ISSN 1683–5603

International Journal of Statistical Sciences Vol. 9(Special Issue), 2009, pp 59-70 © 2009 Dept. of Statistics, Univ. of Rajshahi, Bangladesh

Jackknife and Bootstrap Methods for Variance Estimation from Sample Survey Data

J. N. K. Rao School of Mathematics and Statistics

Carleton University Ottawa, K1S 5B6, Canada

[Received January 1, 2009; Accepted March 30, 2009]

Abstract

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

Keywords and Phrases: Bootstrap, Conditional Properties, Confidence Intervals, Jackknife, MSE Estimation, Small Area Models.

AMS Classification: Primary 62D05; Secondary 62G09.

1 Re-sampling Variance Estimation

The importance of measurement errors in sample surveys was recognized as early as the 1940's. Mahalanobis (1946) developed the technique of interpenetrating sub-samples (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 sub-samples according to the same sampling design such that each sub-sample provides a valid estimate of the finite population total or mean. By assigning the sub-samples 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 $y_1, ..., y_n$, 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 $G_1, ..., G_g$, each of size m, assuming that n = gm. The groups, G_j , are deleted in turn and the "delete-group" estimates $\hat{\theta}_{(j)}, j = 1, ..., g$, are computed, where $\hat{\theta}_{(j)}$ denotes the estimator of θ based on the sample of size n - m = g - m after deleting G_j . Quenouille (1956) showed that the estimator

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

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

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

where $\hat{\theta}_{(.)} = g^{-1} \sum \hat{\theta}_{(j)}$. In the sample survey context, Durbin (1959) applied Quenouille's method to ratio estimation, using g = 2 groups. Rao (1965) 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 $\hat{\theta} = \bar{y}$, 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

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

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

as the "jackknife" variance estimator of $\hat{\theta}_Q$ or $\hat{\theta}$. Note that the implementation of $v_J(\hat{\theta}_Q)$ is computer-intensive if $\hat{\theta}$ requires iterative calculation, because *n* sets of iterative calculation need to be performed to calculate $\hat{\theta}_{(j)}$, j = 1, ..., 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.

Rao: Jackknife and Bootstrap Methods for Variance Estimation

Wu (1986) studied the linear regression model $y_i = x_i^T \beta + \varepsilon_i$, 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(.). Under the weighted jackknife method, proposed by Wu (1986), pairs (y_i, x_j) are deleted in turn for $j = 1, \dots, n$ 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

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

where $w_j = x_j^T (\sum_{i=1}^n x_i x_i^T)^{-1} x_j$. 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 re-sampling was first introduced by Efron (1979). Efron's pioneering 1979 paper on the bootstrap 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, remains 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 under stratified multi-stage cluster sampling is due to McCarthy (1969) and Kish and Frankel (1974). Krewski and Rao (1981) provide 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 re-samples, 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 $\hat{\theta}, \hat{\theta}_1, ..., \hat{\theta}_B$ from the data file, using the full sample weights and the B bootstrap weights. The bootstrap variance estimator of $\hat{\theta}$ is then simply obtained as

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

Bootstrap with B = 500 is currently used in Statistics Canada for variance estimation. Shao (2003) and Lahiri (2003) provide nice accounts of the impact of the bootstrap in survey sampling.

2 Small Area Estimation

Traditional direct estimators for large areas or domains, based only on area-specific data, are not reliable for small areas due to small sample sizes. As a result, 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. Here, we focus on a basic area-level model, called the Fay-Herriot (FH) model. Let $\theta_i = g(Y_i)$ be a suitable function of the small area total Y_i linearly related to area-level predictor variables Z_i , $i = 1, \dots, m$. The linking model is given by

$$\theta_i = z_i^T \beta + v_i, \ i = 1, \cdots, m,$$

where the v_i are independent and identically distributed (IID) as $N(0, \sigma_v^2)$. A matching sampling model is of the form

$$\hat{\theta}_i = g(\hat{Y}_i) = \theta_i + e_i, \quad i = 1, \cdots, m,$$

where the e_i are independently distributed as $N(0, \psi_i)$ with known sampling variance ψ_i , and \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(.) is non-linear. However, we focus here 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 $\hat{\theta}_i^B = E(\theta_i | \hat{\theta}_i, \beta, \sigma_v^2) \equiv h(\hat{\theta}_i, \beta, \sigma_v^2)$. We estimate the model parameters β and σ_v^2 by a suitable method, such as maximum likelihood (ML), residual maximum likelihood (REML) or the FH method of moments. Here we focus on REML estimators of β and σ_v^2 . Substituting the estimators β and σ_v^2 in $\hat{\theta}_i^B$, we get the empirical best (EB) estimator: $\hat{\theta}_i^{EB} \equiv h(\hat{\theta}_i, \hat{\beta}, \hat{\sigma}_v^2) = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) z_i^T \hat{\beta}$ 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 $\hat{\boldsymbol{\theta}}_i^{EB}$ may be written as

