally simpler than the JLW estimator of MSE. Also, its leading term in nonlinear cases is area-specific, in the sense of depending on the area-specific data, unlike the JLW estimator.

To illustrate that Lohr-Rao method, consider the simple case of $y_i \sim_{\text{iid}} B(n_i, p_i)$ given p_i and $p_i \sim_{\text{ind}}$ Beta (α, β), $i = 1, \ldots, m$, and the parameter of interest is p_i. In this case, the best estimator of p_i is $\hat{p}_i^B = E(p_i|y_i, \phi) \equiv h(y_i, \phi)$ and the EB estimator is $\hat{p}_i^{\text{EB}} = h(y_i, \hat{\phi})$, where $\hat{\phi} = (\hat{\alpha}, \hat{\beta})$ is a consistent estimator of $\phi = (\alpha, \beta)$. We have

$$\text{MSE}(\hat{p}_i^{\text{EB}}) = EV(p_i|y_i, \phi) + E(\hat{p}_i^{\text{EB}} - \hat{p}_i^B)^2 \equiv M_{1i} + M_{2i}. \qquad (6)$$

JLW need M_{1i} in (6) as a function of ϕ to get their bias corrected estimator $\hat{M}_{1i,J}$ which is not area-specific. Area-specific estimator, $\hat{M}_{2i,J}$, of M_{2i} is given by (3) with $\hat{\theta}_i$ replaced by y_i. Let $V(p_i|y_i, \phi) = \tilde{g}_{1i}(y_i, \phi)$ which depends on area-specific data, unlike in the FH case studied above. Following a suggestion of Rao (2003, Chapter 9), Lohr and Rao (2007) applied jackknife bias corrections to $\tilde{g}_{1i}(y_i, \hat{\phi})$ to get the following estimator of M_{1i}:

$$\tilde{M}_{1i}(y_i) = \tilde{g}_{1i}(y_i, \hat{\phi}_{(i)}) - \frac{m-1}{m} \sum_{u=1}^{m} \left\{ \tilde{g}_{1i}(y_i, \hat{\phi}_{(u)}) - \tilde{g}_{1i}(y_i, \hat{\phi}) \right\} \qquad (7)$$

The JLW estimator $\hat{M}_{2iJ}$ is used for M_{2i} in (6). The Lohr-Rao (LR) estimator $\tilde{M}_{i,J} = \tilde{M}_{1i}(y_i) + \hat{M}_{2iJ}$ is nearly conditionally unbiased given y_i, unlike the JLW estimator, and also nearly unbiased unconditionally as in the case of JLW. Note that in the FH model case, the posterior variance given ϕ, $V(\theta_i|\hat{\theta}_i, \phi)$, does not depend on $\hat{\theta}$, unlike in the non-linear case. Hence, it is not possible to obtain an area-specific estimator of the leading term $M_{1i} = g_{1i}(\sigma_v^2)$ in the FH case.

Lohr and Rao (2007) conducted a simulation study under the above binomial-beta model. Their results may be summarized as follows: (1) Unconditional relative bias (URB) of both JLW and LR estimators is small, but coefficient of variation (CV) of JLW is smaller, as expected, because $\tilde{g}_{1i}(y_i, \phi)$ differs for each value of y_i unlike $\hat{M}_{1i,J}$. (2) Conditional relative bias (CRB) of LR is small and it decreases as m increases, while JLW exhibits strong pattern for CRB: large and positive when y_i is small or large, and large and negative when y_i is close to the middle. Moreover, CRB of JLW does not necessarily decrease as m increases.

4. Bootstrap MSE Estimation

Parametric bootstrap versions of the JLW jackknife MSE estimator, $\hat{M}_{iJ}$, have been proposed by Butar and Lahiri (2003) and Pfeffermann and Glickman (2004). For the FH model under normality, B bootstrap samples $\{(\hat{\theta}_i^b, z_i); i = 1, \ldots, m\}$, $b = 1, \ldots, B$ are generated as follows: (i) Generate $\hat{v}_i^b$ and $\hat{e}_i^b$ independently from $N(0, \hat{\sigma}_v^2)$ and $N(0, \psi_i)$ respectively, (ii) Let $\hat{\theta}_i^b = z_i'\hat{\beta} + \hat{v}_i^b + \hat{e}_i^b \equiv \theta_i^b + \hat{e}_i^b$. Using the b-th bootstrap sample, we calculate the estimators $\hat{\sigma}_v^2(b)$ and $\hat{\beta}(b)$ and the resulting EB estimators $h(\hat{\theta}_i^b, \hat{\phi}(b))$.

