Accounting for Uncertainty About Variances in Small Area Estimation

William R. Bell, Room 3000-4, U.S. Bureau of the Census, Washington, DC 20233-9100

wbell@census.gov

1. Introduction

The model of Fay and Herriot (1979) for small area estimation can be written

$$
\begin{aligned}
y_i &= Y_i + e_i \qquad i = 1, \ldots, m & (1) \\
&= (\mathbf{x}_i'\boldsymbol{\beta} + u_i) + e_i & (2)
\end{aligned}
$$

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 $1 \times r$ row vectors of regression variables for area i, and $\boldsymbol{\beta}$ is the corresponding vector of regression parameters.

From (2), letting $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_u^2 + v_i)$, $\boldsymbol{\beta}$ can be estimated by generalized least squares (GLS): $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$ with $\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$, where $\mathbf{y} = (y_1, \ldots, y_m)'$, and $\mathbf{X}$ is $m \times r$ with rows $\mathbf{x}_i'$. Then the best linear unbiased predictors (BLUPs) of the Y_i can be formed and their error variances obtained from

$$
\begin{aligned}
\widehat{Y}_i &= h_i y_i + (1 - h_i)\mathbf{x}_i'\widehat{\boldsymbol{\beta}} & (3) \\
\mathrm{Var}(Y_i - \widehat{Y}_i) &= \sigma_u^2(1 - h_i) + (1 - h_i)^2 \mathbf{x}_i' \mathrm{Var}\left(\widehat{\boldsymbol{\beta}}\right) \mathbf{x}_i & (4)
\end{aligned}
$$

where $h_i = \sigma_u^2/(\sigma_u^2 + v_i)$. From (3), the smoothed estimate $\widehat{Y}_i$ is a weighted average of the regression prediction $\mathbf{x}_i'\widehat{\boldsymbol{\beta}}$ and the direct estimate y_i. The first term in (4), $\sigma_u^2(1 - h_i)$, is the inherent prediction error variance that would result if all model parameters were known. The second term in (4) accounts for additional error due to estimating $\boldsymbol{\beta}$. Given the v_i, (4) can be augmented to reflect uncertainty about σ_u^2 using asymptotic formulas (Prasad and Rao 1990, Datta and Lahiri 1997) or a Bayesian approach (Berger 1985, pp. 190-193). When only point estimates y_i and variances v_i are available, uncertainty about sampling error variances is generally ignored, though Bell and Otto (1992) address this problem for a time series application via a Bayesian model-based approach.

This paper considers different approaches to dealing with uncertainty about σ_u^2 in the context of a particular application: estimating annual poverty rates of school-aged (5-17) children for the states of the U.S. using data from the Current Population Survey (CPS). For this problem Fay and Train (1997) developed a Fay-Herriot model for each year where, for each of $m = 51$ "states" i (including the District of Columbia as a "state"), y_i is the direct CPS estimate, Y_i the true poverty rate, and $\mathbf{x}_i$ includes a constant term and three variables derived from administrative sources. (Actually, ratios differing slightly from true poverty rates were modeled.) U.S. Internal Revenue Service income tax return files supplied two variables: an analogue to state child poverty rates and also state rates of nonfiling for income taxes. Data from the U.S. Department of Agriculture were used to develop a variable reflecting state participation rates in the food stamp poverty assistance program. In addition, $\mathbf{x}_i$ includes the residual from regressing 5-17 state poverty rates from the previous (1990) decennial census on the other regression variables for 1989 (the census income reference year). The v_i were obtained from a sampling error model of Otto and Bell (1995) that involved fitting a generalized variance function (GVF) to five years of direct variance and covariance estimates for each state produced by Fay and Train (1995). This application is an important component of the Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program. For information, see the SAIPE web site at http://www.census.gov/hhes/www/saipe.html.

Section 2 of this paper examines, in the context of the Fay and Train (1997) model, different approaches to dealing with uncertainty about σ_u^2 (given the v_i) and their effects on prediction error variances. Future work will explore a Gibbs sampling scheme to also recognize uncertainty about sampling error variances using the model of Otto and Bell (1995).

