Examining Sensitivity of Small Area Inferences to Uncertainty About Sampling Error Variances*

William R. Bell†

Abstract

Small area estimation based on area level models typically assumes that sampling error variances for the direct survey small area estimates are known. In practice we use estimates of the sampling error variances, and these can contain substantial error. This suggests modeling the sampling variances to improve them and to quantify effects of their estimation error on small area inferences. We review papers that have attempted to address these issues. We then provide some results on the latter issue, showing, in a simple framework, how error in estimating sampling variances can affect the accuracy of small area predictions and lead to bias in stated mean squared errors.

Key Words: sampling error model, Fay-Herriot model, Bayesian inference, small area estimation

1. Introduction

A basic area level model used in small area estimation (Fay and Herriot 1979, Rao 2003) is as follows:

$$\begin{aligned} y_i &= Y_i + e_i \qquad i = 1, \ldots, m \\ &= (\mathbf{x}_i'\beta + u_i) + e_i \end{aligned} \tag{1}$$

where the y_i are direct survey estimates of true population quantities Y_i for m small areas, the e_i are sampling errors (of the y_i) independently distributed as $N(0, v_i)$, the u_i are small area random effects (model errors) distributed $i.i.d.$ $N(0, \sigma_u^2)$, the $\mathbf{x}_i$ are $r \times 1$ vectors of regression variables for area i, and β is the corresponding vector of regression parameters. Normality is not an essential assumption. A "standard" assumption, however, is that the sampling variances, v_i, are all known. If the model error variance, σ_u^2, is also known, then the best linear unbiased predictor (BLUP) of Y_i and its mean squared error (MSE) are (Rao 2003, 96-99 and 116-117)

$$\tilde{Y}_i = h_i y_i + (1 - h_i)\mathbf{x}_i'\hat{\beta} \tag{2}$$

$$\mathrm{Var}(Y_i - \tilde{Y}_i) = \sigma_u^2(1 - h_i) + (1 - h_i)^2 \mathbf{x}_i' \mathrm{Var}\left(\hat{\beta}\right) \mathbf{x}_i \tag{3}$$

where $h_i = \sigma_u^2/(\sigma_u^2 + v_i)$ and $\hat{\beta}$ and $\mathrm{Var}(\hat{\beta})$ come from weighted least squares results:

$$\hat{\beta} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{y} \tag{4}$$

$$\mathrm{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1} \tag{5}$$

where $\mathbf{y} = (y_1, \ldots, y_n)'$, $\mathbf{X}$ is $n \times r$ with rows $\mathbf{x}_i'$, and $\mathrm{Var}(\mathbf{y}) \equiv \mathbf{\Sigma} = \mathrm{diag}(\sigma_u^2 + v_i)$.

From (2), the smoothed estimate $\tilde{Y}_i$ is a weighted average of the direct estimate y_i and the regression prediction $\mathbf{x}_i'\hat{\beta}$, with weights h_i and $1 - h_i$ determined by the model error variance σ_u^2 and the sampling variance v_i. The first term in (3), $\sigma_u^2(1 - h_i)$, is the inherent prediction error variance that would result if all model parameters were known. The second term in (3) accounts for additional error due to estimating β. Considerable attention has been given in the literature to augmenting (3) to reflect uncertainty due to estimating σ_u^2 while still assuming the v_i are known. Prasad and Rao (1990) and Datta and Lahiri (2000) provide asymptotic results while Berger (1985, pp. 190-193) provides results from a Bayesian approach. Many other papers have extended these results towards more general models (e.g., Booth and Hobert (1998) consider generalized linear mixed models) and to other approaches to accounting for uncertainty due to estimating σ_u^2 (e.g., Jiang, Lahiri, and Wan (2002) provide a jackknife approach).

Much less attention has been given to dealing with the fact that, in practice, the v_i are not known but are replaced in equations (2)–(5) by estimates $\hat{v}_i$. Typically the $\hat{v}_i$ are direct sampling variance estimators based on survey microdata (Wolter 1985) and, as such, are subject to errors ($\hat{v}_i \neq v_i$.) In fact, if the direct survey point estimates y_i are very imprecise due to small sample sizes for some or all areas (which is what motivates model-based small area estimation in the first place), the corresponding $\hat{v}_i$ can also be expected to be very imprecise due to the

*__Disclaimer:__ This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical, methodological, technical, or operational issues are those of the author and not necessarily those of the U.S. Census Bureau.