$$MSE(\hat{\theta}_i^{EB}) = E(\hat{\theta}_i^{EB} - \theta_i)^2$$

$$= E(\hat{\theta}_i^B - \theta_i)^2 + E(\hat{\theta}_i^{EB} - \theta_i)^2$$

$$\equiv M_{1i}(\sigma_v^2) + M_{2i}(\sigma_v^2). \qquad (2.1)$$

For the FH model, the leading term in (2.1) 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 $MSE(\hat{\theta}_i) = E(\hat{\theta}_i - \theta_i)^2 = \psi_i$. 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 $MSE(\hat{\theta}_i^{EB})$ are of order $O(m^{-2})$.

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

$$mse(\hat{\theta}_{i}^{EB}) = g_{1i}(\hat{\sigma}_{v}^{2}) + g_{2i}(\hat{\sigma}_{v}^{2}) + 2g_{3i}(\hat{\sigma}_{v}^{2})$$
(2.2)

The bias of (2.2) is of lower order than m^{-1} for large m. Note that the MSE estimator (2.2) is not area-specific in the sense that it does not depend on $\hat{\theta}_i$. Alternatives to the term $2g_{3i}(\hat{\sigma}_v^2)$ in (2.2) that make use of $\hat{\theta}_i$ have been proposed (Rao, 2003, section 7.1.5).

If $\theta_i = g(Y_i)$, then the best estimator of Y_i , $E(Y_i|\hat{Y}_i, \beta, \sigma_v^2) \equiv h(\hat{Y}_i, \beta, \sigma_v^2)$ has no closed form expression. As a result, MSE estimation using Taylor linearization becomes complex or difficult. In Section 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 for Small Areas

Jiang, Lahiri and Wan (2002) proposed a jackknife estimator of $MSE(\hat{\theta}_i^{EB})$ that avoids the explicit evaluation of $g_{2i}(.)$ and $g_{3i}(.)$ terms in (2.2), but it still requires the derivation of $g_{1i}(.)$ term which is simple for the EB estimator $\hat{\theta}_i^{EB}$ 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 $\phi; u = 1, \dots, m$. Then, the Jiang, Lahiri and Wan (JLW) un-weighted jackknife estimators of $M_{2i}(\sigma_v^2)$ is given by

$$M_{2i,J} = \frac{m-1}{m} \sum_{u=1}^{m} \{h(\hat{\theta}_i, \hat{\phi}_{(u)}) - h(\hat{\theta}_i, \hat{\phi})\}^2$$
(3.1)

Quenouille's bias reduction method is applied to $M_{1i}(\hat{\sigma}_v^2)$ in (2.1) to get

$$\hat{M}_{1i,J} = g_{1i}(\hat{\sigma}_v^2) - \frac{m-1}{m} \sum_{u=1}^m \{g_{1i}(\hat{\sigma}_{v(u)}^2) - g_{1i}(\hat{\sigma}_v^2)\}^2$$
(3.2)

JLW proved that $\hat{M}_{i,J} = \hat{M}_{1i,J} + \hat{M}_{2i,J}$ is a nearly unbiased estimator of $MSE(\hat{\theta}_i^{EB})$ 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 $w_u = 1 - (z_u^T/\psi_u)(\sum z_i z_i^T/\psi_i)^{-1} z_u$: replace (m-1)/m in (3.1) and (3.2) by $w_u(u = 1, \dots, m)$ and take it inside the summation terms. Note that $\hat{M}_{2i,J}$ and its weighted version are areaspecific 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 (3.2), can take negative values under certain scenarios. Chen and Lahiri (2005) used jackknife linearization, under the REML estimator $\hat{\sigma}_v^2$, to get a weighted version

$$\hat{M}_{i,JL} = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + \frac{\psi_i^2}{(\hat{\sigma}_v^2 + \psi_i)^3} v_{wJ}(\hat{\sigma}_v^2) + \frac{\psi_i^2}{(\hat{\sigma}_v^2 + \psi_i)^4} (\hat{\theta} - z_i^T \hat{\beta})^2 v_{wJ}(\hat{\sigma}_v^2)$$
(3.3)