The components corresponding to $\hat{M}_{1i,J}$ and $\hat{M}_{2i,J}$ are then given by (Butar and Lahiri (2003)):

$$\hat{M}_{1i,B} = g_{1i}(\hat{\sigma}_v^2) - \frac{1}{B}\sum_{b=1}^{B}\{g_{1i}(\hat{\sigma}_v^2(b)) - g_{1i}(\hat{\sigma}_v^2)\}$$

$$= 2g_{1i}(\hat{\sigma}_v^2) - \frac{1}{B}\sum_{b=1}^{B} g_{1i}(\hat{\sigma}_v^2(b)) \quad (8)$$

and

$$\hat{M}_{2i,B} = \frac{1}{B}\sum_{b=1}^{B}\left\{h(\hat{\theta}_i, \hat{\phi}(b)) - h(\hat{\theta}_i, \hat{\phi})\right\}^2, \quad (9)$$

leading to $\hat{M}_{i,B} = \hat{M}_{i1,B} + \hat{M}_{2i,B}$ as the bootstrap MSE estimator of $\hat{\theta}_i^{\text{EB}}$. Pfeffermann and Glickman (2004) proposed a different version of $\hat{M}_{2i,B}$, but $\hat{M}_{1i,B}$ is not changed:

$$\tilde{M}_{2i,B} = \frac{1}{B}\sum_{b=1}^{B}\left\{h(\hat{\theta}_i^b, \hat{\phi}(b)) - \theta_i^b\right\}^2. \quad (10)$$

They provide a heuristic argument that the resulting MSE estimator $\hat{M}_{1i,B} + \tilde{M}_{2i,B}$ has "the advantage of potential robustness against sampling from non-normal distributions". The above bootstrap methods extend to more general models, as in the jackknife case. A possible disadvantage of the bootstrap method is that the bias of the MSE estimator may be sensitive to the choice of number of bootstrap samples, B. It may be advisable to study sensitivity as B changes.

As noted before, for general small area models it is difficult to evaluate the M_{1i} term. Instead, it is possible to develop a bootstrap analogue of the Lohr-Rao method and get a computationally simpler and area-specific MSE estimator that is conditionally as well as unconditionally unbiased. Hall and Maiti (2006) and Chatterjee and Lahiri (2007a) developed a general double bootstrap method that is computer-intensive and avoids the evaluation of the M_{1i}-term. We illustrate the method for the FH model but it is applicable for general parametric models. First, we note that $\text{MSE}(\hat{\theta}_i^{\text{EB}}) = E\{h(\hat{\theta}_i, \hat{\phi}) - \theta_i\}^2$ which suggests that a naive estimator based on the (level 1) bootstrap samples, b, is given by $\tilde{M}_{2i,B}$ in (10). Next, we perform bootstrap bias correction of $\tilde{M}_{2i,B}$ using level 2 bootstrap samples. The c-th level 2 bootstrap sample $\{(\hat{\theta}^b(c), z_i); i = 1, \ldots, m\}$, $c = 1, \ldots, C$ associated with the b-th level 1 bootstrap sample is obtained by generating $v_i^b(c)$ and $e_i^b(c)$ independently from $N(0, \hat{\sigma}_v^2(b)$, and $N(0, \psi_i)$ and then letting $\hat{\theta}_i^b(c) = z_i'\hat{\beta}(b) + \hat{v}_i^b(c) + \hat{e}_i^b(c) \equiv \theta_i^b(c) + \hat{e}_i^b(c)$. Using the (bc)-th level 2 bootstrap sample we calculate the estimators $\hat{\sigma}_v^2(bc)$ and $\hat{\beta}(bc)$ and the resulting EB estimators $h\{\hat{\theta}_i^b(c), \hat{\phi}(bc)\}$. Let

$$\tilde{M}_{2i\text{BC}} = \frac{1}{\text{BC}}\sum_{b=1}^{\text{B}}\sum_{c=1}^{\text{C}}\left\{h(\hat{\theta}_i(bc), \hat{\phi}(b)) - \theta_i^b(c)\right\}^2 \quad (11)$$

Then the bias-corrected estimator of $\text{MSE}(\hat{\theta}_i^{\text{EB}})$ is given by

$$\tilde{M}_{i,\text{BC}} = 2\tilde{M}_{2i,\text{B}} - \tilde{M}_{2i,\text{BC}}. \quad (12)$$

The estimator $\tilde{M}_{i,\text{BC}}$ is nearly unbiased for very large B and C, but its bias may be quite sensitive to the choice of B and C (Tang and Jiang (2007)).

Hall and Maiti (2006) studied MSE estimation under a unit level nested error linear regression model $y_{ij} = x_{ij}'\beta + v_i + e_{ij}$, $j = 1, \ldots, i$; $i = 1, \ldots, m$, with v_i and e_{ij} independent and having zero means and finite second and fourth moments, where n_i is the number of sample observations (y_{ij}, x_{ij}) in small area i and the population mean $\overline{X}_i$ is known. Customary normality assumption on v_i and e_{ij} is thus relaxed. Hall and Maiti (2006) proposed drawing B level 1 bootstrap samples from distributions that match the estimated second and fourth moments of v_i and e_{ij} and then computing the empirical best linear unbiased prediction (EBLUP) estimators of small area means of y from the level 1 bootstrap samples. The resulting MSE estimator of the form (10) is then bias-corrected using a double bootstrap with C level 2 bootstrap samples form each level 1 bootstrap sample using the same moment matching method. The resulting MSE estimator of the form (12) is nearly unbiased for very large B and C. The Hall-Maiti method could also be used under the FH model without normality assumption, but it could be quite involved for general linear mixed models, such as two level models, because the fourth moments need to be estimated. Again, the bias of the MSE estimator could be quite sensitive to the choice of B and C.