†U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

same sample size limitations. This suggests modeling the $\hat{v}_i$ to develop improved estimates of the v_i and to quantify their estimation error. It also suggests translating these results into improved measures of the MSE, or, from a Bayesian perspective, improved measures of the (posterior) uncertainty of Y_i. Section 2 discusses papers that have attempted to address these issues.

Section 3 provides some results giving a rough indication of the extent to which estimation error in the $\hat{v}_i$ can affect small area prediction results in regard to (i) increase in MSE relative to use of the true (unknowable) v_i, and (ii) misstatement of prediction MSE. These topics are considered both in a conditional sense (for given values of the $\hat{v}_i$) and in an unconditional sense (averaging over the distribution of the the $\hat{v}_i$).

Before proceeding a couple points are worth noting. First, the issues of concern here do not arise for unit level small area models (Rao 2003), such as the nested error regression model of Battese, Harter, and Fuller (1988). We can convert (1) to such a model by replacing the single index i by the double index ij, so that y_{ij} denotes the observation from subject (ultimate sampling unit) j in area i, except we leave u_i as is, so that it denotes a random effect common to all observations from area i. With such models sampling variability is reflected in the variation of y_{ij} within area i. The challenge is to specify a unit level model that adequately reflects the sampling variability, including important features of the sample design. This challenge is made more severe if the sample design is complex and/or the model is nonlinear. For example, while a number of papers have explored use of unit level generalized linear mixed models for small area estimation, the papers have generally assumed, effectively, that the data arose from simple random sampling (or basically they have ignored any consideration of the sample design). An exception to this limitation is the paper by Malec, Davis, and Cao (1999).

Second, as noted by Wang and Fuller (2003), the asymptotic MSE results given by Prasad and Rao (1990) are sufficiently general to cover the case where the sampling variances v_i (as well as $\text{Var}(u_i)$) depend on a finite number of parameters and, potentially, on some additional covariates. (This is also true of the asymptotic MSE results of Datta and Lahiri (2000).) In such a case more information accumulates about the sampling variances as the number of observations $m \to \infty$, and their parametric estimates should thus converge to the true values subject to an assumption of correct specification of the parametric form of the variances. This differs fundamentally from the situation where each individual variance estimate $\hat{v}_i$ is used for the true v_i, all of which are treated as distinct. However, even if the parametric specification of the v_i is reasonable these asymptotic results may not be very helpful in practice since they assume that the parameters determining the v_i are estimated using only the direct point estimates y_i as data (with no use made of direct survey variance estimates). Even if identifiability conditions hold for the v_i and $\text{Var}(u_i)$ in a strict mathematical sense, these quantities may be very weakly identified in a statistical sense with any finite amount of data $\{y_i\}$. In such cases the parameters determining the v_i and $\text{Var}(u_i)$ will be highly correlated and poorly estimated (Bell (1997) provides such an example), which may compromise practical application of the asymptotic MSE results.

2. Literature Review: Dealing with Estimation Error in Sampling Variances

We group the papers reviewed here into three subsections: (2.1) approximate MSE results; (2.2) small area modeling including sampling variance modeling; and (2.3) additional papers on modeling sampling variances.

2.1 Approximate MSE results

Wang and Fuller (2003, Theorem 1) provide an asymptotic result for the mean squared error of small area predictors from the model (1) when small area sampling variances are estimated. Let $\hat{Y}_i$ be the predictor for area i using estimated sampling variances $\hat{v}_i$. The $\hat{v}_i$ are assumed unbiased estimators of the v_i that are independent of the model and sampling errors u_i and e_i. In the notation being used here, under suitable assumptions their result can be written:

$$\text{MSE}(Y_i - \hat{Y}_i) \approx \sigma_u^2(1-h_i) + (1-h_i)^2 \mathbf{x}_i' V\left(\hat{\beta}\right)\mathbf{x}_i + (\sigma_u^2 + v_i)^{-3}\{\sigma_u^4 V(\hat{v}_i) + v_i^2 V_A(\hat{\sigma}_u^2)). \quad (6)$$

$V\left(\hat{\beta}\right)$ is the variance of $\hat{\beta}$, but differs from (5) because in Wang and Fuller (2003) $\hat{\beta}$ is assumed not to depend on the unknown variances σ_u^2 and v_i. See their paper for the expressions for this and for $V(\hat{v}_i)$ and $V_A(\hat{\sigma}_u^2)$. The expression (6) is analogous to the asymptotic MSE result of Prasad and Rao (1990), but with addition of the term $\sigma_u^4 V(\hat{v}_i)$ to reflect error in the estimates $\hat{v}_i$. Wang and Fuller derive two estimators of the MSE and examine their performance in a simulation study. The estimators perform well in many of the cases considered, but do poorly when σ_u^2 is quite small relative to the sampling variances v_i.

