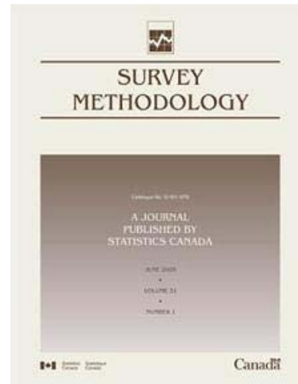
Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

A conditional mean squared error of small area estimators



by Louis-Paul Rivest and Eve Belmonte

June 2000

Survey Methodology, June 2000
Vol. 26, No. 1, pp. 67-78
Statistics Canada, Catalogue No. 12-001

A conditional mean squared error of small area estimators

Louis-Paul Rivest and Eve Belmonte [1]

Abstract

This paper suggests estimating the conditional mean squared error of small area estimators to evaluate their accuracy. This mean squared error is conditional in the sense that it measures the variability with respect to the sampling design for a particular realization of the smoothing model underlying the small area estimators. An unbiased estimator for the conditional mean squared error is easily constructed using Stein's Lemma for the expectation of normal random variables. This estimator can be calculated for any shrinking strategy; composite and empirical Bayes estimators are considered in this work. It can be calculated when the small area estimators are benchmarked to coincide with direct estimators at high level of aggregation. It can accommodate skewness in the data and estimated variances. The conditional mean squared error estimator does not rely on any smoothing model. The price to pay for this property is a high variance; the new estimator is unstable under heavy shrinking. In these situations, it still provides useful diagnostic information about the shrinking model. It can also be seen as a building block for estimators of unconditional mean squared errors such as Prasad and Rao's (1990). Examples dealing with the estimation of the under-coverage in the Canadian Census illustrate the application of this new estimator.

Key Words: Census under-coverage; Diagnostics; Empirical Bayes estimation; Estimated variances; Skewness; Stein's lemma; Survey sampling.

1. Introduction

In survey sampling, the need to develop accurate methods of estimation for small areas poses challenging statistical problems. For small areas, direct survey estimates have too large a variance to be reliable. Small area techniques "improve" direct estimates by shrinking them towards model based smoothed values. Simple shrinking estimators are proposed by Purcell and Kish (1979). In a pioneering paper, Fay and Herriot (1979) demonstrate that this can lead to interesting gains in precision. The review papers of Ghosh and Rao (1994) and of Singh, Gambino and Mantel (1994) provide convincing evidence of the vitality of this area.

The estimation of the errors in small area estimation is receiving an increasing attention, see Singh, Stukel and Pfeffermann (1998) and Booth and Hobert (1998). This paper suggests estimating the conditional mean squared errors of small area estimators. The conditional mean squared error can be estimated for all shrinking strategies, either empirical Bayes or decision theoretic (Purcell and Kish 1979). Other mean squared errors, such as Prasad and Rao's (1990), and Singh, Stukel and Pfeffermann (1998) frequentist proposals measure the variability with respect to both, the sampling design and the smoothing model. The mean squared error of this paper is conditional in the sense that it measures variability with respect to the sampling design for a particular realization of the smoothing model. This feature is attractive since the conditional estimator reflects the conditions under which the survey has been carried out (see Särndal, Swensson, and Wretman 1992, chapter 7). The drawback of this property is a high variability. In some instances, the proposed estimator is too variable for practical use.

When shrinking is important, the conditional mean squared error estimators are highly unstable. An unconditional assessment of the precision of small area estimators must be used. In this situation, the conditional estimator proposed in this paper still provides some useful information. It can be looked at as a diagnostic for comparing smoothing models. It can also be a building block for constructing Monte Carlo estimates of unconditional mean squared errors in situations where closed form formulas, such as Prasad and Rao's (1990), are not available.

The assessment of the accuracy of estimators for the under-coverage, at the provincial and sub-provincial levels, of the Canadian Census motivated this work. Alternatives to the direct estimates for provincial under-coverage are discussed by Royce (1992) andRivest (1995). Dick (1995) applies empirical Bayes methods to sub-provincial under-coverage estimates. These two examples are treated in section 5.

An estimator of the conditional mean squared error is presented in section 2. Its construction relies on the multivariate version of Stein's Lemma for the expectation of normal deviates. Section 3 suggests changes to the conditional estimator to accommodate skewness in the distribution of the direct estimators and estimated variances. Section 4 discusses the application of the new estimator to empirical Bayes estimators. Its relationship with Prasad and Rao (1990) prediction variance is highlighted. Examples are treated in section 5.

1. Louis-Paul Rivest and Eve Belmonte, Département de mathématiques et de statistique, Université de Laval, Ste-Foy, Québec, Canada, G1K 7P4.

2. A conditional mean squared error estimator

Suppose that there are n small areas and let $\mu = (\mu_1, ..., \mu_n)^t$ denote the unknown population characteristics for these small areas. The direct survey estimates for the n small areas are $y = (y_1, ..., y_n)^t$ where the distribution of y is $N_n(\mu, \Sigma)$, a n-variate normal distribution with mean vector μ and known variance-covariance matrix Σ. As pointed out by Ghosh and Rao (1994), the normality assumption is likely to hold for many surveys since direct survey estimates are usually functions of sums of variables. The $n \times n$ matrix Σ is a design based measure of precision for y. For the time being, this matrix is assumed to be known. This assumption is relaxed in section 3.2. The uncertainty in y comes from the random selection of the sampling units. Subscript S, for sampling design, denotes expectations taken with respect to the distribution of y.

In a typical application of small area techniques, one has,

$$y_i = \frac{\sum_j w_{ij} y_{ij}}{\sum_j w_{ij}}$$

where y_{ij} is the y-value for the j^{th} sample unit in small area i, w_{ij} is its sampling weight and the sum is over all the sample units in small area i. In many instances, the variance covariance matrix Σ is diagonal; its (i, i) term, is $\sigma_{ii} = \text{Var}_S(y_i)$; when they are non null, the off diagonal elements of Σ are denoted by σ_{ij}, $i, j = 1, ..., n$.

Several methods have been proposed to improve the accuracy of direct survey estimators. They involve shrinking y_i towards some indirect estimator of μ_i. The resulting estimators can be written as

$$\hat{\mu}_i = y_i + g_i(y_1, ..., y_n), \; i = 1, ..., n \tag{1}$$

where the g_i's are functions depending on the shrinking strategy.

In vector form, one can write (1) as $\hat{\mu} = y + g(y)$ where g, whose i^{th} component is equal to g_i, is a function defined from R^n to R^n. We assume that for each i, the right partial derivative and the left partial derivative of g_i with respect to y_j exists for any y in R^n. When they are equal, $\partial g_i(y)/\partial y_j$ denotes the common value; if they differ $\partial g_i(y)/\partial y_j$ is the average between the two values. The component of $g(y)$ and their partial derivatives are assumed to have finite variances. A conditional assessment of the precision of $\hat{\mu}$ as an estimator for μ is given by the matrix of the mean product errors which is given by

$$E_S\{(\hat{\mu}-\mu)(\hat{\mu}-\mu)^t\} = \Sigma + E_S\{(y-\mu)\,g(y)^t\} + E_S\{g(y)(y-\mu)^t\} + E_S\{g(y)g(y)^t\}.$$

On the right hand side of this equality, the only quantities for which there are no obvious estimators are $E_S\{(y-\mu)g(y)^t\}$ and $E_S\{g(y)(y-\mu)^t\}$. Their evaluations are eased by the following result which is a multivariate extension of Stein's lemma (Stein 1981). Its proof is given in the appendix together with the proofs for Propositions 2, 3, and 4.

PROPOSITION 1: Let y be a $N_n(\mu, \Sigma)$ random vector then,

$$E_S\{(y-\mu)\,g(y)^t\} = \Sigma\, E_S\{\nabla g(y)\},$$

where $\nabla g(y)$ is an $n \times n$ matrix whose $(i, j)^{\text{th}}$ element is given by $g_i^j(y) = \partial g_i(y)/\partial y_j$.