Bootstrap methods seem to have the potential to provide second order accurate confidence intervals on small area parameters (Chatterjee, Lahiri and Li (2007b)). Further work on this topic would be practically useful and theoretically challenging.

REFERENCES

Bell, W. (2001), Discussion of "Jackknife in the Fay-Herriot model with an example", *Proceedings of the Seminar on Funding Opportunity in Survey Research*, 98–104.

Butar, F.B. and Lahiri, P. (2003), "On measures of uncertainty of empirical Bayes small area estimators". Model selection, model diagnostics, empirical Bayes and hierarchical Bayes, Special Issue II (Lincoln, NE, 1999), *Journal of Statistical Planning and Inference*, **112**(1-2), 63–76.

Chatterjee, S. and Lahiri, P. (2007a), "A simple computational method for estimating mean squared prediction error in general small-area models," paper presented at the JSM 2007, Salt Lake City.

Chatterjee, S. and Lahiri, P. and Li, H. (2007b), "On small-area prediction interval problems," paper presented at the JSM 2007, Salt Lake City.

Chen, S. and Lahiri, P. (2002), "A weighted jackknife MSPE estimator in small area estimation," *ASA Proceedings of the Survey Research Methods Section*, pp. 473–477.

Chen, S. and Lahiri, P. (2005), "On mean squared prediction error estimation in small area estimation problems," *ASA Proceedings of the Survey Research Methods Section*.

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.

Deming, W.E. (1960). *Sample Design in Business Research*, New York: Wiley.

Durbin, J. (1959), "A note on the application of Quenouille's method of bias reduction to the estimation of ratios," *Biometrika*, **46**, 477–480.

Efron, B. (1979), "Bootstrap methods: another look at the jackknife'" *Annals of Statistics*, **7**, 1–26.

Fay, R.E. and 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.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

Hall, P. (2003), "A short prehistory of the bootstrap," *Statistical Science*, **18**, 158–167.

Hall, P. and Maiti, T. (2006), "Nonparametric estimation of mean-squared prediction error in nested-error regression models," *Annals of Statistics*, **34**, 1733-1750.

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

Kish, L. and Frankel, M.R. (1974), "Inference from complex samples," *Journal of the Royal Statistical Society*, series B, **36**, 1–37.

Krewski, D. and Rao, J.N.K. (1981), "Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods," *Annals of Statistics*, **9**, 1010–1019.

Lahiri, P. (2003), "On the impact of bootstrap in survey sampling and small area estimation," *Statistical Science*, **18**, 199–210.

Lohr, S. and Rao, J.N.K. (2007), "Jackknife estimation of mean squared error of small area predictors in nonlinear mixed models," submitted for publication.

Mahalanobis, P.C. (1946), "Recent experiments in statistical sampling in the Indian Statistical Institute'" *Journal of the Royal Statistical Society*, **109**, 325–378.

Miller, R.G. (1964), "A trustworthy jackknife," *Annals of Mathematical Statistics*, **35**, 1594–1605.

Pfeffermann, D. and Glickman, H. (2004), "Mean square error approximation in small area estimation by use of parametric and nonparametric bootstrap," *ASA Proceedings of the Survey Research Methods Section*,

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

Quenouille, M.H. (1956), "Notes on bias in estimation," *Biometrika*, **43**, 353–360.

Rao, J.N.K. (1965), "A note on the estimation of ratios by Quenouille method," *Biometrika*, **52**, 647–649.

Rao, J.N.K. (2003), *Small Area Estimation*, New York: Wiley.

Rao, J.N.K. and Webster (1966), "On two methods of bias reduction," *Biometrika*, **53**, 571–577.

Rao, J.N.K. and Wu, C.F.J. (1988), "Resampling inference with complex survey data," *Journal of the American Statistical Association*, **83**, 231–241.

Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992), "Some recent work on resampling methods for complex surveys," *Survey Methodology*, **18**, 209–217.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*, New York: Springer.

Sitter, R.R. (1992), "A resampling procedure for complex survey data," *Journal of the American Statistical Association*, **87**, 755–765.

Tang, E. and Jiang, J. (2007), "On comparison of different MSE estimators in small-area estimation," paper presented at the JSM 2007, Salt Lake City.

Tukey, J.W. (1958), "Bias and confidence in not quite large samples," *Annals of Mathematical Statistics*, **29**, 614.

Wu, C.F.J. (1986), "Jackknife, bootstrap and other resampling methods in regression analysis (with discussions)," *Annals of Statistics*, **14**, 1261–1350.