Catalogue no. 11-522-XIE

Statistics Canada International Symposium Series - Proceedings

Symposium 2003: Challenges in Survey Taking for the Next Decade

2003

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2003
Challenges in Survey Taking for the Next Decade

RESAMPLING METHODS FOR MSE ESTIMATION WITH NONLINEAR SMALL AREA MODELS

Sharon L. Lohr and J. N. K. Rao[1]

ABSTRACT

Small area estimation methods, used when the sample sizes in subpopulations of interest are too small for direct estimates to be reliable, have been applied to a wide variety of problems in sample surveys and biostatistics. Many models that are used for small area estimation give the approximate theoretical value of the mean squared error (MSE) for the characteristic of interest in each small area. Estimating the MSE, however, is more complicated, particularly when the model used is not a linear model. We develop a jackknife method proposed in Rao (2003) for estimating MSE's when generalized linear models or other nonlinear models are used for the response of interest, and demonstrate its performance in a simulation study. The jackknife method, unlike some other methods in current use, gives area-specific estimates of the MSE and reduces the amount of computation needed.

KEYWORDS: Beta-Binomial Model; Bootstrap; Jackknife; Mean Squared Error.

1. INTRODUCTION

Small area estimation methods are commonly used to make inferences for subpopulations in which the sample sizes are too small for direct estimates to be sufficiently reliable. Examples include estimating school-age poverty rates for each county, estimating diabetes prevalence among demographic subgroups, finding health insurance coverage rates for each state, or estimating leukemia survival rates at different hospitals. Typically, models are employed to "borrow strength" from other small areas using covariates such as census and administrative information. The recent book by Rao (2003) provides an overview of different methods and models currently used in small area estimation.

Suppose there are m subpopulations, or small areas, of interest. For area i, the characteristic of interest is considered to be θ_i. For example, θ_i might be the poverty rate for county i, the mean blood pressure for subpopulation i, or the proportion of persons with diabetes in demographic subgroup i.

In many commonly used small area estimation methods, a two-stage hierarchical model is adopted. In the first stage, the data vector from area i, $\mathbf{y}_i$, follows $\mathbf{y}_i \mid \theta_i \sim f_1(\mathbf{y}_i \mid \theta_i, \lambda)$. The second stage model relates the characteristic of interest θ_i to other areas and to covariates. A general form for such a model is $\theta_i \sim f_2(\mathbf{x}_i, v_i, \phi)$, where $\mathbf{x}_i$ is a vector of covariates for area i, ϕ is a vector of unknown parameters, and v_i is a random variable. The pairs $(\mathbf{y}_i, \theta_i)$ are assumed to be independent. Under this model, θ_i is a random quantity. We want to find the best predictor of θ_i and an estimator of its mean squared error.

As an example, consider the Fay-Herriot (1979) area level mixed model. For the simplest case, let $\bar{y}_i$ denote the estimator of the population mean in area i from the survey; for example, $\bar{y}_i$ might be the direct estimator of the poverty rate from the Current Population Survey. The first-stage model is $\bar{y}_i \mid \theta_i = \theta_i + e_i$, where $e_i \sim N(0, \psi_i)$ represents the sampling error from the survey. For the second stage, it is assumed that $\theta_i = \mathbf{x}_i^T\beta + v_i$ where $v_i \sim N(0,$

[1]Sharon L. Lohr, Department of Mathematics and Statistics, Arizona State University, Tempe, Arizona, USA 85287-1804 (sharon.lohr@asu.edu), J. N. K. Rao, School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada K1S 5B6 (jrao@math.carleton.ca)

σ^2) represents the error from the model relating the small area means. In this case, then, the vector of unknown parameters is $\phi = (\beta, \sigma^2)$. Under this setup, the joint distribution of $\bar{y}_i$ and θ_i is known. Consequently, the best predictor—i.e., the predictor with the smallest mean squared error—of θ_i if the values of β and σ^2 are known is the mean of the posterior distribution,

$$\hat{\theta}_i^B = E\left[\theta_i \mid \bar{y}_i, \beta, \sigma^2\right] = \gamma_i \bar{y}_i + (1-\gamma_i)\mathbf{x}_i^T \beta, \tag{1.1}$$

where $\gamma_i = \sigma^2/(\sigma^2 + \psi_i)$. The posterior variance of θ_i if the values of β and σ^2 are known is

$$V\left[\theta_i \mid \bar{y}_i, \beta, \sigma^2\right] = E\left[(\hat{\theta}_i^B - \theta_i)^2 \mid \beta, \sigma^2\right] = \gamma_i \psi_i. \tag{1.2}$$

The posterior variance in (1.2) is the MSE of $\hat{\theta}_i^B$ if β and σ^2 are known.