where $v_{wJ}(\hat{\sigma}_v^2) = \sum_{u=1}^m w_u(\hat{\sigma}_{v(u)}^2 - \hat{\sigma}_v^2)^2$ is a weighted jackknife variance estimator of $\hat{\sigma}_v^2$. The estimator (3.3) is always non-negative, unlike $\hat{M}_{i,J}$ or its weighted version, but requires additional analytical work as in the case of (2.2). A simulation study indicated superior performance of the proposed jackknife linearization MSE estimator (3.3).

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 small area parameter of interest, given the model parameters ϕ . But it may not have a close 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 (2009) 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 $y_i \sim_{iid} B(n_i, p_i)$, given p_i and $p_i \sim_{iid} Beta(\alpha, \beta)$, $i = 1, \dots, m$, and the parameter of interest is p_i . In this case, the best estimator of p_i is $\hat{p}_i^B = E(p_i|y_i, \phi_i) \equiv h(y_i, \phi)$ and the EB estimator is $\hat{p}_i^{EB} = h(y_i, \hat{\phi})$, where $\hat{\phi} = (\hat{\alpha}, \hat{\beta})$ is a consistent estimator of $\phi = (\alpha, \beta)$. We have

$$MSE(\hat{p}_i^{EB}) = EV(p_i|y_i, \phi) + E(\hat{p}_i^{EB} - \hat{p}_i^B)^2 \equiv M_{1i} + M_{2i}$$
(3.4)

JLW need M_{1i} in (3.4) as function of ϕ to get their bias correction estimator $\hat{M}_{1i,J}$ which is not area-specific. Area specific estimator, $\hat{M}_{2i,J}$, of M_{2i} is given by (3.1) 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, section 9.4), Lohr and Rao (2009) applied a jackknife bias correction to $\tilde{g}_{1i}(y_i, \hat{\phi})$ to get the following estimator of M_{1i} :

$$\tilde{M}_{1i}(y_i) = \tilde{g}_{1i}(y_i, \hat{\phi}) - \sum_{j \neq i=1}^m \{ \tilde{g}_{1i}(y_i, \hat{\phi}_{(u)}) - \tilde{g}_{1i}(y_i, \hat{\phi}) \}^2.$$
(3.5)

The JLW estimator $\hat{M}_{2i,J}$ is used for M_{2i} in (3.4). The Lohr-Rao (LR) estimator $\tilde{M}_{i,J} = \tilde{M}_{1i}(y_i) + \hat{M}_{2i,J}$ is nearly conditionally unbiased given y_i , unlike the JLW estimator, and also nearly unbiased unconditionally as in the case of JLW, but it is less stable unconditionally than the JLW estimator. 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.

4 Bootstrap MSE and Interval Estimation

4.1 MSE Estimation

Parametric bootstrap versions of the JLW jackknife MSE estimator, $\hat{M}_{i,J}$, have been proposed by Butar and Lahiri (2003) and Pfeffermann and Glickman (2004). For the FH model under normality, *B* parametric bootstrap samples $(\hat{\theta}_i^b, z_i)$; $i = 1, \dots, m, b = 1, \dots, B$ 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 $\hat{\theta}_i^b = z_i^T \hat{\beta} + \hat{v}_i^b + \hat{e}_i^b \equiv \theta_i^b + \hat{e}_i^b$. 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)):

$$\hat{M}_{1i,B} = g_{1i}(\hat{\sigma}_v^2) - \frac{1}{B} \sum_{b=1}^{B} \{g_{1i}(\hat{\sigma}_v^2(b)) - g_{1i}(\hat{\sigma}_v^2)\}$$
$$= 2g_{1i}(\hat{\sigma}_v^2) - \frac{1}{B} \sum_{b=1}^{B} g_{1i}(\hat{\sigma}_v^2(b))$$
(4.1)

$$\hat{M}_{2i,B} = \frac{1}{B} \sum_{b=1}^{B} \{h(\hat{\theta}_i, \hat{\phi}(b)) - h(\hat{\theta}_i, \hat{\phi})\}^2,$$
(4.2)

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

$$\tilde{M}_{2i,B} = \frac{1}{B} \sum_{b=1}^{B} \{h(\hat{\theta}_i^b, \hat{\phi}(b)) - \theta_i^b\}^2.$$
(4.3)