Rivest and Vandal (2003) provide an essentially similar MSE estimator, but obtained under the assumption that the $\hat{v}_i$ are approximately normally distributed. They provide some simulation results on its performance, showing some improvements over the MSE estimator of Prasad and Rao (1990), which ignores error in the $\hat{v}_i$.

The improvements are more pronounced as the degrees of freedom of the $\hat{v}_i$ (assumed distributed as χ^2 in their simulations) gets small (they use a minimum of 4). However, in most cases that they consider the bias of the Prasad-Rao MSE estimator is also small. Their simulations cover a relatively narrow range of values of v_i/σ_u^2 (smallest is 1/3, largest is 2.5) compared to Wang and Fuller (smallest is 1/4, largest is 160).

An interesting feature of Wang and Fuller's result is that it is asymptotic in both the number of small areas m and the degrees of freedom d of the $\hat{v}_i$. They show that the error of the approximation (6) is $O(r_{m,d})$ where $r_{m,d} = \max(m^{-1.5}, m^{-1}d^{-1}, d^{-1.5})$. One would not ordinarily think of the degrees of freedom of direct small area variance estimates being large, which their theorem suggests is needed to make the approximation error in (6) small. Similarly, Rivest and Vandal's assumption of approximate normality of the $\hat{v}_i$ obviously improves as d increases. However, even with relatively small values of d (8 and 17 for Wang and Fuller, down to 4 for Rivest and Vandal), the simulation results show fairly low bias of the MSE estimator in many cases, with the exception noted by Wang and Fuller of the cases where σ_u^2 is very small relative to the v_i.

Note also that papers mentioned in the next section that feature Bayesian treatments of small area models which include sampling variance models provide posterior variances reflecting uncertainty about the true sampling error variances.

2.2 Small area modeling including sampling variance modeling

Arora and Lahiri (1997) examined theoretically a unit level small area model with random area variances proposed by Kleffe and Rao (1992), but in analysis of an empirical example they used an area level model with direct sampling error variance estimates assumed unbiased, independent of the direct survey point estimates, and distributed proportional to a chi-squared random variable with known degrees of freedom. They then developed a Gibbs sampling scheme to make Bayesian inferences for the model. The model is related to that of Otto and Bell (1995) (which is mentioned in Section 2.3), though with a much simpler model for the sampling variances. They applied their model to data from the Consumer Expenditure Survey on milk consumption for 43 small areas and, drawing eight 12.5 percent samples from the survey data treated as a finite population, they compared direct and model-based small area estimators from these samples. The hierarchical Bayes model yielded lower MSE than the direct survey estimates for all eight samples, and lower MSE than an EBLUP for six of the eight samples. They didn't attempt to assess the contribution of uncertainty about the sampling error variances to their results.

You and Chapman (2006) analyzed an essentially similar area level model to that used by Arora and Lahiri (1997), applying it also to their data, as well as to small area data on amount of land planted with corn and soybeans taken from the paper of Battese, Harter, and Fuller (1988). You and Chapman provided results both for the model that assumed the $\hat{v}_i$ were distributed via $d_i\hat{v}_i/v_i \sim \chi^2_{d_i}$ (with the degrees of freedom d_i assumed equal to $n_i - 1$ where n_i is the sample size for area i), and for a model that assumed the $\hat{v}_i$ were equal to the v_i. Comparing posterior standard deviations or coefficients of variation for the two models showed, for the corn and soybean example, substantially larger uncertainty surrounding the small area predictions from the model with the v_i assumed unknown, but, for the milk consumption example, almost identical results for the two models. The latter result was due to the large small area sample sizes for that example, while for the corn and soybean example the small area samples were quite small.