2. Accounting for Uncertainty About the Model Error Variance (σ_u^2)

Three estimation approaches are considered here: maximum likelihood (ML), restricted ML (REML), and the less-familiar mean likelihood (MEL). A Bayesian analysis is also explored. First, note REML maximizes the restricted likelihood (Harville 1977, p. 325)

$$L(\sigma_u^2) \propto |\mathbf{\Sigma}|^{-\frac{1}{2}} \left|\mathbf{X}'\Sigma^{-1}\mathbf{X}\right|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})'\Sigma^{-1}(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})} \tag{5}$$

where $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\sigma_u^2)$ is given by GLS. Omitting the term $\left|\mathbf{X}'\Sigma^{-1}\mathbf{X}\right|^{-1/2}$ from (5) gives the concentrated likelihood (for σ_u^2) maximized by ML. Also, (5) normalized to integrate to 1 is the Bayesian posterior density of σ_u^2 under the flat prior $p(\beta, \sigma_u^2) \propto$ constant (Berger 1985, p. 192). The corresponding posterior mean of σ_u^2 is the same as the mean likelihood estimate (Barnard 1949).

For the Fay and Train (1997) model of U.S. 5-17 year-old state poverty rates, Table 1 shows the three estimates of σ_u^2 for 1989–1993. Focusing first on the left half of Table 1, note that the ML and REML estimates are both zero in the first four years.

Table 1. Alternative Estimates of σ_u^2 for Five Years

	Updating v_i to Convergence			No Updating of v_i		
year	ML	REML	MEL	ML	REML	MEL
1989	0	0	1.7	4.6	4.9	6.1
1990	0	0	2.2	1.9	2.5	3.7
1991	0	0	1.6	0	0	1.6
1992	0	0	1.6	0	0	1.4
1993	.4	1.7	3.4	3.3	2.1	3.6

Having $\widehat{\sigma}_u^2 = 0$ has several unreasonable implications. First, it implies that if the Y_i were observed (if the CPS were a complete census every year), then the model would fit this data perfectly. (Note: The 1990 census data are not the Y_i for 1989 because of CPS-census measurement differences.) Second, since $\widehat{\sigma}_u^2 = 0$ implies $h_i = 0$ for all i, (3) implies that each $\widehat{Y}_i$ is just the regression prediction, $\mathbf{x}_i'\widehat{\boldsymbol{\beta}}$; the direct estimates y_i get no weight. Third, $\widehat{\sigma}_u^2 = 0$ implies that the first term on the right hand side of (4) is zero, and the prediction error variance comes entirely from the error in estimating $\boldsymbol{\beta}$. These results tend to look unreasonable, as will be seen later in Table 2.

Getting $\widehat{\sigma}_u^2 = 0$ for ML could motivate consideration of REML, which is intended to remove the downward asymptotic bias of ML (Datta and Lahiri 1997, p. 8). Table 1 shows that in this application, however, REML is of little help. The mean likelihood estimates, or Bayesian posterior means of σ_u^2, look more reasonable. (These were computed for Table 1 as $\int \sigma_u^2 L(\sigma_u^2) d\sigma_u^2 / \int L(\sigma_u^2) d\sigma_u^2$, with the integrations done numerically by Simpson's rule over 100 equal subintervals of $\sigma_u^2 \in [0, 20]$, an interval judged from graphs to contain essentially all the posterior probability for σ_u^2 for all years.) The reason for the differences between the estimators of σ_u^2 is easy to see from graphs (not shown) of the marginal posterior density ($L(\sigma_u^2) / \int L(\sigma_u^2) d\sigma_u^2$), which reveal a long right tail in all years. Since the marginal posteriors are not concentrated near $\sigma_u^2 = 0$, the posterior means substantially exceed the posterior modes (mean likelihood estimators exceed REML estimators).

The estimation scheme used by Fay and Train (1997) involved iteratively updating the v_i given each new estimate of $(\boldsymbol{\beta}, \sigma_u^2)$. If superscript (k) denotes the kth iteration, the update of v_i used was $v_i^{(k)} = v_i^{(0)}[\mathbf{x}_i'\widehat{\boldsymbol{\beta}}^{(k)}(1 - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}^{(k)})]/[y_i(1 - y_i)]$, where $v_i^{(0)}$ are the original estimated v_i from the sampling error model of Otto and Bell (1995). The idea was to adjust the v_i at each iteration to be consistent with the current estimate $\mathbf{x}_i'\widehat{\boldsymbol{\beta}}$. To find $\widehat{\boldsymbol{\beta}}^{(k)}$ and $\widehat{\sigma}_u^{2(k)}$ the $v_i^{(k-1)}$ were used. Convergence was effectively achieved in two iterations. For comparison, the right half of Table 1 shows the estimates of σ_u^2 without updating the v_i. The results are different in some cases, though some zero estimates for σ_u^2 still occur. For the remainder of this paper, results from updating v_i to convergence are used.