If β and σ^2 are unknown, then $\hat{\theta}_i^B$ cannot be used as a predictor of θ_i. Instead, the empirical best predictor $\hat{\theta}_i^{\mathrm{EB}} = \hat{\gamma}_i \bar{y}_i + (1-\hat{\gamma}_i)\mathbf{x}_i^T \hat{\beta}$ may be used, where $\hat{\gamma}_i = \hat{\sigma}^2/(\hat{\sigma}^2 + \psi_i)$ and the statistics $\hat{\beta}$ and $\hat{\sigma}^2$ are calculated from the data. The additional uncertainty introduced by estimating β and σ^2 rather than knowing their values gives $\hat{\theta}_i^{\mathrm{EB}}$ a larger MSE than $\hat{\theta}_i^B$, since

$$\begin{aligned} \mathrm{MSE}\left[\hat{\theta}_i^{EB}\right] &= E\left[(\hat{\theta}_i^{EB} - \theta_i)^2\right] \\ &= E\left[(\hat{\theta}_i^B - \theta_i)^2\right] + E\left[(\hat{\theta}_i^{EB} - \hat{\theta}_i^B)^2\right] \\ &= \quad \gamma_i \psi_i \quad + \quad O(1/m)\,. \end{aligned} \tag{1.3}$$

Thus, if the "naive" estimator $\hat{\gamma}_i \psi_i$ were used to estimate the MSE of $\hat{\theta}_i^{\mathrm{EB}}$, ignoring the error in estimating β and σ^2, the estimated MSE would be negatively biased.

Similar issues occur for models that are not linear. For example, the beta-binomial model is often used for binary data. In this model, the characteristic of interest θ_i is a proportion, often denoted p_i. The n_i observed values in area i from the survey, y_{ij}, are assumed to be independently generated from Bernoulli random variables with parameter p_i, so that for $y_{i+} = \sum_{j=1}^{n_i} y_{ij}$, we have $y_{i+} \mid p_i \sim B(n_i, p_i)$. The second-stage model linking the proportions p_i across areas is $p_i \sim Beta(\alpha, \beta)$, with $\alpha > 0$ and $\beta > 0$. As in the linear model, the pairs (y_{i+}, p_i) are assumed to be independent across areas. If the parameters α and β are known, then $p_i \mid y_{i+}, \alpha, \beta \sim \mathrm{Beta}(y_{i+} + \alpha, n_i - y_{i+} + \beta)$. Consequently, if α and β are known the best predictor of p_i is

$$\hat{p}_i^B = E\left[p_i \mid y_{i+}, \alpha, \beta\right] = \frac{y_{i+} + \alpha}{n_i + \alpha + \beta} \tag{1.4}$$

with

$$V\left[p_i \mid y_{i+}, \alpha, \beta\right] = E\left[(\hat{p}_i^B - p_i)^2 \mid y_{i+}, \alpha, \beta\right] = \frac{(y_{i+} + \alpha)(n_i - y_{i+} + \alpha)}{(n_i + \alpha + \beta + 1)(n_i + \alpha + \beta)^2}. \tag{1.5}$$

If α and β are unknown then the empirical best predictor of p_i, obtained by substituting estimators of α and β for the unknown true values, is $\hat{p}_i^{EB} = (y_i + \hat{\alpha})/(n_i + \hat{\alpha} + \hat{\beta})$. As in the case of the linear model, if the error in estimating α and β is ignored when estimating the MSE of $\hat{p}_i^{EB}$, the resulting estimator of the MSE will be negatively biased.