Liu, Lahiri, and Kalton (2007) considered four alternative models for small area proportions. Two of these took sampling variances as known, while the other two parameterized the sampling variances as $[p_i(1-p_i)/n_i]deff_i$, with p_i denoting the unknown true proportion for area i, n_i the sample size, and $deff_i$ design effects that were estimated and treated as known. Since the p_i are unknown, the sampling variances were, to this extent, treated as unknown. They applied the models to data on all registered births for 2002 and made predictions about the prevalence of low birth weight for states from 1,000 samples drawn from the full data set. They examined coverage of 95 percent credible intervals, finding that the models whose sampling variances depended on the true (unknown, to the models) state proportions produced intervals with better coverage overall, though coverage rates for these and one of the other two models were observed to increase with increasing sample size. They suggested, "that the credible intervals are not adequately reflecting the effect of the greater precision of the direct estimates in the states with large sample sizes."

You (2008) considered a cross-sectional and time series model for estimated unemployment rates of small areas in Canada, with sampling variances of the direct estimates parameterized in the same way as in the model of Liu, Lahiri, and Kalton (2007), again with design effects estimated and then treated as known. He did not, however, compare results to those from a model that assumed sampling variances were known.

Bell and Otto (1992) and Bell (1995) investigated time series models with sampling error components, with the sampling error variances treated as unknown. A Bayesian approach was implemented via accept/reject sampling and used to produce posterior means and variances of the time series components, a prediction problem analogous to small area estimation. In the application considered the sampling errors were approximately uncorrelated over time, and were assumed to have constant relative variance over time, with logarithms taken of the time series so

the relative variance was taken as an approximation to the sampling variance in the log scale. Thus, the sampling variances depended on this one unknown parameter, of which multiple estimates were available. The Bayesian approach was used to reflect uncertainty about the common sampling relvariance and the other model parameters.

Nguyen, Bell, and Gomish (2002) applied a similar modeling approach to an application for which the sampling errors did not appear uncorrelated over time (a second order autoregressive model was used), and for which the sampling relvariances varied over time due to sample size fluctuations. Later, Bell (2005) provided some results from a Bayesian treatment of this model.

2.3 Additional papers on modeling sampling variances

Several recent papers by (mostly) staff of the U.S. Bureau of Labor Statistics have explored fitting generalized variance functions (GVFs) to direct survey variance estimates in a formal modeling context. Particularly relevant to our concerns here is the paper by Gershunskaya and Lahiri (2005). They examined alternative approaches to variance estimation for domain estimates in the Current Employment Statistics (CES) survey. In a simulation study with subsamples drawn from part of the CES sample they noted that design-based variance estimates have low bias but are unstable (very high CVs), while a synthetic model-based variance estimator was more stable but had substantial bias. They developed simple models for design-based variance estimates or their logarithms that permitted empirical Bayes smoothing of the variance estimates. They found the resulting variance estimates had low bias and were much more stable than the design-based variances.

Additional papers on GVF modeling of direct variance estimates, though not doing empirical Bayes smoothing, include Huff, Eltinge, and Gershunskaya (2002), Cho et al. (2002), and Eltinge, Cho, and Hinrichs (2002). An earlier paper by Valliant (1987) connected use of a generalized variance function (GVF) with sample design by showing that a commonly used GVF is consistent with a particular class of prediction models for estimating totals from stratified, two-stage cluster samples.

Otto and Bell (1995) developed a model with state random effects for sampling covariance matrices of CPS estimated poverty ratios. This model provides for a Bayesian or empirical Bayes smoothing of the direct sampling variance estimates analogous to what was done by Gershunskaya and Lahiri (2005).

3. Examining How Error in Sampling Variance Estimates Can Affect Small Area Predictions

To get a rough idea of how error in sampling variance estimates can affect small area predictions – point predictions and prediction error variances – we consider the simple case where the parameters β and σ_u^2 are known, leaving only the sampling error variances v_i as unknown parameters. This assumption also applies as the number of small areas m grows sufficiently large so that the estimation error of β and σ_u^2 becomes small. We start by computing the MSE treating the $\hat{v}_i$ as fixed (i.e., conditional on the $\hat{v}_i$). Let $\hat{Y}_i$ be given by (2) but using the estimated sampling variance $\hat{v}_i$, that is, with weight $\hat{h}_i = \sigma_u^2/(\sigma_u^2 + \hat{v}_i)$ on the direct estimate y_i. The MSE of $\hat{Y}_i$ can be obtained by writing $Y_i - \hat{Y}_i = (Y_i - \tilde{Y}_i) + (\tilde{Y}_i - \hat{Y}_i)$, noting that the error $Y_i - \tilde{Y}_i$ in the optimal predictor $\tilde{Y}_i$ is orthogonal to $\tilde{Y}_i - \hat{Y}_i$, which is a linear function of the data, so that $E[(Y_i - \hat{Y}_i)^2] = E[(Y_i - \tilde{Y}_i)^2] + E[(\tilde{Y}_i - \hat{Y}_i)^2]$. After a little algebra, we have that