Now according to Proposition 1, $\Sigma \nabla g(y)$ is an unbiased estimator for $E_S\{(y-\mu)g(y)^t\}$. Thus the conditional estimator (index "c" stands for conditional) for the matrix of the mean product errors is given by

$$\text{mpe}_c(\hat{\mu}) = \Sigma + \Sigma \nabla g(y) + \nabla g(y)^t \Sigma + g(y)g(y)^t. \tag{2}$$

The diagonal terms of (2) can be negative. Since they estimate mean squared errors, a better estimator for the mean squared error of $\hat{\mu}_i$ is

$$\text{mse}_c^+(\hat{\mu}_i) = \max\big(0, \sigma_{ii} + \sum_j \sigma_{ij}\{g_j^i(y) + g_i^j(y)\} + g_i(y)^2\big).$$

It generalizes an estimator proposed by Bilodeau and Srivastava (1988) for James-Stein estimator, and by Robert (1992 page 292) for empirical Bayes estimators. When the y_i's are independent, with $\sigma_{ij} = 0$ when $i \neq j$, then

$$\text{mse}_c(\hat{\mu}_i) = \sigma_{ii} + 2\sigma_{ii}\frac{\partial g_i(y)}{\partial y_i} + g_i(y)^2, \tag{3}$$

and $\text{mse}_c^+(\hat{\mu}_i) = \max\{\text{mse}_c(\hat{\mu}_i), 0\}$.

Kott's (1989) small area estimator has $g_i(y) = \hat{\alpha}_i(\hat{\gamma}_i - y_i)$, where $\hat{\gamma}_i$ is a measure of location for the y's and $\hat{\alpha}_i$ is a smoothing parameter. These two statistics involve variance estimates calculated at the "unit" level, that is using the y_{ij}'s Kott's (1989) conditional mean squared error is

$$v(\hat{\mu}_i) = \sigma_{ii}(1 - 2\hat{\alpha}) + (\hat{\alpha}_i(y_i - \hat{\gamma}_i))^2.$$

This is equal to (3) when both $(d/dy_i)\hat{\alpha}_i$ and $(d/dy_i)\hat{\gamma}_i$ are null. Thus Kott's (1989) estimator for the conditional mean squared error does not account for the estimation for the variance components. This may account for the biases that it exhibited in the simulations reported by Prasad and Rao (1999).

The estimates mse_c and mpe_c can be evaluated numerically by taking

$$\frac{\partial g_i(y)}{\partial y_j} = \frac{g_i(y_1, ..., y_{j-1}, y_j + \epsilon, y_{j+1}, ..., y_n) - g_i(y_1, ..., y_{j-1}, y_j - \epsilon, y_{j+1}, ..., y_n)}{2\epsilon}$$

where ϵ is a small positive number. Thus mse_c and mpe_c can be calculated in all circumstances, even when g has no explicit form.

To illustrate the flexibility of the conditional estimator, consider $\hat{\mu}^* = \hat{\mu}(\sum y_i)/(\sum \hat{\mu}_i)$, an estimator bench-marked to agree with the direct estimator for the y-total. One has $\hat{\mu}^* = y + g^*(y)$ where

$$g^*(y) = \frac{\sum y_i}{\sum \hat{\mu}_i} g(y) + \left(\frac{\sum y_i}{\sum \hat{\mu}_i} - 1\right) y.$$

It might be difficult to derive an analytical formula for $\mathrm{mpe_c}(\hat{\mu}^*)$, however this expression is easily evaluated using numerical derivates. Modifications of the conditional estimator to account for non-normality in the y_i's and for estimated variances σ_{ii} are given next.

3. Sensitivity analysis

In many surveys, especially those in the business sector, the study variables are skewed. Some of this skewness might still be left in the direct estimators y_i. This section suggest a correction to the conditional mean squared error to account for skewness in the distribution of y. It also proposes ways to account for the estimation of the variances σ_{ii} in the mean squared error calculations.

In practice the variances σ_{ii} are estimated. Several authors (Dick 1995; Hogan 1992) smooth the variances before calculating the small area estimates. They then consider the smoothed variances as the true variances in the small area calculations. Section 3.2 gives a condition under which replacing the estimated variances by their smoothed values yields unbiased mean squared error estimators. It also consider situations where the sampling variances are estimated with random groups (Wolter 1985 chapter 2). This method consists in carrying a certain number, say k, of replications of the survey design. This yields, for each i, k estimates of μ_i; $\hat{\sigma}_{ii}$ is then equal to the sampling variance of these k estimates divided by k. Assuming that these k estimates are normally distributed, one can consider that, suitably normalized, the distribution of $\hat{\sigma}_{ii}$ is chi-squared with $k-1$ degrees of freedom. A conditional mean squared error, adjusted for variances estimated with random groups, is proposed in this section. To keep the discussion simple, we assume in this section that Σ is a diagonal matrix; in other words the y_i's are assumed to be independent random variables.

3.1 Non-normality in the distribution of y_i

In many applications of small area estimation, the distributions of the y_i's are not exactly normal. A simple adjustment to (3) is proposed to deal with asymmetry in the distribution of the y_i's.

Suppose that the skewness of y_i, $\rho_i = E_S\{(y_i - \mu_i)^3\}/\sigma_{ii}^{3/2}$ is small and non-zero. A first order Edgeworth series for the distribution of y_i is given by (see for instance Reid 1991):

$$f(t) = \frac{\exp\{-(t-\mu_i)^2/(2\sigma_{ii})\}}{\sqrt{(2\sigma_{ii}\pi)}} \times \left[1 + \frac{\rho_i}{6}\left\{\left(\frac{t-\mu_i}{\sqrt{\sigma_{ii}}}\right)^3 - 3\left(\frac{t-\mu_i}{\sqrt{\sigma_{ii}}}\right)\right\}\right].$$

Such an expansion is used to correct for skewness in the direct estimators (Barndorff-Nielsen and Cox 1989, remark 2 page 92). Expansions involving additional terms are used for correcting for both skewness and kurtosis; they will not be considered in this section. The evaluation of $E\{(y_i - \mu_i) g_i(y)\}$ under f, needed for the construction of the conditional mean squared error estimator, is given in Proposition 2.

PROPOSITION 2: When y_i distributed according to $f(t)$, as ρ_i tends to 0.

$$E_S\{(y_i - \mu_i) g_i(y)\} = \sigma_{ii} E_S\left\{\frac{\partial g_i(y)}{\partial y_i}\right\} + \frac{\sigma_{ii}^{3/2}\rho_i}{2} E_S\left\{\frac{\partial^2 g_i(y)}{\partial y_i^2}\right\} + O(\rho_i).$$

A mean squared error estimator corrected for asymmetry is therefore given by $\mathrm{mse}_c^+(\hat{\mu}_i) = \max\{0, \mathrm{mse_c}(\hat{\mu}_i)\}$ where

$$\mathrm{mse_c}(\hat{\mu}_i) = \sigma_{ii} + 2\sigma_{ii}\frac{\partial g_i(y)}{\partial y_i} + \sigma_{ii}^{3/2}\rho_i \frac{\partial^2 g_i(y)}{\partial y_i^2} + g_i(y)^2.$$

In practice, it might be difficult to find individual skewness coefficients ρ_i for each i. A better strategy might be to combine all the data points to come up with a common ρ-value.

3.2 Estimated variances

Consider first a survey where the $\hat{\sigma}_{ii}$'s are estimated using k random groups. Assuming normality, one can consider that $\{(k-1)\hat{\sigma}_{ii}/\sigma_{ii}: i = 1, ..., n\}$ is a sequence of independent χ^2_{k-1} random variables which is independent of y. Evaluating the conditional mean squared error (3) with variance estimates $\hat{\sigma}_{ii}$ yields potentially biased estimators, since $g_i(y)$ and its derivatives depend on $\hat{\sigma}_{ii}$. The potential bias can be expressed as

$$2E\left\{\hat{\sigma}_{ii}\frac{\partial g_i(y)}{\partial y_i}\right\} - 2\sigma_{ii} E\left\{\frac{\partial g_i(y)}{\partial y_i}\right\}. \tag{4}$$

As shown in the Appendix, this bias is $O(1/k)$. The next proposition suggests a small change to (3) that reduces its bias (4).