Tables 2 and 3 show some alternative prediction error variances for 1992 (when $\widehat{\sigma}_u^2 = 0$ for ML and REML) and for 1993, respectively. Also shown are CPS sample sizes n_i (number of households in the March CPS sample), CPS direct poverty rate estimates y_i, and direct sampling variance

estimates $v_i^{(0)}$. Results are shown for four states in increasing order of $v_i^{(0)}$: California (CA), the largest state with the largest sample size and lowest direct variance; North Carolina (NC); Indiana (IN); and Mississippi (MS). The tables show variances (ML[1], REML[1], and MEL[1]) obtained by plugging $\widehat{\sigma}_u^2$ (and corresponding fully updated v_i) into (4) for $\widehat{\sigma}_u^2$ given by ML, REML, and MEL. For ML and REML the tables also show prediction error variances (ML[2] and REML[2]) augmented as in Datta and Lahiri (1997) to asymptotically account for error in estimating σ_u^2. The tables also show two "Bayesian posterior variances" described later.

Table 2. Alternative Prediction Error Variances for Four States for 1992

state	n_i	y_i	$v_i^{(0)}$	ML[1]	ML[2]	REML[1]	REML[2]	MEL[1]	Bayes[1]	Bayes[2]
CA	4,927	20.9	1.9	1.3	3.6	1.3	2.8	1.5	1.4	1.4
NC	2,400	23.0	5.5	.6	2.0	.6	1.2	1.6	1.4	2.0
IN	670	11.8	9.3	.3	1.4	.3	.6	1.8	1.6	1.7
MS	796	29.6	12.4	2.8	3.8	2.8	3.0	4.1	3.9	4.0

Table 3. Alternative Prediction Error Variances for Four States for 1993

state	n_i	y_i	$v_i^{(0)}$	ML[1]	ML[2]	REML[1]	REML[2]	MEL[1]	Bayes[1]	Bayes[2]
CA	4,639	23.8	2.3	1.5	3.2	1.6	2.2	1.7	1.7	1.7
NC	2,278	17.0	4.5	1.0	2.4	1.7	2.2	2.2	2.0	2.0
IN	650	10.3	8.5	.8	1.9	1.8	2.2	2.9	2.7	3.0
MS	747	30.5	13.6	3.2	4.3	4.2	4.5	5.2	5.0	5.1

First consider ML[1] and REML[1] in 1992, which are the same since $\widehat{\sigma}_u^2 = 0$ for both. When $\sigma_u^2 = 0$, (4) reduces to $\mathrm{Var}(Y_i - \widehat{Y}_i) = \mathbf{x}_i' \mathrm{Var}(\widehat{\boldsymbol{\beta}})\mathbf{x}_i$, and variation in (4) over states results solely from variations in the regression variables $\mathbf{x}_i$. Hence, the small values for NC and IN, despite their having smaller sample sizes and higher sampling variances than CA. In fact, many other states have values for (4) lower than that for CA. While these results would not be unexpected if we really believed $\sigma_u^2 = 0$, since $\sigma_u^2 = 0$ seems questionable so do these prediction error variances. Now comparing the ML[1] results from 1992 and 1993, we see substantial increases in 1993 for NC and IN. Similar large increases occur for many other states. In general, the differences between the ML[1] results in the two years seem overly large and not very plausible (suggesting problems particularly for 1992). The REML[1] results for 1993 show even more dramatic increases due to the larger REML estimate of $\widehat{\sigma}_u^2 = 1.7$ for 1993, and in contrast cast doubt on the 1993 ML[1] results.

Augmenting the ML and REML prediction variances as in Datta and Lahiri (1997) to reflect error in estimating σ_u^2 (ML[2] and REML[2] results) yields large increases, suggesting that ignoring this term can significantly underestimate prediction error variance. Note the largest contributions from estimating σ_u^2 go to the states with the lowest sampling variances. This makes some sense as the lower v_i is the more weight goes to the direct estimate y_i in (3) when $\sigma_u^2 > 0$, so uncertainty about σ_u^2 means more to states with fairly precise direct estimates. However, this also means that the unappealing pattern of many states having smaller prediction error variances than CA persists in ML[2] in both years and REML[2] in 1992.