$$E[(\tilde{Y}_i - \hat{Y}_i)^2 | \hat{v}_i] = (h_i - \hat{h}_i)^2(\sigma_u^2 + v_i).$$

We shall examine the percentage increase in MSE from using $\hat{Y}_i$ instead of $\tilde{Y}_i$. Since, with β and σ_u^2 known the MSE of $\tilde{Y}_i$ is given by the first term in (3), which can be written as $\sigma_u^2 v_i/(\sigma_u^2 + v_i)$, this percentage increase in MSE turns out to be

$$\begin{aligned} \text{MSE pct diff} &\equiv 100 \times \frac{\text{MSE}(Y_i - \hat{Y}_i) - \text{MSE}(Y_i - \tilde{Y}_i)}{\text{MSE}(Y_i - \tilde{Y}_i)} \\ &= \frac{E[(\tilde{Y}_i - \hat{Y}_i)^2 | \hat{v}_i]}{\sigma_u^2 v_i/(\sigma_u^2 + v_i)} \\ &= \frac{(h_i - \hat{h}_i)^2}{h_i(1 - h_i)}. \end{aligned} \tag{7}$$

We also examine the extent to which the MSE would be misstated by assuming that the $\hat{v}_i$ are the true sampling variances. The reported MSE would be, from (3), $\sigma_u^2(1 - \hat{h}_i)$, while the actual MSE is, from above, $\sigma_u^2(1 - h_i) +$

$(h_i - \hat{h}_i)^2(\sigma_u^2 + v_i)$. The percentage difference between these two is

$$\begin{aligned} \text{MSE relbias} &= 100 \times \left\{ \frac{\sigma_u^2(1-\hat{h}_i)}{\sigma_u^2(1-h_i) + (h_i - \hat{h}_i)^2(\sigma_u^2 + v_i)} - 1 \right\} \\ &= 100 \times \left\{ \frac{h_i(1-\hat{h}_i)}{h_i(1-h_i) + (h_i - \hat{h}_i)^2} - 1 \right\} \end{aligned} \tag{8}$$

upon dividing the numerator and denominator of the ratio by $\sigma_u^2 + v_i$ and simplifying. Notice that we can write $h_i = (1 + v_i/\sigma_u^2)^{-1}$ as a function of just the "noise-to-signal" ratio v_i/σ_u^2. Similarly, $\hat{h}_i = (1 + \hat{v}_i/\sigma_u^2)^{-1}$. Thus, both (7) and (8) can be computed given the ratios v_i/σ_u^2 and $\hat{v}_i/\sigma_u^2$.

We shall examine the MSE pct diff and MSE relbias for various multiplicative errors in $\hat{v}_i$ as an estimate of v_i. These can be specified in terms of the ratio $\hat{v}_i/v_i = (\hat{v}_i/\sigma_u^2)/(v_i/\sigma_u^2)$. For various degrees of *underestimation* of v_i we set $\hat{v}_i/v_i$ to one of the values $(.75, .50, .25)$, reflecting underestimation by 25, 50, or 75 percent. For various degrees of *overestimation* of v_i we use the reciprocal values $(4/3, 2, 4)$. We then computed MSE pct diff and MSE relbias for values of the true ratio v_i/σ_u^2 ranging from 1 to 50, and their corresponding reciprocals ranging from 1 down to .02. Results are plotted in Figures 1.a and 1.b. The dotted curves correspond to underestimation with the ratio $\hat{v}_i/v_i$ set to .75 (green), .50 (blue), or .25 (red). The solid curves correspond to overestimation by the reciprocal factors 4/3 (green), 2 (blue) and 4 (red). The x-axis of both plots, for v_i/σ_u^2, is on a log scale. Note that the MSE pct diff curves are symmetric in that the MSE pct diff for $(\hat{v}_i/\sigma_u^2, v_i/\sigma_u^2) = (r_1, r_2)$ is equal to that for $(\hat{v}_i/\sigma_u^2, v_i/\sigma_u^2) = (r_1^{-1}, r_2^{-1})$.