There is a difference in the expressions for the posterior variance for the linear model in (1.2) and for the nonlinear model in (1.5). For the linear model, the expression $V\left[\theta_i \mid \bar{y}_i, \beta, \sigma^2\right]$ does not depend on $\bar{y}_i$, and consequently $E\left[(\hat{\theta}_i^B - \theta_i)^2 \mid \beta, \sigma^2\right] = V\left[\theta_i \mid \bar{y}_i, \beta, \sigma^2\right]$. The conditional variance expression in (1.5), however, does depend on the value of y_{i+}, so that in general the conditional mean squared error CMSE= $E\left[(\hat{p}_i^B - p_i)^2 \mid y_{i+}, \alpha, \beta\right]$ is not the same as the unconditional mean squared error UMSE = $E\left\{E\left[(\hat{p}_i^B - p_i)^2 \mid y_{i+}, \alpha, \beta\right]\right\}$. An estimator that is unbiased for UMSE will not necessarily be unbiased for CMSE.

In this paper, we study the conditional and unconditional bias and stability of several estimators that have been proposed for estimating UMSE for small area models. In Section 2, we review the major approaches that may be taken to estimate the mean squared error in linear and nonlinear models. Section 3 presents and gives properties of a jackknife estimator of CMSE proposed in Rao (2003, p. 199). Section 4 presents results from a small simulation study, and Section 5 contains a discussion.

2. APPROACHES TO ESTIMATING THE MEAN SQUARED ERROR

The mean squared error (MSE) of $\hat{\theta}_i^{\mathrm{EB}}$ for the linear mixed model is given in (1.3). Several approaches have been taken for estimating the MSE. The simplest estimator, hereafter referred to as the "naive" estimator of the MSE, simply substitutes estimates for the unknown parameters in (1.2) and ignores the extra variability due to estimating those parameters. For the linear model, the naive estimator of the MSE is $\hat{\gamma}_i \psi_i$.

Prasad and Rao (1990) showed that the naive estimator is biased downwards and derived an analytical approach for estimating the MSE in linear mixed models. They estimated the MSE by

$$\mathrm{mse}_{\mathrm{PR}}\left[\hat{\theta}_i^{EB}\right] = \hat{\gamma}_i \psi_i + g_{2i} + 2g_{3i},$$

where the terms g_{2i} and $2g_{3i}$ account for the extra variability due to estimating β and σ^2. Prasad and Rao (1990) proved that under certain regularity conditions, $E\left[\mathrm{mse}(\hat{\theta}_i^{\mathrm{EB}})\right] = \mathrm{MSE}(\hat{\theta}_i^{\mathrm{EB}}) + o(1/m)$, so that the estimator of the MSE is approximately unbiased. Lahiri and Rao (1995) showed that the Prasad-Rao estimator of the MSE is approximately unbiased even when the assumption of normality on v_i is not met, provided that sufficient moments of v_i exist.

The Prasad-Rao analytic results have been extended to other situations such as generalized linear mixed models (see Rao, 2003, Section 5.6 for a review). For an analytic approach, however, results for each case must be derived separately, and then code must be written for each specific situation. Resampling methods avoid some of those problems, and various authors have proposed using the bootstrap and jackknife for estimating the MSE of small area estimators.

Pfeffermann and Tiller (2001) proposed a parametric bootstrap estimator of MSE[$\hat{\theta}_i^{\mathrm{EB}}$] in the context of a state-space model. They estimated the extra variability due to estimating the unknown parameter ϕ by generating bootstrap series from the model derived from the original data, and showed that under proposed regularity conditions the bias of the bootstrap method is $o(1/m)$.

Jiang, Lahiri and Wan (2002) derived properties of a jackknife estimator of the MSE in the general setting. They used the orthogonal decomposition

$$\begin{aligned} \mathrm{MSE}\left[\hat{\theta}_i^{EB}\right] &= E\left[(\hat{\theta}_i^{B}-\theta_i)^2\right] + E\left[(\hat{\theta}_i^{EB}-\hat{\theta}_i^{B})^2\right] \\ &= \quad M_{1i} \quad + \quad M_{2i} \end{aligned} \tag{2.1}$$