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 (2006a) and Chatterjee and Lahiri (2007) 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 the $MSE(\hat{\theta}_i^{EB}) = E\{h(\hat{\theta}_i, \hat{\phi}) - \theta_i\}^2$ which suggests that a naïve estimator based on the (level 1) bootstrap samples, b, is given by $\tilde{M}_{2i,B}$ in (4.3). Next, we perform bootstrap bias correction of $\tilde{M}_{2i,B}$ using level 2 bootstrap samples. The c-th level 1 bootstrap sample is obtained by generating $v_i^b(c)$ and $v_i^b(c)$ independently from $N(0, \hat{\sigma}_v^2(c))$, and $N(0, \psi_i)$ and then letting $\hat{\theta}_i^b(c) = z_i^T \hat{\beta}(b) + \hat{v}_i^b(c) + \hat{e}_i^b(c) = \theta_i^b(c) + \hat{e}_i^b(c)$. 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

$$\tilde{M}_{2i,BC} = \frac{1}{BC} \sum_{b=1}^{B} \sum_{c=1}^{C} \left\{ h\{\hat{\theta}_{i}^{b}(c), \hat{\phi}(bc)\} - \theta_{i}^{b}(c) \right\}^{2}.$$
(4.4)

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

$$\tilde{M}_{i,BC} = 2\tilde{M}_{2i,B} - \tilde{M}_{2i,BC}$$
 (4.5)

The estimator $\tilde{M}_{i,BC}$ is nearly unbiased as B and C tend to infinity.

Hall and Maiti (2006a) studied MSE estimation under a unit level nested error linear regression model $y_{ij} = x_{ij}^T \beta + v_i + e_{ij}$, $j = 1, \dots, i$; $i = 1, \dots, m$, 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 \bar{X}_i is known. Customary normality assumption on v_i and e_{ij} is thus relaxed. Hall and Maiti (2006a) 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 (4.3) is then bias-corrected using a double bootstrap with C level 2 bootstrap samples from each level 1 bootstrap sample using the same moment matching method. The resulting MSE estimator of the form (4.5) 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.

4.2 Interval Estimation

Normal approximation $(1-\alpha)$ -level confidence intervals on the small area parameter θ_i , based on $\hat{\theta}_i^{EB}$ and a nearly unbiased MSE estimator, $mse(\hat{\theta}_i^{EB})$ or a re-sampling estimator $\hat{M}_{i,J}$ or $\tilde{M}_{i,BC}$, are of the form $\{\hat{\theta}_i^{EB} - z_{\alpha/2}(mse)^{1/2}, \hat{\theta}_i^{EB} + z_{\alpha/2}(mse)^{1/2}\}$, where $z_{\alpha/2}$ is the upper $\alpha/2$ -point of a N(0,1) variable and mse denotes a second-order MSE estimator. However, the normal theory intervals are not second-order accurate, even under normality of v_i and e_i , in the sense that the error in the coverage probability is of the order $O(m^{-1})$ for large m. Chatterjee, Lahiri and Li (2008) used the parametric bootstrap samples $\{(\hat{\theta}_i^b, z_i); i =$ $1, \dots, m\}, b = 1, \dots, B$ and the leading term $g_{1i}(\hat{\sigma}_v^2) \equiv \hat{g}_{1i}$ of the MSE estimator (2.2) to determine quantiles t_{1i} and t_{2i} such that the resulting bootstrap-based interval $\{\hat{\theta}_i^{EB} - t_{1i}(g_{1i})^{1/2}, \hat{\theta}_i^{EB} + t_{2i}(g_{1i})^{1/2}\}$ has coverage error of order $m^{-3/2}$ so that the proposed interval is second-order accurate. Hall and Maiti (2006b) proposed a different bootstrap-based interval that is also second-order accurate, but it does not make use of the area-specific estimator $\hat{\theta}_i^{EB}$.

It would be more appealing to the practitioner to use a nearly unbiased MSE estimator in constructing confidence intervals on θ_i . However, replacing \hat{g}_{1i} by *mse* in the Chatterjee *et al.* (2008) bootstrap calibration method does not seem to yield second-order accurate intervals. Further research on accurate confidence interval estimation for small areas would be useful.