PROPOSITION 3: Replacing $\hat{\sigma}_{ii}$ by $(k-1)\hat{\sigma}_{ii}/(k+1)$ in the evaluation of $\partial g_i(y)/\partial y_i$ for calculating the mean squared error estimator (3) yields an estimator with an $O(1/k^2)$ bias.

The correction factor $(k-1)/(k+1)$ has been proposed in a different context by Scott and Smith (1971). Other methods are available for correcting the bias for estimating variances, depending on the way in which σ_{ii} is estimated. For instance if the $\hat{\sigma}_{ii}$ are independent $N\{\sigma_{ii}, \operatorname{var}(\hat{\sigma}_{ii})\}$ random variables distributed independently of y, then by Stein's lemma, (4) is equal to $2\operatorname{var}(\hat{\sigma}_{ii})E\{\partial^2 g_i(y)/\partial y_i \partial\hat{\sigma}_{ii}\}$.

Suppose now that the variances are estimated, not necessarily with random groups. In surveys, such as those considered in Dick (1995) and Hogan (1992), explanatory variables are available to model estimated variances. Small area estimators are then calculated with the predicted variances $\tilde{\sigma}_{ii}$ under the smoothing model; this means that $\tilde{\sigma}_{ii}$ enters in the calculation of g_i in (1). Considering (4), the mean squared error estimated with the smoothed variance,

$$\tilde{\sigma}_{ii} + 2\tilde{\sigma}_{ii}\frac{\partial g_i(y)}{\partial y_i} + g_i(y)^2$$

is unbiased provided that

$$2E\left\{(\tilde{\sigma}_{ii} - \sigma_{ii})\frac{\partial g_i(y)}{\partial y_i}\right\} = 0.$$

When $g_i(y)$ is calculated with smoothed variances, (4) should be small; the above condition holds provided that

$$E_V\left\{(\hat{\sigma}_{ii} - \tilde{\sigma}_{ii})\frac{\partial g_i(y)}{\partial y_i}\right\} = 0, \tag{5}$$

where index V refers to the model for smoothing the variances. One can easily test whether this condition holds by calculating the correlation between the variance residuals and the partial derivatives of the functions g_i. Since, as shown in Proposition 5 of the next section, unconditional mean squared errors can be derived as expectations of $\mathrm{mse}_c(\hat{\mu}_i)$ testing whether (5) is true is relevant even when unconditional measures of accuracy, such as Prasad and Rao's are calculated. Indeed, replacing variances by their predicted values biases the mean squared error estimators, conditional or unconditional, when (5) is violated.

4. Mean squared error estimation for empirical bayes estimators

4.1 Model construction

This section assumes that the y_i's are independent, *i.e.*, that Σ is diagonal. In an empirical Bayes setting, the model (M) for smoothing direct estimators expresses the parameters μ_i's as random variables whose distributions depend on a p-variate auxiliary variable x_i (Maritz and Lwin 1989, chapter 3),

$$\mu_i = x_i'\beta + v_i, \tag{6}$$

where β is a $p \times 1$ vector of unknown regression parameters and the v_i's are independent random variables with mean 0 and variance σ_v^2. Often the v_i's are assumed to be normally distributed; the marginal distribution of y_i, with respect to both the sampling design S and the smoothing model M, is then $N(x_i^t\beta, \sigma_{ii} + \sigma_v^2)$. The empirical Bayes estimators are obtained by shrinking the direct estimators y_i towards their predicted values under (6).

The extent of the shrinking depends on estimators of the parameters of (6) calculated from the marginal distribution of y_i. Several methods are available for parameter estimation (Cressie 1992). A popular estimator for σ_v^2 (see Lahiri and Rao (1995)) is

$$\hat{\sigma}_v^2 = \max\left[0, (n-p)^{-1}\left\{\sum_{i=1}^{n}(y_i - x_i^t\hat{\beta})^2 - \sum_{i=1}^{n}\sigma_{ii}(1-h_{ii})\right\}\right]$$

where $\hat{\beta} = (X^tX)^{-1}X^ty$, $h_{ii} = x_i^t(X^tX)^{-1}x_i$, and $X = (x_1, \ldots, x_n)^t$. The weighted least squares estimator of β is

$$\hat{\beta}_w = \hat{A}^{-1}\sum_{i=1}^{n}\frac{x_i y_i}{(\hat{\sigma}_v^2 + \sigma_{ii})},$$

where

$$\hat{A} = \sum_{i=1}^{n}\frac{x_i x_i^t}{(\hat{\sigma}_v^2 + \sigma_{ii})}.$$

The empirical Bayes estimator for μ_i is then

$$\begin{aligned}\hat{\mu}_i &= x_i^t\hat{\beta}_w + \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \sigma_{ii}}(y_i - x_i\hat{\beta}_w)\\ &= y_i - \frac{\sigma_{ii}}{\hat{\sigma}_v^2 + \sigma_{ii}}(y_i - x_i^t\hat{\beta}_w).\end{aligned} \tag{7}$$

Thus for empirical Bayes estimators, one has

$$g_i(y) = -\frac{\sigma_{ii}}{\hat{\sigma}_v^2 + \sigma_{ii}}(y_i - x_i^t\hat{\beta}_w).$$

4.2 The conditional mean squared error estimator

An explicit form for (3) can be obtained from the following formula for the derivative of the functions g_i for empirical Bayes estimators,

$$\frac{\partial g_i(y)}{\partial y_i} = \frac{\partial\hat{\sigma}_v^2}{\partial y_i}\frac{\partial g_i(y)}{\partial\hat{\sigma}_v^2} - \frac{\sigma_{ii}}{\hat{\sigma}_v^2 + \sigma_{ii}}\left\{1 - \frac{x_i^t\hat{A}^{-1}x_i}{(\hat{\sigma}_v^2 + \sigma_{ii})}\right\}, \tag{8}$$

The partial derivatives appearing in (8) can be evaluated using standard methods. They are given by

$$\frac{\partial\hat{\sigma}_v^2}{\partial y_i} = \frac{2}{(n-p)}(y_i - x_i^t\hat{\beta}),$$

and

$$\frac{\partial g_i(y)}{\partial \hat{\sigma}_v^2} = \frac{\sigma_{ii}}{(\hat{\sigma}_v^2 + \sigma_{ii})^2}(y_i - x_i^t\hat{\beta}_w) + \frac{\sigma_{ii}}{(\hat{\sigma}_v^2 + \sigma_{ii})} x_i^t \frac{\partial \hat{\beta}_w}{\partial \hat{\sigma}_v^2},$$

where

$$\frac{\partial \hat{\beta}_w}{\partial \hat{\sigma}_v^2} = -\hat{A}^{-1}\sum_1^n \frac{x_k(y_k - x_k^t\hat{\beta}_w)}{(\hat{\sigma}_v^2 + \sigma_{kk})^2}.$$

From (8), one has a closed form expression for $\text{mse}_c(\hat{\mu}_i)$. This statistics is an estimator of mean squared error for the empirical Bayes estimator for the i^{th} small area with respect to the sampling design only. It is valid for any sample size n; it relies on the sole assumption that the direct estimators y_i are normally distributed. When $\hat{\sigma}_v^2 = 0$, $\hat{\mu}_i = x_i^t\hat{\beta}_w$ and the derivatives in (8) simplify substantially. Since $\partial\hat{\sigma}_v^2/\partial y_i = 0$, one has

$$\text{mse}_c(\hat{\mu}_i) = (y_i - x_i^t\hat{\beta}_w)^2 - \sigma_{ii} + 2x_i^t\hat{A}^{-1}x_i.$$