and used the jackknife to estimate M_{1i} and M_{2i} separately. The jackknife is employed as follows for the Fay-Herriot linear mixed model: in the jth jackknife iteration, area j is deleted. The data from the remaining $(m-1)$ areas are used to calculate $\hat{\beta}_{(-j)}$, $\hat{\sigma}^2_{(-j)}$, and $\hat{\gamma}_{i(-j)} = \hat{\sigma}^2_{(-j)}/(\hat{\sigma}^2_{(-j)} + \psi_i)$. These quantities are then used to calculate a delete-j estimator of θ_i as $\hat{\theta}^{EB}_{i(-j)} = \hat{\gamma}_{i(-j)}\bar{y}_i + (1-\hat{\gamma}_{i(-j)})\mathbf{x}_i^T\hat{\beta}_{(-j)}$. Jiang, Lahiri and Wan (2002) estimated M_{1i} by

$$\hat{M}_{1i} = \hat{\gamma}_i\psi_i - \frac{m-1}{m}\sum_{j=1}^{m}\left[\hat{\gamma}_{i(-j)}\psi_i - \hat{\gamma}_i\psi_i\right]. \tag{2.2}$$

The quantity $\hat{M}_{1i}$ in (2.2) equals the naive estimator $\hat{\gamma}_i\psi_i$ plus a correction for the bias. The second term in (2.1) is estimated by

$$\hat{M}_{2i} = \frac{m-1}{m}\sum_{j=1}^{m}\left[\hat{\theta}^{EB}_{i(-j)} - \hat{\theta}^{EB}_{i}\right]^2. \tag{2.3}$$

Jiang, Lahiri and Wan (2002) showed that the jackknife estimator of the MSE, $\mathrm{mse}_{\mathrm{JLW}} = \hat{M}_{1i} + \hat{M}_{2i}$, has bias of order $o(1/m)$ if regularity conditions are met.

The Jiang, Lahiri and Wan (2002) jackknife estimator of the MSE is defined similarly for nonlinear models. In general, if we let $k_i(\phi) = M_{1i} = E\left[(\hat{\theta}_i^{B}-\theta_i)^2\right]$, then

$$\hat{M}_{1i} = k_i(\hat{\phi}) - \frac{m-1}{m}\sum\nolimits_j\left[k_i(\hat{\phi}_{(-j)}) - k_i(\hat{\phi})\right] \tag{2.4}$$

and $\hat{M}_{2i}$ is as in (2.3). For the beta-binomial model described in Section 1,

$$k_i(\alpha,\beta) = E\left[\left(\hat{p}_i^{B} - p_i\right)^2 \mid \alpha,\beta\right] = \frac{\alpha}{(n_i+\alpha+\beta+1)(n_i+\alpha+\beta)^2}\left\{n_i+\beta+\frac{n_i(n_i-1)\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}+\frac{n_i\alpha(\beta-\alpha)}{\alpha+\beta}\right\}.$$

For the beta-binomial model, $k_i(\phi)$, with $\phi = (\alpha,\beta)$, can be calculated analytically. For most other nonlinear situations, including most generalized linear mixed models, calculating $k_i(\phi)$ requires numerical integration, and that numerical integration must be performed $(m+1)$ times to estimate the M_{1i} term.

3. AREA-SPECIFIC JACKKNIFE

The MSE estimators described in Section 2 estimate the unconditional MSE, $\mathrm{UMSE} = E\left[(\hat{\theta}_i^{EB}-\theta_i)^2\right]$. For linear models, the unconditional and conditional MSE's coincide. For nonlinear models, however, the conditional MSE (CMSE) differs from the UMSE. In some situations, it may be preferable to estimate the CMSE. Booth and Hobert (1998) argued that if the pairs ($\mathbf{y}_i$, θ_i) are independent across areas and if ϕ is known, then inference about θ_i should be based on the conditional distribution of $\theta_i \mid \mathbf{y}_i, \phi$. They stated: "only the observations from the ith location are

relevant for predicting [θ_i] when the parameters are known" (p. 265). Fuller (1990) also presented a conditional MSE estimator for prediction of true values from the measurement error model. We are not using hierarchical Bayesian methods in this paper, but we note that from a Bayesian perspective, the CMSE is the posterior variance if the parameters ϕ are known. An estimator of CMSE for nonlinear mixed models is area-specific in that it depends on the data observed in area *i*.