Examining first Figure 1.a we see that the increase in MSE from underestimating v_i by 25 or 50 percent, or overestimating v_i by multiplicative factors of 4/3 or 2, is not very large, being no more than about 10 percent for all values of v_i/σ_u^2. With the more extreme estimation errors by factors of 1/4 or 4, consequences for MSE are more severe. Note that when the sampling variance becomes larger than the model error variance ($v_i/\sigma_u^2 > 1$), underestimation of v_i is a more severe problem than overestimation. When the sampling variance is smaller than the model error variance ($v_i/\sigma_u^2 < 1$), the reverse is true. Focusing on underestimation of v_i, note that this implies that the weight, $\hat{h}_i = (1 + \hat{v}_i/\sigma_u^2)^{-1}$, given to the direct estimate y_i, exceeds the optimal weight $h_i = (1 + v_i/\sigma_u^2)^{-1}$. When $v_i/\sigma_u^2 < 1$ the optimal weight exceeds 1/2, and giving a still larger weight to y_i does not incur much increase in MSE. However, when $v_i/\sigma_u^2 > 1$ the optimal weight is less than 1/2, and putting substantially more weight than this on y_i can lead to a substantial increase in MSE. The MSE increase peaks around $v_i/\sigma_u^2 = 5$, at which point underestimating v_i by a factor of 1/4 leads to $\hat{h}_i = (1 + 5/4)^{-1} \approx .44$ compared to the optimal $h_i = (1 + 5)^{-1} \approx .17$, with about a 55 percent increase in MSE. In other words, substantially underestimating the sampling variance v_i when it is large can substantially increase MSE. Note also, however, that for the largest values shown of v_i/σ_u^2, even underestimating v_i at 1/4 its true value does not increase MSE so much. For example, if $v_i/\sigma_u^2 = 50$ the optimal weight on y_i is about .02, and if $\hat{v}_i = v_i/4$ the weight is about .074, still quite small, so the increase in MSE is less than 20 percent. Parallel comments clearly apply to overestimation of v_i.

Turning to Figure 1.b we see that over- or underestimation of v_i has more substantial effects on error in the reported MSE than it did on the true MSE. The effects are largest for small values of v_i/σ_u^2. Note that if $\hat{v}_i/\sigma_u^2$ is small (which it will be when v_i/σ_u^2 is small unless v_i is substantially overestimated) $\hat{h}_i$ is close to 1 and $\hat{Y}_i$ is close to y_i, whose MSE is v_i, so that errors in $\hat{v}_i$ translate directly into errors in the reported MSE. With overestimation of v_i by factors of 2 or 4 the bias in the reported MSE is very large for small v_i/σ_u^2, but declines fairly rapidly as v_i/σ_u^2 increases, becoming much less important when $v_i/\sigma_u^2 > 2$. With underestimation of v_i substantial bias in the reported MSE persists into large values of v_i/σ_u^2. When v_i/σ_u^2 gets sufficiently large then, even when v_i is underestimated to some extent, we have $\hat{h}_i$ near zero, $\hat{Y}_i$ close to the regression prediction $\mathbf{x}_i'\hat{\beta}$, and the MSE becomes the variance of the regression prediction error, which is $\sigma_u^2 + \mathbf{x}_i'\text{Var}(\hat{\beta})\mathbf{x}_i$. This depends on the v_i only through $\text{Var}(\hat{\beta})$, which may not be severely affected by error in any individual v_i (though here we are actually assuming $\text{Var}(\hat{\beta})$ is small due to m being large).

Notice that underestimation of v_i is the more serious problem (in regard to both increased MSE and bias in reported MSE) when v_i/σ_u^2 is large, while overestimation of v_i is the more serious problem when v_i/σ_u^2 is small. However, small values of v_i generally result from large sample sizes, which also lead to more precise variance estimates $\hat{v}_i$, making substantial error in the $\hat{v}_i$ less likely. Thus, situations where overestimation of v_i causes serious problems seem generally less likely to arise than situations where underestimation of v_i causes serious problems.