The properties of the conditional mean squared error estimator are best investigated in the simple situation where all the parameters of the smoothing model are assumed to be known. In this situation, $\partial g_i(y)/\partial y_i = -\sigma_{ii}/(\sigma_{ii} + \sigma_v^2)$ and the conditional mean squared error estimator is equal to $\text{mse}_c^+(\hat{\mu}_i) = \max\{(\text{mse}_c(\hat{\mu}_i), 0\}$ where

$$\text{mse}_c(\hat{\mu}_i) = \frac{\sigma_{ii}\sigma_v^2}{\sigma_{ii} + \sigma_v^2} + \left(\frac{\sigma_{ii}}{\sigma_{ii} + \sigma_v^2}\right)^2 \{(y_i - x_i^t\beta)^2 - \sigma_{ii}\sigma_v^2\}.$$

The model based alternative to this estimator is the posterior variance, $\sigma_{ii}\sigma_v^2/(\sigma_{ii} + \sigma_v^2)$, which coincides with $E_M[E_S\{\text{mse}_c(\hat{\mu}_i)\}]$. This estimator is a special case of Prasad and Rao (1990) estimator and is denoted $\text{mse}_{\text{PR}}(\hat{\mu}_i)$. Estimator $\text{mse}_c^+(\hat{\mu}_i)$ is highly variable when σ_v^2 is small. Indeed, when σ_v^2 is close to 0, about 50% of the conditional mean squared error estimates are null. To further compare the 2 mean squared error estimators, conditional and unconditional, observe that when all the parameters of the smoothing model are known, the conditional mean squared error of $\hat{\mu}_i$ is

$$E_S\{\text{mse}_c(\hat{\mu}_i)\} = \frac{\sigma_{ii}\sigma_v^2}{\sigma_{ii} + \sigma_v^2} + \left(\frac{\sigma_{ii}}{\sigma_{ii} + \sigma_v^2}\right)^2 \{(u_i - x_i^t\beta)^2 - \sigma_v^2\}.$$

The next proposition compares the average mean squared errors of the estimators, conditional or unconditional, of $E_S\{\text{mse}_c(\hat{\mu}_i)\}$.

PROPOSITION 4: When $\sigma_{ii} = \sigma^2$, for $i = 1, ..., n$ and when the small area means are μ_i's are drawn using (6), the efficiency of the posterior variance with respect to the conditional mean squared error estimator for estimating the conditional mean squared error is

$$\frac{E_M[\sum \text{MSE}_S\{\text{mse}_c(\hat{\mu}_i)\}/n]}{E_M[\sum \text{MSE}_S\{\text{mse}_{\text{PR}}(\hat{\mu}_i)\}/n]} = \frac{\sigma^4 + 2\sigma^2\sigma_v^2}{\sigma_v^4}$$

where $\text{MSE}_S(\cdot)$ denote a mean squared error taken with respect to the distribution of the y_i's which are independent $N(\mu_i, \sigma^2)$ random variables.

The above efficiency is larger than 1 provided that $\sigma_v^2/\sigma^2 < 2.41$. Proposition 4 shows under heavy shrinking, the unconditional mean squared error estimator is a better estimator of the conditional mean squared error than the conditional estimator. This surprising result is caused by the large variance of the conditional estimator; when shrinking is extensive, it is a poor estimator.

In some situations, such as that consider in section 5.1, shrinking is light and the use of the conditional mean squared error estimator is appropriate. The conditional efficiency of $\hat{\mu}_i$ with respect to the direct estimator y_i is given by $\sigma_{ii}/\text{mse}_c^+(\hat{\mu}_i)$. This is larger than one provided that $(y_i - x_i^t\beta)^2/(\sigma_{ii} + \sigma_v^2) < 2$. Assuming that the smoothing model holds true, conditional efficiencies less than 1 can be expected for approximatively 16% ($= P[N(0.1)^2 < 2]$) of the small area estimators. This percentage should be higher if the smoothing model is deficient. Conditional efficiencies less than 1 occur in small areas having large residuals. On the other hand, the unconditional efficiencies, calculated with the posterior variance are, in this situation, less than 1 for all small areas. This shows that it is practically impossible for all the conditional efficiencies to be less 1; this had already been noted by Rao and Shinozaki (1978) for James-Stein estimators.

Many of the observations made in the unrealistic situation where all the parameters are known also apply when parameters are estimated. The unconditional alternative to the conditional mean squared error estimator is Prasad and Rao's (1990) estimator,

$$\text{mse}_{\text{PR}}(\hat{\mu}_i) = \frac{\sigma_{ii}\hat{\sigma}_v^2}{\sigma_{ii} + \hat{\sigma}_v^2} + \frac{\sigma_{ii}^2 x_i^t\hat{A}^{-1}x_i}{(\sigma_{ii} + \hat{\sigma}_v^2)^2} + 2\frac{\sigma_{ii}^2\widehat{\text{Var}}(\hat{\sigma}_v^2)}{(\sigma_{ii} + \hat{\sigma}_v^2)^3}, \quad (9)$$

where $\widehat{\text{Var}}(\hat{\sigma}_v^2) = 2\sum(\hat{\sigma}_{ii} + \hat{\sigma}_v^2)^2/n^2$. To investigate the extent to which Proposition 4 holds when parameters are estimated, a small Monte Carlo study was carried out along the lines of the approach ii) simulation study of Prasad and Rao (1999). In all the simulations, $n = 30$ and $\sigma_{ii} = 1$, for $i = 1, ..., n$. The smoothing model (6) was $\mu_i = \mu + v_i$ and various values of σ_v^2 were used. The results reported in Table 1 are based on $m = 5{,}000$ Monte Carlo replications.

The simulations used 5 sets of μ_i-values whose variances are reported in Table 1. For each set, y_i was simulated repeatedly as a $N(\mu_i, 1)$ random variable, $i = 1, ..., n$. The empirical Bayes estimate $\hat{\mu}_i$ was calculated for each small area and the mean squared error for small area i was calculated as $\text{MSE}_i = \sum^*(\hat{\mu}_I - \mu_I)^2/m$ where $\sum^*$ denotes the sum on the m Monte Carlo replications. The efficiency of the empirical Bayes estimator for small area i is $1/\text{MSE}_i$. The mean and the median of the $n = 30$ small area efficiencies are given in Table 1. The 2 mean squared errors, conditional and unconditional, were calculated for each small area in the m Monte Carlo replications; from (9), $\text{mse}_{\text{PR}}(\hat{\mu}_i) = (\hat{\sigma}_v^2 + 5/n)/(1+\hat{\sigma}_v^2)$ for each small area. Table 1 presents the mean and the median of their absolute relative biases, defined a $|\sum^*(\text{mse}_i(\hat{\mu}_i) - \text{MSE}_i)|/(m\text{MSE}_i)$ and of their coefficients of variation which are equal to $(\sum^*(\text{mse}_i(\hat{\mu}_i) - \text{MSE}_i)^2/m)^{1/2}/\text{MSE}_i$.

Table 1
Relative Efficiency of the empirical Bayes estimators (RE), absolute Relative Bias (RB) and Coefficient of Variation (CV) of two mse estimators (n = 30). All results are expressed in percentage

$\Sigma(\mu_i - \hat{\mu})^2/29$		RE%	RB$_c$%	RB$_{PR}$%	CV$_c$%	CV$_{PR}$%
1.3	mean	212	1	47	97	51
	median	214	1	40	100	43
2.53	mean	149	2	30	37	31
	median	163	2	31	37	32
3.7	mean	129	2	20	23	20
	median	133	1	21	24	21
4.24	mean	125	2	19	19	20
	median	131	1	22	20	22
4.93	mean	122	1	17	15	18
	median	133	1	17	13	17

As shown in section 2, $\text{mse}_c(\hat{\mu}_i)$ is unbiased; the biases reported in Table 1 are caused by Monte Carlo errors. When $n = 30$, the condition $\sigma_v^2/\sigma^2 > 2.4$ derived in Proposition 4 for the conditional estimator to improve on the unconditional estimator is not sufficient; the stronger condition $\sigma_v^2/\sigma^2 > 4$ is needed. Noteworthy is the fact that in Table 1, for $\sum(\mu_i - \bar{\mu})^2/29 > 2.5$, the CV of $\text{mse}_{\text{PR}}(\hat{\mu}_i)$ is only bias. Table 1 confirms that, when $\hat{\sigma}_y^2$ is of the same order of magnitude as σ_{ii} or smaller, the squared residual dominates the distribution of the conditional mean squared error estimator; in such cases Prasad and Rao (1990) unconditional estimator is a better estimator of conditional mean squared error. Even in situations when $\text{mse}_c(\hat{\mu}_i)$ cannot be recommended as an estimator for the conditional mean squared error, it still provides interesting diagnostic information: changes in the conditional estimators give a basis for comparing two smoothing models. This is illustrated in section 5.2.