In this section, we explore a jackknife estimator of the MSE that estimates the conditional (and hence also the unconditional) MSE, proposed briefly in Rao (2003, Section 9.4). The area-specific jackknife estimator of the MSE is

$$\text{mse}_{AS}(\hat{\theta}_i^{EB}) = \hat{M}_{A1i}(\mathbf{y}_i) + \hat{M}_{2i}, \tag{3.1}$$

where

$$\hat{M}_{A1i}(\mathbf{y}_i) = g_i(\hat{\phi}, \mathbf{y}_i) - \frac{m-1}{m}\sum_j [g_i(\hat{\phi}_{(-j)}, \mathbf{y}_i) - g_i(\hat{\phi}, \mathbf{y}_i)] \tag{3.2}$$

and $g_i(\phi, \mathbf{y}_i) = V(\theta_i \mid \mathbf{y}_i, \phi)$ is the posterior variance of θ_i given the data from area *i* and the values of ϕ. The quantity $\hat{M}_{2i}$ is defined in (2.3). Under appropriate regularity conditions, it may be shown that the area-specific jackknife in (3.1) has conditional bias of order $o_p(1/m)$, and that the bias for estimating the UMSE is of order $o(1/m)$.

For the linear Fay-Herriot model, the area-specific jackknife in (3.1) is identical to the Jiang, Lahiri, and Wan (2002) jackknife since in that case $g_i = k_i$. In nonlinear models, however, the area-specific jackknife has low bias for the CMSE and the UMSE while the Jiang, Lahiri, Wan (2002) jackknife only has low bias for the UMSE. In addition, the area-specific jackknife is much easier to compute, since it does not require finding the expected value of $g_i(\phi, \mathbf{y}_i)$.

The reduced computational complexity for the area-specific jackknife leads not only to less computation time but also possibly to fewer numerical errors. Few jackknife papers discuss numerical errors of the methods, but the cumulative effect of numerical errors can be large. The summands in (2.3), (2.4), and (3.2) are typically small, and are thus sensitive to numerical errors. Maximum likelihood estimation (MLE) is often used to calculate $\hat{\phi}$, and typically this is done using an iterative algorithm such as Newton-Raphson (see Press et al., 1992, for a discussion of some of the numerical inaccuracies that can affect the results). If the estimate $\hat{\phi}$ from the full data set has numerical errors, and those errors are propagated in the jackknifed quantities $\hat{\phi}_{(-j)}$, the jackknife correction in (3.2) can result in an estimate that is less accurate than the naive estimate. This is a particular danger if subroutines written by other researchers are used to calculate parameter estimates, since sometimes these subroutines specify iterating only until a tolerance of 0.001—or even larger value—is attained. For use with the jackknife, the estimate of the unknown parameters ϕ from the full data set should be calculated to as close to machine precision as possible.

With a numerically accurate estimate $\hat{\phi}$, the Newton-Raphson algorithm may be used to accelerate the computation of $\hat{\phi}_{(-j)}$. In the jackknife iterations, the MLE's from the full data may be used as initial values, and Newton-Raphson with numerical derivatives used to calculate $\hat{\phi}_{(-j)}$ for each jackknife iteration. Generally, convergence is achieved in one to four steps of Newton-Raphson, and convergence for $\hat{\phi}_{(-j)}$ is faster when *m* is larger.

In practice, $\hat{M}_{1i}$ in (2.4) or $\hat{M}_{A1i}(\mathbf{y}_i)$ in (3.2) may be negative. In that case, we recommend substituting $k_i(\hat{\phi})$ for $\hat{M}_{1i}$, and $g_i(\hat{\phi}, \mathbf{y}_i)$ for $\hat{M}_{A1i}(\mathbf{y}_i)$.

4. SIMULATION STUDY

We performed a small simulation study to compare the area-specific jackknife with the naive estimator of MSE and with the jackknife estimator proposed in Jiang, Lahiri, and Wan (2002). The beta-binomial model described in Section 1 was used to generate the data. The simulation study used a factorial design with three values of *m* (10, 30, and 60) and three values for each of α and β (0.1, 1.0, and 10.0). The within-area sample sizes took on the values $n_i \in \{1, 2, 3, 4, 5\}$; each value was used an equal number of times among the *m* small areas.