We now examine results for unconditional MSE obtained by assuming a distribution for the $\hat{v}_i$, integrating with respect to this distribution to get $E[(h_i - \hat{h}_i)^2]$ and $E(\hat{h}_i)$, and substituting these quantities into (7) and (8). These results are consistent with the approximate unconditional MSE results of Wang and Fuller (2003) and Rivest and Vandal (2003) discussed in Section 2, though again under the assumption that m is sufficiently large that σ_u^2 and β are essentially known. For this purpose we assume a χ_d^2 distribution for $d\hat{v}_i/v_i$. We choose d so that the lower 5

percent point of the χ^2_d/d distribution for $\hat{v}_i/v_i$ roughly corresponds to the (.75, .50, .25) underestimation factors used in Figures 1.a and 1.b. The corresponding 95 percent points of the χ^2_d/d distribution (Chemical Rubber Company 1971) are then substantially less than the reciprocals of (.75, .50, .25), as can be seen from Table 1, which also shows the coefficients of variation (CVs) of the distributions.

Table 1. 5% and 95% points and CVs for the χ^2_d/d distribution

d	5% point	95% point	CV
6	.27	2.10	.82
16	.50	1.64	.50
80	.75	1.27	.22

While these assumed distributions make a loose connection with the results in Figures 1.a and 1.b in relation to underestimation of $\hat{v}_i$ (the more serious concern), these assumptions are just for illustration, and clearly other assumptions could be used. In particular, if we instead assumed a lognormal distribution with median 1 for $\hat{v}_i/v_i$ (the mean would then exceed 1), then its 5 percent and 95 percent points would be reciprocals of one another, more analogous to the calculations for Figures 1.a and 1.b.

Results from these unconditional calculations are shown in Figures 1.c and 1.d, with the red, blue, and green curves corresponding to the values 6, 16, and 80 for d. From Figure 1.c we see that the increase in unconditional MSE when $d = 16$ or $d = 80$ is quite mild, and is not very large (less than 10 percent) even for $d = 6$. What increases there are in unconditional MSE are largest when v_i/σ_u^2 exceeds 1. (This is due to the nature of the χ^2_d/d distribution, which makes severe underestimation of v_i more likely than severe overestimation, in a multiplicative sense, as can be seen from Table 1.) Examining Figure 1.d we see only slightly larger effects on the bias of the reported MSE, with estimation error in $\hat{v}_i$ leading to downward biases in the reported MSE (consistent with the asymptotic result (6) of Wang and Fuller (2003)). Again, the effects are more pronounced when v_i/σ_u^2 exceeds 1, and are larger for smaller values of d, reflecting larger amounts of estimation error in $\hat{v}_i$.

Clearly the most serious concerns arising from the results presented here are those of Figure 1.b on the percent bias in the reported conditional MSE due to error in $\hat{v}_i$. Whether or not one takes comfort in the relatively mild effects of estimation error in $\hat{v}_i$ on the unconditional MSE and on the bias in the reported unconditional MSE, in contrast to the potentially larger effects for a specific observed sample, probably depends on whether one views things from a conditional or unconditional perspective.

REFERENCES

Arora, Vipin and Lahiri, Partha (1997), "On the Superiority of the Bayesian Method Over the BLUP in Small Area Estimation Problems," *Statistica Sinica*, **7**, 1053-1063.

Battese, George E., Harter, Rachel M., and Fuller, Wayne A. (1988), "An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data," *Journal of the American Statistical Association*, **83**, 28-36.

Bell, William R. (1995), "Bayesian Sampling Error Modelling with Application," Statistical Society of Canada, Proceedings of the Survey Methods Section, pp. 19-28.

Bell, William R. (1997), "A County CPS Model with Census Residuals," internal note, Statistical Research Division, U.S. Census Bureau, July, 23, 1997.

Bell, William R. (2005), "Some Consideration of Seasonal Adjustment Variances," Proceedings of the American Statistical Association, Survey Research Methods Section, [CD-ROM], 2747-2758.

Bell, William R. and Otto, Mark C. (1992), "Bayesian Assessment of Uncertainty in Seasonal Adjustment with Sampling Error Present," Research Report 92/12, Statistical Research Division, U.S. Census Bureau, available at http://www.census.gov/srd/papers/pdf/rr92-12.pdf.