4.3 Conditional mean squared error and prediction variance

This section explores the relationship between the conditional mean squared error proposed in this paper and the prediction variance which is an unconditional measure of accuracy. Using the rotation of (6), the prediction variance is $\text{MSE}(\hat{\mu}_i) = E_M[E_S\{(\hat{\mu}_i - x_i^t\beta - v_i)^2\}]$. From the construction of presented in section 2, one has

$$E_S\{\text{mse}_c(\hat{\mu}_i)\} = E_S\{(\hat{\mu}_i - x_i^t\beta - v_i)^2\}.$$

Thus we have the following result,

PROPOSITION 5: The conditional mean squared error of empirical Bayes small area estimators satisfies,

$$E_M[E_S\{\text{mse}_c(\hat{\mu}_i)\}] = \text{MSE}(\hat{\mu}_i),$$

where $\text{MSE}(\hat{\mu}_i)$ is the unconditional prediction variance.

Proposition 5 shows that $\text{mse}_c(\hat{\mu}_i)$ can be looked at as an intermediate step in the evaluation of the unconditional mean squared error of $\hat{\mu}_i$. Consider for instance the calculation of Prasad and Rao (1990) $o(1/n)$ approximation to $\text{MSE}(\hat{\mu}_i)$,

$$\text{MSE}_{\text{PR}}(\hat{\mu}_i) = \frac{\sigma_{ii}\sigma_v^2}{\sigma_{ii}+\sigma_v^2} + \frac{\sigma_{ii}^2 x_i^t A^{-1} x_i}{(\sigma_{ii}+\sigma_v^2)^2} + \frac{\sigma_{ii}^2\,\text{Var}(\hat{\sigma}_v^2)}{(\sigma_{ii}+\sigma_v^2)^3},$$

where $\text{Var}(\hat{\sigma}_v^2) = 2\sum(\sigma_{ii}+\sigma_v^2)^2/n^2$. The standard derivation, as reviewed in section 3.2 of Singh, Stukel, and Pfeffermann (1998), is based on Kackar and Harville (1984). An alternative derivation, presented in Belmonte (1998, 1999), is to take the expectation of $\text{mse}_c(\hat{\mu}_i)$, obtained using (8), with respect to the marginal distribution of the y_i's, which are independent $N(x_i^t\beta, \sigma_{ii}+\sigma_v^2)$ deviates and to retain only the higher order terms.

Proposition 5 holds in situations where the small area estimators are bench-marked, or where corrections suggested in section 3 are implemented. These are cases for which there are no closed form formulas for the prediction variances. Proposition 4 suggests a simple method for constructing unconditional Monte Carlo estimates. It suffices to generate a large number of replicates of $\{y_i,\ i = 1, ..., n\}$ where y_i follows a $N(x_i^t\hat{\beta}_w, \hat{\sigma}_v^2 + \sigma_{ii})$ and to calculate $\text{mse}_c(\hat{\mu}_i)$ for each one. Averaging the $\text{mse}_c(\hat{\mu}_i)$'s gives a plug-in unconditional prediction variance, equal to the MSE of Proposition 4 evaluated at estimates $\hat{\beta}_w, \hat{\sigma}_v^2$ of the unknown parameters. Unfortunately, this estimate is biased (this is a first order estimate in the terminology of Singh, Stukel and Pfeffermann (1998)). For the empirical Bayes estimator given by (7), according to (9) the bias of the Monte Carlo estimate derived from Proposition 4 is $-\sigma_{ii}^2\,\text{Var}(\hat{\sigma}_v^2)/(\sigma_{ii}+\hat{\sigma}_v^2)^3$. Further work is needed for constructing, using Proposition 4, unbiased unconditional prediction variance estimators.

5. Estimating the under-coverage in the 1991 Canadian Census

In 1991, the under-coverage of the Canadian Census was estimated using two surveys, the Over-coverage Study, which estimates the number of persons double counted or erroneously counted in the Census and the Reverse Record Check (Burgess 1988) for the persons missed in the Census. Combining these figures gives estimates of the under-coverage of the Census. This section investigates several estimators of census under-coverage.

5.1 Provincial estimations

The 1991 under-coverage rates for the ten Canadian provinces and the two territories with their coefficients of variation, expressed in percentage, are given in Table 2. The proportion p_i of the population living in each province (the word province is used in this section to denote the 10 Canadian provinces and the two territories) is also provided. The coefficients of variation (CV) of Table 2 were calculated from variances estimated with 5 random groups. Thus, one can consider that the sampling variances have a χ_4^2 distribution. Throughout this section, we assume that the provincial under-coverage estimates and their variances are independent.

Several estimators for provincial under-coverage are proposed by Royce (1992). Rivest (1995) proposed a composite estimator that shrinks the provincial under-coverage rate towards the national rate. It is given by:

$$r_i^c = \hat{\alpha} r_i + (1-\hat{\alpha}) r_N,$$

where $r_N = \sum p_i r_i$ is the national under-coverage rate and the shrinking parameter $\hat{\alpha}$ is given by:

$$\hat{\alpha} = \frac{\sum p_i r_i^2 - r_N^2}{\sum p_i (1-p_i) \sigma_i^2 + \sum p_i r_i^2 - r_N^2}.$$

This is the value of α that is optimal for loss functions for the estimation of provincial totals and of provincial shares of the population; see Royce (1992) and Rivest (1995) for details. One has $r_i^c = r_i + g_i(r)$, where

$$g_i(r) = -\frac{\sum p_i (1-p_i) \sigma_i^2}{\sum p_i (1-p_i) \sigma_i^2 + \sum p_i r_i^2 - r_N^2} (r_i - r_N).$$

A closed form expression for the conditional mean square error estimator can be calculated easily by noting that

$$\frac{\partial g_i(r)}{\partial r_i} = 2 p_i (r_i - r_N)^2 \frac{\sum p_i (1-p_i) \sigma_i^2}{\left[\sum p_i (1-p_i) \sigma_i^2 + \sum p_i r_i^2 - r_N^2\right]^2} - (1-p_i)(1-\hat{\alpha}).$$

The second partial derivative of $g_i(r)$ can also be calculated; it has the same sign as $r_i - r_N$. Thus positive skewness in the under-coverage rate, that is likely when estimating rare events such as being missed by the census, increases the conditional mean squared error in provinces where the under-coverage is above the national rate.

For 1991, $\hat{\alpha} = 0.874$ and the national under-coverage rate is $r_N = 2.872\%$. Table 2 gives the provincial composite under-coverage estimates, r_i^c together with their efficiencies $\text{eff}_{ic}^c = \sigma_{ii} / \text{mse}_c(r_i^c)$, where $\text{mse}_c(r_i^c)$ is calculated as defined in section 2, with the correction proposed in section 3.2 to account for estimated variances. The composite estimator is an improvement over the direct estimators in all cases except three, that correspond to the provinces with the most extreme under-coverage rates.

Table 2 also gives the empirical Bayes estimator r_i^B calculated with a location smoothing model. Under model (M), the true under-coverage rate θ_i is assumed to be distributed as a $N(\beta, \sigma_v^2)$. The parameter estimates are $\hat{\sigma}_v^2 = 1.45 \times 10^{-4}$ and $\hat{\beta}_w = 2.61\%$. Two efficiencies with respect to direct estimators are presented, eff_{ic}^B which is calculated with the conditional mean squared error estimator for r_i^B, including the adjustment of section 3.2 to account for estimated variances, and $\text{eff}_{i\text{PR}}^B$ which is calculated with Prasad-Rao unconditional estimator. The large under-coverage rate in the N.W. Territories is responsible for the large estimate for $\hat{\sigma}_v^2$; this makes the empirical Bayes estimators r_i^B much closer to the direct estimators r_i than the composite estimators. Redoing the analysis without the N.W. Territories and Yukon changes the empirical Bayes estimates drastically.