All computations were performed in R, version 1.6.1 (obtained from www.r-project.org). One thousand simulation runs were performed for each factor combination. For computational simplicity in the simulations, method of moments estimators were used for α and β. Let $\hat{p} = \sum y_{i+} / \sum n_i$ and $\hat{\sigma}^2 = [\sum y_{i+}(y_{i+} - 1)]/[\sum n_i (n_i - 1)] - \hat{p}^2$, then $\hat{\alpha} = \hat{p}[\hat{p}(1-\hat{p})/\hat{\sigma}^2 - 1]$ and $\hat{\beta} = (1-\hat{p})\hat{\alpha}/\hat{p}$ are method of moments estimators of α and β when both $\hat{\alpha}$ and $\hat{\beta}$ are defined and positive. For simulation runs in which $\hat{\alpha}$ or $\hat{\beta}$ given above was not positive or not defined, we set $\hat{\alpha}$ and $\hat{\beta}$ to be large values with $\hat{\alpha}/(\hat{\alpha} + \hat{\beta}) = \hat{p}$.

For each setting of the experimental factors, we estimated the true UMSE separately for each value of the small area sample size n_i as $\text{UMSE}(n_i) = (1/1000)\sum_{k=1}^{1000}(\hat{p}_{ik}^{EB} - p_{ik})^2$, where p_{ik} is the true value of p_i generated for the *k*th simulation run and $\hat{p}_{ik}^{EB}$ is the estimate of p_i obtained in that run. The CMSE(y_{i+}, n_i) was estimated similarly, by averaging the squared deviations over the simulation runs that had the specified value of y_{i+}. We then calculated the unconditional relative bias of an MSE estimator, mse, for each sample size as $100[\overline{\text{mse}}(n_i) - \text{UMSE}(n_i)]/\text{UMSE}(n_i)$, where $\overline{\text{mse}}(n_i)$ is the average of the MSE estimates for that sample size. The conditional relative bias was calculated similarly, using $100[\overline{\text{mse}}(y_{i+}, n_i) - \text{CMSE}(y_{i+}, n_i)]/\text{CMSE}(y_{i+}, n_i)$ with $\overline{\text{mse}}(y_{i+}, n_i)$ the average of the MSE estimates for runs with that value of y_{i+}. We calculated the coefficient of variation (CV) as (standard deviation of mse)/MSE.

Partial results, for $\alpha = \beta = 1$, are summarized in Tables 1 to 3. The entries in the tables represent the averages of the absolute values of relative bias (ARB) and of the CV's over all of the sample sizes. For the table rows reporting conditional values, to make comparisons easier, we first averaged over the values obtained from different y_{i+}'s.

Table 1. Results from simulation when *m* = 10.

	Naive estimator	Jiang et al. (2002)	Area-specific jackknife
Unconditional ARB	61 (all negative)	73 (most positive)	72 (most positive)
Unconditional CV	0.3	2.6	2.7
Conditional ARB	60 (all negative)	96 (most positive)	88 (most positive)
Conditional CV	0.3	2.3	2.2

Table 2. Results from simulation when *m* = 30.

	Naive estimator	Jiang et al. (2002)	Area-specific jackknife
Unconditional ARB	31.8 (all negative)	3.8	2.6
Unconditional CV	0.30	0.51	0.53
Conditional ARB	32.4 (all negative)	22.3	7.5
Conditional CV	0.24	0.53	0.57

Table 3. Results from simulation when $m = 60$.

	Naive estimator	Jiang et al. (2002)	Area-specific jackknife
Unconditional ARB	17.1 (all negative)	1.9	1.6
Unconditional CV	0.24	0.17	0.24
Conditional ARB	17.8 (all negative)	20.4	5.5
Conditional CV	0.16	0.16	0.21

Tables 1 to 3 demonstrate the theoretical results that the naive estimator of the MSE underestimates the true value. The naive estimate of the MSE was less than the true value of UMSE or CMSE for every value of n_i (and of y_{i+} for the CMSE), so that every value of the relative bias was negative. For $m = 10$ small areas, the two jackknife methods also had large bias, but most of the values overestimated the UMSE or CMSE. This overestimation occurred largely because of the instability of the method of moments estimator with small sample sizes. With only ten areas, sometimes deleting one of the areas resulted in a big change in the estimates of α and β, particularly when one of the moments estimates was nonpositive for a jackknife iteration. These large changes led to an overly large estimate of the MSE.