Berger, James O. (1985), *Statistical Decision Theory and Bayesian Analysis*, New York: Springer-Verlag.

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

Chemical Rubber Company, (1971), *Basic Statistical Tables*, ed. William H. Beyer, Cleveland: The Chemical Rubber Company.

Cho, Moon, Eltinge, John, Gershunskaya, Julie and Hud', Larry (2002), "Evaluation of Generalized Variance Function Estimators for the U.S. Current Employment Survey," Proceedings of the American Statistical Association, Survey Research Methods Section, 534-539.

Datta, Gauri S. and Lahiri, Partha (2000), "A Unified Measure of Uncertainty of Estimated Best Linear Unbiased Predictors in Small-Area Estimation Problems," *Statistica Sinica*, **10**, 613-628.

Eltinge, John, Cho, Moon, and Hinrichs, Paul (2002), "Use of Generalized Variance Functions in Multivariate Analysis," Proceedings of the American Statistical Association, Survey Research Methods Section, 904-912.

Fay, Robert E. and Herriot, Roger 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.

Gershunskaya, Julie B. and Lahiri, Partha (2005) "Variance Estimation for Domains in the U.S. Current Employment Statistics Program," *Proceedings of the American Statistical Association, Survey Research Methods Section*, 3044-3051.

Hud', Larry, Eltinge, John and Gershunskaya, Julie (2002), "Exploratory Analysis of Generalized Variance Function Models for the U.S. Current Employment Survey," Proceedings of the American Statistical Association, Survey Research Methods Section, 1519-1524.

Jiang, Jiming, Lahiri, Partha, and Wan, Shu-Mei (2002), "A Unified Jackknife Theory for Empirical Best Prediction with M-estimation," *The Annals of Statistics*, **30**, 1782-1810.

Kleďe, J. and J. N. K. Rao (1992), "Estimation of Mean Square Error of Empirical Best Linear Unbiased Predictors under a Random Error Variance Linear Model," *Journal of Multivariate Analysis*, **43**, 1-15.

Liu, Benmei, Lahiri, Partha, and Kalton, Graham (2007), "Hierarchical Bayes Modeling of Survey-Weighted Small Area Proportions," Proceedings of the American Statistical Association, Survey Research Methods Section, 3181-3186.

Malec, Donald, Davis, William W., and Cao, Xin (1999), "Model-Based Small Area Estimates of Overweight Prevalence Using Sample Selection Adjustment," *Statistics in Medicine*, **18**, 3189-3200.

Nguyen, Thuy Tran T., Bell, William R., and Gomish, James M. (2002), "Investigating Model-Based Time Series Methods to Improve Estimates from Monthly Value of Construction Put-in-Place Surveys," Proceedings of the American Statistical Association, Survey Research Methods Section [CD-ROM], 2470-2475.

Otto, Mark C. and Bell, William R. (1995), "Sampling Error Modelling of Poverty and Income Statistics for States," *Proceedings of the American Statistical Association, Government Statistics Section*, pp. 160-165.

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

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

Rivest, Louis-Paul and Vandal, Nathalie (2003), "Mean Squared Error Estimation for Small Areas When the Small Area Variances are Estimated," in *Proceedings of the International Conference on Recent Advances in Survey Sampling*, ed. J.N.K. Rao.

U.S. Bureau of Labor Statistics and U.S. Census Bureau (2002), "Current Population Survey: Design and Methodology," Technical Paper 63RV, available at http://www.census.gov/prod/2002pubs/tp63rv.pdf.

Valliant, Richard (1987), "Generalized Variance Functions in Stratified Two-Stage Sampling," *Journal of the American Statistical Association*, **82**, 499-508.

Wang, Junyuan and Fuller, Wayne A. (2003), "The Mean Squared Error of Small Area Predictors Constructed with Estimated Error Variances," *Journal of the American Statistical Association*, **98**, 716-723.

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

You, Yong (2008), "An Integrated Modeling Approach to Unemployment Rate Estimation for Sub-Provincial Areas of Canada," *Survey Methodology*, **34**, 19-27.

You, Yong and Chapman, Beatrice (2006), "Small Area Estimation Using Area Level Models and Estimated Sampling Variances," *Survey Methodology*, **32**, 97-103.

Fig. 1. Percent difference in MSE and percent bias in reported MSE from using estimated versus true sampling variance