Table 2
Two estimators of provincial under-coverage and their efficiencies

Province	p_i	r_i	CV	r_i^c	eff_{ic}^c	r_i^B	eff_{ic}^B	$\text{eff}_{i\text{PR}}^B$
Newfoundland	2.06	1.994	15.96	2.105	1.12	2.038	1.07	1.04
Prince Edward Island	0.47	0.931	30.00	1.176	0.65	1.025	0.93	1.03
Nova Scotia	3.26	1.889	20.05	2.013	1.11	1.959	1.09	1.06
New Brunswick	2.66	3.245	13.73	3.198	1.29	3.162	1.14	1.09
Québec	25.19	2.605	8.35	2.639	1.16	2.605	1.04	1.02
Ontario	37.24	3.641	8.46	3.544	0.87	3.572	1.02	1.04
Manitoba	3.96	1.86	20.83	1.987	1.10	1.936	1.09	1.06
Saskatchewan	3.58	1.798	18.87	1.933	1.04	1.863	1.06	1.05
Alberta	9.24	1.995	14.57	2.106	1.01	2.032	1.06	1.03
British Columbia	12.01	2.733	9.86	2.751	1.26	2.727	1.07	1.03
Yukon	0.10	3.83	15.99	3.709	1.27	3.56	1.05	1.17
N.W. Territories	0.22	5.439	11.28	5.116	0.96	4.813	0.49	1.18

In Table 2, the composite estimator performs better than the empirical Bayes estimator; it provides gains in conditional efficiency larger than 10% in 7 of 12 provinces. Three efficiencies are smaller than 1; the discussion in section 4.2 suggests that efficiencies less than 1 are unavoidable. The relatively poor precision of $\hat{\sigma}_{ii}$ (they are estimated using only 4 degrees of freedom), lowers the conditional efficiencies of the empirical Bayes estimators. It does not affect the composite estimator as much since it uses the same shrinking parameter for all provinces. The conditional efficiencies capture the poor performances of the r_i^c and r_i^B in the provinces with the most extreme under-coverage rates. This is missed by the Prasad Rao efficiencies. They highlight the gains that smoothing brings to the two territories where the under-coverage rates are highly variable. The Prasad Rao efficiencies are meaningful only if one accepts the hypothesis of provincial exchangeability underlying the smoothing model. This is doubtful since under-coverage tends to be higher in large urban provinces than in small rural areas.

5.2 Sub-provincial estimations

Dick (1995) considered the estimation of the adjustment factors for census under-coverage for age × sex categories within each province for the 1991 census. The adjustment factor for a small area is defined as F = 1 + (estimated under-coverage)/(census count). With four age categories, 0-19, 20-29, 30-44, 45+, and two sexes, there are 96 small areas. The explanatory variables are interactions between the indicator variables for the 12 provinces, the 4 age groups and the two sexes, and the proportions of renters (R) and of people that do not speak either official language (L) in the 96 small areas. In each one, the estimated variance was given by $\hat{\sigma}_{ii}$ = (under-coverage variance)/(census count) ^{2}.

Dick (1995) regressed the log-variances on the census count to smooth the variance. He considered the exponentials of the predicted values for the log-variances $(\tilde{\sigma}_{ii})$ as the known variances. This underestimates the variability.

Indeed, the average predicted variance $\tilde{\sigma}_{ii}$ represents only 68% of the average unsmoothed variance. Multiplying $\tilde{\sigma}_{ii}$ by $\exp(\hat{\sigma}_r^2/2) = 1.54$, where $\hat{\sigma}_r^2$ is the residual variance of the smoothing model, corrects this problem. Fitting Dick's (1995) model using the "unbiased" smoothed variance yields $\hat{\sigma}_v^2 = 0$. This is a degenerate situation where empirical Bayes estimators are equal to linear model predicted values. Note also the correlation between the variance residuals and the partial derivatives of g_i, calculated as if $\hat{\sigma}_v^2 > 0$, is 0.25. This suggests that (5) is violated. Using $\tilde{\sigma}_{ii}\exp(\hat{\sigma}_r^2/2)$ in the calculation is likely to over-estimate the precision the small area estimates. To illustrate the application of the conditional mean squared error estimator, these problems are ignored and the remainder of this section assumes that the sampling variances σ_{ii} are known and equal to their smoothed values $\tilde{\sigma}_{ii}$.

The model fitted by Dick (1995) has ten independent variables; the weighted least squares estimates and their standard errors, given by the square roots of the elements on the diagonal matrix of $\hat{A}^{-1}$, appear in Table 3. The conditional mean squared errors $\mathrm{mse}_c^+(\hat{\mu}_i)$ for the 96 small areas can be calculated using (8). One had $\mathrm{mse}_c^+(\hat{\mu}_i) = 0$ and $\mathrm{mse}_c^+(\hat{\mu}_i) > \sigma_{ii}$ for respectively 51 and 15 small areas. The 15 small areas with large conditional mean squared errors need special attention: can the prediction model be improved for these areas? Two systematic features among the 15 corresponding residuals are noteworthy: there are 2 large positive residuals in the M/0-19 category and 2 large negative residuals in the F/45+ category. This suggests adding M/0-19 and F/45+ as independent variables. The additional column to the X matrix for M/0-19 contains 1's for the 12 small areas for males between 0 and 19 years old and 0 elsewhere; that for F/45+ is constructed in a similar way. Only F/45+ improves the fit; adding this explanatory variable gives the modified Dick model of Table 3. The absolute value of the t-statistic for F/45+ is 3; this is clearly significant.

It is interesting to compare the conditional mean squared errors obtained with the modified Dick model with those for Dick's model. Using the modified model decreases mse_c^+ in 26 small areas and increases it in 21; showing a slight improvement with the modified model.

The sub-provincial empirical Bayes adjustment factors can be aggregated at the provincial level. Provincial adjustment factors F_p are given by

$$\hat{F}_p = \frac{\sum_p C_i \hat{F}_i}{\sum_p C_i}$$

where C_i represents the census count for the i^{th} small area and $\sum_p$ is the summation over the 8 small areas in province p. A mean squared error for the provincial adjustment factor, either conditional or unconditional, can be calculated using a mean product error matrix mpe as

$$\mathrm{mse}(\hat{F}_p) = \frac{1}{\left(\sum_p C_i\right)^2} \sum_p \sum_p C_i C_j \,\mathrm{mpe}(\hat{F}_i,\ \hat{F}_j).$$

Conditional mean squared errors are obtained by using formula (2) for mpe. Lahiri and Rao (1995) give a formula for the off-diagonal terms of the unconditional mean product error matrix whose diagonal is given by Prasad Rao (1990) mean squared errors.