For $m = 30$ or 60, a different pattern emerges. The naive estimator still underestimates the true value of the UMSE and CMSE for every situation studied. Both the Jiang, Lahiri, and Wan (2002) jackknife and the area-specific jackknife have small bias for estimating UMSE, as expected. The two jackknifes perform differently for estimating the CMSE, however. The area-specific jackknife has small conditional bias, and it decreases with increasing m. On the other hand, the Jiang, Lahiri, Wan (2002) jackknife has a strong pattern for the conditional bias—the conditional bias is large and positive when y_{i+}/n_i is small or large, and the conditional bias is large and negative when y_{i+}/n_i is close to 0.5. The conditional bias of the Jiang, Lahiri, and Wan (2002) method, moreover, does not decrease with increasing m.

The area-specific jackknife estimator has small unconditional and conditional bias, but the lower conditional bias is achieved at some cost of stability of the MSE estimator. The Jiang, Lahiri, and Wan (2002) jackknife has somewhat lower CV when there are a large number of areas, although even when $m = 60$ the difference in CV between the two jackknife methods is not large. The area-specific jackknife is much easier to calculate even for the beta-binomial situation when the extra integral $E[V(\theta_i \mid \mathbf{y}_i, \phi)]$ can be evaluated analytically.

5. DISCUSSION

In this paper, we have discussed and compared three methods for estimating the MSE in linear and nonlinear small area models—a naive estimator, the jackknife method proposed by Jiang, Lahiri, and Wan (2002), and an area-specific jackknife. The two jackknife methods account for the extra variability due to not knowing the values of unknown parameters such as β and σ^2 in the linear model or α and β in the beta-binomial model. While both jackknife methods result in more accuracy for estimating the unconditional MSE than achieved by the naive estimator, the area-specific jackknife has several desirable properties. It has low conditional bias as well as low unconditional bias, and thus provides an area-specific estimator of the MSE. The area-specific jackknife also has fewer computational steps, and is thus less susceptible to numerical errors.

The area-specific jackknife is a promising estimator for the MSE in small area estimation problems. With the simpler computations, it can also be adapted to problems such as estimating the MSE of other quantities such as the difference between two subpopulation characteristics $\theta_i - \theta_j$.

ACKNOWLEDGEMENTS

This research was partially supported by Grant No. 0105852 from the U.S. National Science Foundation and by a grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Booth, J. and J. Hobert (1998), "Standard Errors of Prediction in Generalized Linear Mixed Models", *Journal of the American Statistical Association,* 93, pp. 262-272.

Fay, R. E. and R. A. Herriot (1979), "Estimates of Income for Small Places: An Empirical Bayes Application of James-Stein Procedures to Census Data", *Journal of the American Statistical Association*, 78, pp. 269-277.

Fuller, W. A. (1990), "Prediction of True Values for the Measurement Error Model", in P. J. Brown and W. A. Fuller (eds.), *Statistical Analysis of Measurement Error Models and Applications, Contemporary Mathematics Vol. 112,* Providence, RI: AMS, pp. 41-57.

Jiang, J., P. Lahiri and S.-M. Wan (2002), "A Unified Jackknife Theory for Empirical Best Prediction with *M*-Estimation", *Annals of Statistics*, 30, pp. 1782-1810.

Lahiri, P. and J. N. K. Rao (1995), "Robust Estimation of Mean Squared Error of Small Area Estimators", *Journal of the American Statistical Association,* 82, pp. 758-766.

Pfeffermann, D. and R. B. Tiller (2001), "Bootstrap Approximation to Prediction MSE for State-Space Models with Estimated Parameters", Technical report, Jerusalem Israel: Hebrew University.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992), *Numerical Recipes in C, 2nd ed.,* Cambridge: Cambridge University Press.

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

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