Acknowledgments

A preliminary version of this paper has appeared in the 2007 Proceedings of the Survey Research Section, American Statistical Association. This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. It is a great personal pleasure to contribute this paper to the special felicitation volume for Professor A. K. Md. Saleh. As a colleague of Dr. Saleh for the past 25 years at Carleton University, I have enjoyed collaborating with him on organizing conferences and talking to him on research topics of mutual interest.

References

- [1] Bell, W. (2001). Discussion of "Jackknife in the Fay-Herriot model with an example". *Proceedings of the Seminar on Funding Opportunity in Survey Research*, pp. 98-104.
- [2] Butar, F. B. and Lahiri, P. (2003). On measures of uncertainty of empirical Bayes small area estimators. In Model Selection, Model Diagnostics, Empirical Bayes and Hierarchical Bayes, Special Issue II. *Journal of Statistical Planning and Inference*, **112** (1-2), 63-76.
- [3] Chatterjee, S. and Lahiri, P. (2007). A simple computational method for estimating mean squared prediction error in general small-area models. *Proceedings of the Survey Research Methods Section*, American Statistical Association, Washington, DC.
- [4] Chatterjee, S., Lahiri, P. and Li, H. (2008). Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models. *Annals* of *Statistics*, 36, 1221-1245
- [5] Chen, S. and Lahiri, P. (2002). A weighted jackknife MSPE estimator in small area estimation. *Proceedings of the Survey Research Methods Section*, American Statistical Association, Washington, DC, pp. 473-477.
- [6] Chen, S. and Lahiri, P. (2005). On mean squared prediction error estimation in small area estimation problems. *Proceedings of the Survey Research Methods Section*, American Statistical Association, Washington, DC, pp. 473-477.
- [7] 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.
- [8] 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.
- [9] Deming, W. E. (1960). Sample Design in Business Research. Wiley, New York.
- [10] Durbin, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika*, **46**, 477-480.
- [11] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7, 1-26.
- [12] 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.
- [13] Hall, P. (1992). The Bootstrap and Edgeworth Expansion. Springer-Verlag, New York.

- [14] Hall, P. (2003). A short prehistory of the bootstrap. *Statistical Science*, 18, 158-167.
- [15] Hall, P. and Maiti, T. (2006a). Nonparametric estimation of mean squared prediction error in nested-error regression models. *Annals of Statistics*, 34, 1733-1750.
- [16] Hall, P. and Maiti, T. (2006b). On parametric bootstrap methods for small-area prediction. *Journal of the Royal Statistical Society*, series B, 68, 221-238.
- [17] 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.
- [18] Kish, L. and Frankel, M. R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society*, series B, **36**, 1-37.
- [19] 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.
- [20] Lahiri, P. (2003). On the impact of bootstrap in survey sampling and small area estimation. *Statistical Science*, 18, 199-210.
- [21] Lohr, S. and Rao, J. N. K. (2009). Jackknife estimation of mean squared error of small area predictors in nonlinear mixed models. *Biometrika*, 96, 457-468.
- [22] Mahalanobis, P. C. (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society*, **109**, 325-378.
- [23] McCarthy, P. J. (1969). Pseudo-replication. *Review of the International Statistical Institute*, 37, 239-264.
- [24] Miller, R. G. (1964). A trustworthy jackknife. *Annals of Mathematical Statistics*, **35**, 1594-1605.
- [25] Pfeffermann, D. and Glickman, H. (2004). Mean square error approximation in small are estimation by use of parametric and nonparametric bootstrap. *Proceedings of the Survey Research Methods Section*, American Statistical Association, Washington, DC.
- [26] Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of mean squared errors of small area estimators. *Journal of the American Statistical Association*, 85, 163-171.
- [27] Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43, 353-360.
- [28] Rao, J. N. K. (1965). A note on the estimation of ratios by Quenouille method. *Biometrika*, **52**, 647-649.
- [29] Rao, J. N. K. (2003). Small Area Estimation. Wiley, New York.

- [30] Rao, J. N. K. and Webster, J. T. (1966) On two methods for bias reduction. *Biometrika*, 53, 571-577.
- [31] 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.
- [32] 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.
- [33] Shao, J. and Tu, D. (1995). The Jackknife and Bootstrap. Springer, New York.
- [34] Sitter, R. R. (1992). A resampling procedure for complex survey data. *Journal of the American Statistical Association*, **87**, 755-765.
- [35] Tukey, J. W. (1958). Bias and confidence in note quite large samples. *Annals of Mathematical Statistics*, **29**, 614.
- [36] Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Annals of Statistics*, **14**, 1261-1350.