Table 3
Two linear models for small area correction factors: dick (p = 11) and modified dick (p = 12). Parameter estimates are given with their standard errors in parentheses

Category	Variable	Dick	modified Dick
mean	intercept	1.0076 (0.0018)	1.0099 (0.0018)
Age* Sex Interaction	M/20-29	0.0563 (0.0038)	0.0541 (0.0037)
	M/30-44	0.0207 (0.0036)	0.0185 (0.0035)
	F/20-20	0.0243 (0.0038)	0.02223 (0.0037)
	F/45+	-	-0.0102 (0.0037)
Province* Renters Interaction	BC*R	0.0436 (0.0115)	0.0433 (0.0110)
	Ontario*R	0.0791 (0.0100)	0.0789 (0.0102)
	Québec*R	0.0253 (0.0097)	0.0259 (0.0090)
	N.-B*R	0.1039 (0.0194)	0.1032 (0.0186)
	Yukon*R	0.0633 (0.0179)	0.0634 (0.0175)
	NWT*R	0.0687 (0.0117)	0.0680 (0.0285)
Language*Sex*Age Interaction	L*F/0-19	0.0802 (0.0293)	0.0680 (0.0285)
Variance		3.3681e-05 (2.45e-05)	2.21e-05 (2.30e-05)

Table 4
Direct (F_p) and empirical Bayes (F_p^b) estimates of the provincial correction factors with their conditional (eff_{pc}) and their unconditional $(\text{eff}_{p\text{PR}})$ efficiencies. A conditional efficiency is ∞ when the conditional mean squared error estimator is null

PROVINCE	F_p	F_p^b	eff_{pc}	$\text{eff}_{p\text{PR}}$
Newfoundland	1.0203	1.0176	6.49	2.94
Prince Edward Island	1.0094	1.0153	1.03	4.52
Nova Scotia	1.0193	1.0171	25.3	2.59
New Brunswick	1.0335	1.0367	0.67	1.11
Québec	1.0268	1.0262	1.12	0.93
Ontario	1.0378	1.0396	0.68	0.93
Manitoba	1.0190	1.0176	∞	2.46
Saskatchewan	1.0183	1.0166	∞	2.54
Alberta	1.0204	1.0187	7.37	1.98
British Columbia	1.0281	1.0293	1.09	1.03
Yukon	1.0396	1.0400	1.41	1.17
N.W. Territory	1.0575	1.0550	1.40	1.32

Direct and empirical Bayes aggregated estimates are presented in Table 4 with two efficiencies. The empirical Bayes estimates retain much of the interprovincial differences. This suggest that the explanatory variables of the smoothing model have captured most of the differences between the provincial under-coverage rates. A notable exception is Prince Edward Island's small correction factor which is not accounted for by the explanatory variables. This is the only province for which the two efficiencies differ substantially. The conditional efficiencies are more unstable than the Prasad Rao efficiencies. Except in Prince Edward Island, both tell similar stories: in New Brunswick, Quebec, Ontario, and British Columbia, the aggregated empirical Bayes estimates do not improve much on the direct estimators.

6. Conclusions

The estimator of the conditional mean squared error proposed in this paper has several interesting features. It can be implemented with any shrinking strategy. It is conditional on the realization of the smoothing model used to produce the small area characteristics; thus the conditional estimator has a large sampling variance. Simple modifications to the estimator are available to handle skewness in the data and estimated variances. In an empirical Bayes setting, it provides diagnostic information concerning the smoothing model. It can also be used as building blocks for estimators of the prediction variances when this variance has no closed form expression.

Acknowledgements

We are grateful to Peter Dick for providing the data set analyzed in section 5.2, and to Jon Rao for pointing out the instability of the conditional estimator under heavy smoothing. The financial contributions of the Fonds pour la formation des chercheurs et l'aide à la recherche du Québec and of the National Science and Engineering Research Council of Canada are gratefully acknowledged.

Appendix

Proof of proposition 1

Let $\Sigma^{1/2}$ be a symmetric square root for Σ, such that $(\Sigma^{1/2})^2 = \Sigma$ and $z = \Sigma^{-1/2}(y - \mu)$. Note that z has a $N_n(0, I)$ distribution. In terms of the random vector $z, E\{(y - \mu)\, g(y)^t\} = \Sigma^{1/2} E\{zg(\mu + \Sigma^{1/2} z)\}$. Now the conditional expectation of $z_i\, g_j(\mu + \Sigma^{1/2} z)$ given $z_1, ..., z_{i-1}, z_{i+1}, ..., z_n)$ is equal to

$$\int_R \frac{z_i \exp(-z_i^2/2)}{\sqrt{2\pi}} g_j\left(\mu + \Sigma^{1/2} z\right) dz_i.$$

Integrating by parts shows that the above integral is equal to

$$\int_R \frac{\exp(-z_i^2/2)}{\sqrt{2\pi}} \frac{\partial g_j\left(\mu + \Sigma^{1/2} z\right)}{\partial z_i} dz_i,$$

Observe that

$$\frac{\partial g_j\left(\mu + \Sigma^{1/2} z\right)}{\partial z_i} = \sum_{k=1}^{n} \Sigma_{ki}^{1/2} g_j^k\left(\mu + \Sigma^{1/2} z\right).$$

Since $\Sigma^{1/2}$ is symmetric, $\Sigma_{ki}^{1/2} = \Sigma_{ik}^{1/2}$. Thus the above expression is the scalar product between $e_i^t \Sigma^{1/2}$, the i^{th} row of $\Sigma^{1/2}$(e_i represents a $n \times 1$ vector of 0's except for the i^{th} component which is 1), and $\nabla g(\mu + \Sigma^{1/2} z) e_j$, the j^{th} column of $\nabla g(y)$, evaluated at $y = \mu + \Sigma^{1/2} z$. We have

$$E\left\{z_i\, g_j\left(\mu + \Sigma^{1/2} z\right) \mid z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n\right\} =$$
$$e_i^t \Sigma^{1/2} E\left\{\nabla g\left(\mu + \Sigma^{1/2} z\right) e_j \mid z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n\right\}.$$

This equality also holds unconditionally, $E\{z_i g_j(\mu + \Sigma^{1/2} z)\} = e_i^t \Sigma^{1/2} E\{\nabla g(\mu + \Sigma^{1/2} z)\} e_j$. In other words,

$$E\left\{z g\left(\mu + \Sigma^{1/2} z\right)\right\} = \Sigma^{1/2} E\left\{\nabla g\left(\mu + \Sigma^{1/2} z\right)\right\}.$$

This completes the proof.

Proof of proposition 2

Let E_i denote the conditional expectation with respect to y_i, given $(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$ and $h(y_i) = g_i(y)$, for fixed values of $(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$. One has

$$E_i\{(y_i - \mu_i) h(y_i)\} = \int_R (t - \mu_i) h(t) f(t)\, dt.$$

To evaluate this expression, one can integrate by parts. Integrating $(t - \mu_i) \exp\{-(t - \mu_i)^2/(2\sigma_{ii})\}/(2\pi\sigma_{ii})^{1/2}$ in the above integrand yields

$$E_i\{(y_i - \mu_i) h(y_i)\} = \sigma_{ii}\, E_i\{h'(y_i)\} + \frac{\sigma_{ii}^{1/2} \rho_i}{2}$$
$$\times \int_R h(t) \left\{\frac{(t - \mu_i)^2}{\sigma_{ii}} - 1\right\} \frac{\exp\{(t - \mu_i)^2/(2\sigma_{ii})\}}{(2\pi\sigma_{ii})^{1/2}} dt,$$

where $h'(t)$ is the derivative of $h(t)$. Repeated integrations by parts show that

$$\int_R h(t) \frac{(t - \mu_i)^2}{\sigma_{ii}} \frac{\exp\{(t - \mu_i)^2/(2\sigma_{ii})\}}{(2\pi\sigma_{ii})^{1/2}} dt$$
$$= \int_R \{h'(t)(t - \mu_i) + h(t)\} \frac{\exp\{(t - \mu_i)^2/(2\sigma_{ii})\}}{(2\pi\sigma_{ii})^{1/2}} dt$$
$$= \int_R \{\sigma_{ii} h''(t) + h(t)\} \frac{\exp\{(t - \mu_i)^2/(2\sigma_{ii})\}}{(2\pi\sigma_{ii})^{1/2}} dt$$

where $h''(t)$ is the second derivative of $h(t)$. This yields

$$E_i\{(y_i - \mu_i) h(y_i)\} =$$
$$\sigma_{ii} E_i\{h'(y_i)\} + \frac{\sigma_{ii}^{3/2} \rho_i}{2} E_i\{h''(y_i)\} + o(\rho_i).$$

Taking, on both sides, the expectation with respect to the distribution of $(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$ completes the proof.

Proof of proposition 3

Let E_i denote the expectation taken with respect to the distribution of $\hat{\sigma}_{ii}$, given all the other random quantities $(y, \hat{\sigma}_{jj}, j \neq i)$. In this context one can write $(\partial g_i(y))/(\partial y_i) = h(\hat{\sigma}_{ii})$, where h is a function possibly depending on $(y, \hat{\sigma}_{jj}, j \neq i)$. A Taylor series expansion of h gives:

$$h(\hat{\sigma}_{ii}) = h(\sigma_{ii}) + h'(\sigma_{ii})(\hat{\sigma}_{ii} - \sigma_{ii})$$
$$+ h''(\sigma_{ii}) = \frac{(\hat{\sigma}_{ii} - \sigma_{ii})^2}{2} + O((\hat{\sigma}_{ii} - \sigma_{ii})^3).$$

Since $(k-1)\hat{\sigma}_{ii}/\sigma_{ii}$ follows a χ_{k-1}^2 distribution, $E_i\{(\hat{\sigma}_{ii} - \sigma_{ii})^2\} = 2\sigma_{ii}^2/(k-1)$, and the centered moments of higher orders are $O(1/k^2)$. The above expansion reduces to,

$$\sigma_{ii} E_i\{\partial g_i(y)/\partial y_i\} = \sigma_{ii} h(\sigma_{ii}) + h''(\sigma_{ii}) \frac{\sigma_{ii}^3}{k-1} + O(1/k^2).$$

It is clear that the bias of $\hat{\sigma}_{ii} h(\hat{\sigma}_{ii})$ as an estimator of this expression is $O(1/k)$, provided that $h'(\sigma_{ii}) \neq 0$. One has, neglecting $O(1/k^2)$ terms,

$$E_i\left\{\hat{\sigma}_{ii} h\left(\frac{(k-1)\hat{\sigma}_{ii}}{k+1}\right)\right\}$$
$$\approx \sigma_{ii} h(\sigma_{ii}) + h'(\sigma_{ii}) E_i\left\{\hat{\sigma}_{ii}\left(\frac{(k-1)\hat{\sigma}_{ii}}{k+1} - \sigma_{ii}\right)\right\}$$
$$+ \frac{h''(\sigma_{ii})}{2} E_i\left\{\hat{\sigma}_{ii}\left(\frac{(k-1)\hat{\sigma}_{ii}}{k+1} - \sigma_{ii}\right)^2\right\}.$$

Elementary manipulations show that in the above formula, the coefficient of $h'(\sigma_{ii})$ is null and

$$E_i\left\{\hat{\sigma}_{ii}\left(\frac{(k-1)\hat{\sigma}_{ii}}{k+1}-\sigma_{ii}\right)^2\right\}=2\frac{\sigma_{ii}^3}{k-1}+O(1/k^2).$$

This shows that

$$E_i\left\{\hat{\sigma}_{ii}h\left(\frac{(k-1)\hat{\sigma}_{ii}}{k+1}\right)\right\}=\sigma_{ii}E_i\{\partial g_i(y)/\partial y_i\}+O(1/k^2).$$

The proof is completed by noting that this equality holds for the unconditional expectation, taken with respect to the joint distribution of $(y, \hat{\sigma}_{ii}, i = 1,\ldots,n)$.

Proof of proposition 4

The mean squared error of the posterior variance as an estimator of the conditional mean squared error has only a bias term,

$$\left(\frac{\sigma^2}{\sigma^2+\sigma_v^2}\right)^4\{(\mu_i-x_i^t\beta)^2-\sigma_v^2\}^2,$$

while the mean squared error of $\text{mse}_c(\hat{\mu}_i)$ has only a variance component which is given by

$$\left(\frac{\sigma^2}{\sigma^2+\sigma_v^2}\right)^4\text{Var}_S\{(y_i-x_i^t\beta)^2\}$$
$$=\left(\frac{\sigma^2}{\sigma^2+\sigma_v^2}\right)^4\{2\sigma_{ii}^2+4(\mu_i-x_i^t\beta)^2\sigma^2\}.$$

The efficiency reported in Proposition 4 can be evaluated as the ratio of the 2 average mean squared errors defined above. It is given by,

$$\frac{2\sigma^4+4\sigma^2\sum(\mu_i-x_i^t\beta)^2/n}{\sum\{(\mu_i-x_i^t\beta)^2-\sigma_v^2\}^2/n}.$$

Taking expectations of the numerator and of the denominator with respect to model (6) yields the result.

References

Barndorff-Nielsen, O.E., and Cox, D.R. (1989). *Asymptotic Techniques for Use in Statistics*. New York: Chapman and Hall.

Belmonte, E. (1998). Estimation dans les petites régions : une nouvelle dérivation de l'erreur quadratique moyenne de Prasad-Rao. *1998 Proceedings of the Survey Methods Section, Statistical Society of Canada*, 165-170.

Belmonte, E. (1999). *L'estimation dans les petites regions* : *survol des méthodes de Bayes et présentation d'un estimateur conditionnel de l'EQM*. Mémoire de maîtrise. Département de mathématiques et de statistique, Université Laval.

Bilodeau, M., and Srivastava, M.S. (1988). Estimation of the MSE matrix of the Stein estimator. *Canadian Journal of Statistics*, 16, 153-159.

Booth, J.G., and Hobert, J.P. (1998). Standard errors of predictions in generalized linear mixed models. *Journal of the American Statistical Association*, 93, 262-272.

Burgess, R.D. (1988). Evaluation of the reverse record check estimates of under-coverage in the Canadian Census of Population. *Survey Methodology*, 14, 137-156.

Cressie, N. (1992). REML estimation in empirical Bayes smoothing of census undercount. *Survey Methodology*, 18, 75-94.

Dick, P. (1995). Modeling net under-coverage in the 1991 Canadian Census. *Survey Methodology*, 21, 44-55.

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.

Ghosh, M., and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, 9, 55-93.

Hogan, H. (1992). The 1990 Post-Enumeration Survey: An overview. *The American Statistician*, 46, 261-269.

Kackar, R.N., and Harville, D.A. (1984). Approximations for standard errors of estimators for fixed and random effects in mixed models. *Journal of the American Statistical Association*, 79, 853-862.

Kott, P.S. (1989). Robust small domain estimation using random effect modeling. *Survey Methodology*, 15, 3-12.

Lahiri, P., and Rao, J.N.K. (1995). Robust estimation of mean squared error of small area estimators. *Journal of the American Statistical Association*, 90, 758-766.

Maritz, J.S., and Lwin, T. (1989). *Empirical Bayes Methods*. (Second Edition), London: Chapman and Hall.

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.

Prasad, N.G.N., and Rao, J.N.K. (1999). On robust small area estimation using a simple random effect model. *Survey Methodology*, 25, 163-171.

Purcell, N.J., and Kish, L. (1979). Estimation for small domains. *Biometrics*, 35, 365-384.

Rao, C.R., and Shinozaki, N. (1978). Precision of individuals estimators in simultaneous estimation of parameters. *Biometrika*, 65, 23-30.

Reid, N. (1991). Approximation and asymptotics. *In Statistical Theory and Modeling. In Honor of Sir David Cox, FRS*, (Eds. D.V. Hinkley, N. Reid and E.J. Snell), 287-305.

Rivest, L.P. (1995). A composite estimator for provincial under-coverage in the Canadian census. *1995 Proceedings of the Survey Methods Section*, *Statistical Society of Canada*, 33-38.

Robert, C. (1992). *L'Analyse Statistique Bayesienne*. Paris: Economica.

Royce, D. (1992). A comparison of some estimators for a set of population totals. *Survey Methodology*, 18, 109-125.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Scott, A., and Smith, T.M.F. (1971). Interval estimates for linear combinations of means. *Applied Statistics*, 20, 276-285.

Singh, M.P., Gambino, J. and Mantel, H.J. (1994). Issues and strategy for small data. *Survey Methodology*, 20, 1-22.

Singh, A.C., Stukel, D.M. and Pfeffermann, D. (1998). Bayesian versus frequentist measures of error in small area estimation. *Journal of the Royal Statistical Society*, Series B, 60, 377-396.

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

Wolter, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.