Plugging the much larger mean likelihood estimates of σ_u^2 into (3) produces much larger prediction error variances than those from ML[1] and REML[1] for NC, IN, and MS (and for many other states not shown). It also yields a more intuitively appealing pattern with prediction error variance increasing with sampling variance.

Bayesian posterior variances can be computed as

$$\mathrm{Var}(Y_i|\mathbf{y}) = E[\mathrm{Var}(Y_i|\mathbf{y}, \sigma_u^2)] + \mathrm{Var}[E(Y_i|\mathbf{y}, \sigma_u^2)] \tag{6}$$

where the outer expectation and variance on the right hand side are taken over the marginal posterior distribution of σ_u^2. These were computed by Simpson's rule in the same manner as the posterior means of σ_u^2 discussed above. Bayes[1] in Tables 2 and 3 denotes $E[\mathrm{Var}(Y_i|\mathbf{y}, \sigma_u^2)]$, while Bayes[2] denotes $\mathrm{Var}(Y_i|\mathbf{y})$. (The results shown are still conditional on the sampling variances v_i, set at their fully updated values from the mean likelihood estimation.) Note that Bayes[1] is fairly

close to MEL[1], i.e., averaging $\text{Var}(Y_i|\mathbf{y}, \sigma_u^2)$ over the posterior of σ_u^2 gives about the same result as evaluating $\text{Var}(Y_i|\mathbf{y}, \sigma_u^2)$ at the posterior mean of σ_u^2. Given this, $\text{Var}[E(Y_i|\mathbf{y}, \sigma_u^2)] = \text{Bayes}^2 - \text{Bayes}^1$ can be thought of as accounting for uncertainty about σ_u^2, and as a Bayesian analogue to the term augmenting (4) to account for error in estimating σ_u^2. For most states $\text{Var}[E(Y_i|\mathbf{y}, \sigma_u^2)]$ is quite small, so $\text{Var}(Y_i|\mathbf{y})$ is close to $E[\text{Var}(Y_i|\mathbf{y}, \sigma_u^2)]$, and to MEL[1]. Note, however, the large difference between Bayes[1] and Bayes[2] for NC in 1992. This arises because the regression prediction for NC in 1992 (which varies little over the different estimates of σ_u^2) is $\mathbf{x}_i'\widehat{\boldsymbol{\beta}} = 17.7$, which differs substantially from the direct estimate $y_i = 23.0$. When such large differences between the direct estimate and the regression prediction occur, the conditional mean ($E(Y_i|\mathbf{y}, \sigma_u^2)$ as given by (3)) is sensitive to variation in σ_u^2. Hence, the posterior variances reflect this. A similar, though less pronounced effect occurs for IN in 1993 when $\mathbf{x}_i'\widehat{\boldsymbol{\beta}} = 15.3$ versus $y_i = 10.3$. Such occurrences are rare in this example, and when $\mathbf{x}_i'\widehat{\boldsymbol{\beta}}$ and y_i are close, $\text{Var}[E(Y_i|\mathbf{y}, \sigma_u^2)]$ is close to zero. Note the difference in this result from the frequentist results, which do not depend in such a direct way on the realized data values.

Without overinterpreting the results from this particular example, they nonetheless show the potential difficulties for the frequentist approaches when the model error variance is estimated at or near zero. By averaging over the posterior of σ_u^2, the Bayesian approach avoids unreasonable results from fixing σ_u^2 at a single value near 0, and gives more intuitively plausible results.

References

[1] Barnard, George A. (1949), "Statistical Inference," *Journal of the Royal Statistical Society*, **B**, **11**, p. 116.

[2] 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, Bureau of the Census.

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

[4] Datta, Gauri S. and Lahiri, Partha (1997), "A Unified Measure of Uncertainty of Estimated Best Linear Unbiased Predictor in Small-Area Estimation Problems," Technical Report 97-7, University of Georgia, Department of Statistics.

[5] Fay, R. E. and Herriott, R. A. (1979), "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data," *Journal of the American Statistical Association*, **74**, 269-277.

[6] Fay, Robert E. and Train, George F. (1995), "Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties," American Statistical Association, Proceedings of the Section on Government Statistics.

[7] Fay, Robert E. and Train, George F. (1997), "Small Domain Methodology for Estimating Income and Poverty Characteristics for States in 1993," American Statistical Association, Proceedings of the Social Statistics Section, 183-188.

[8] Harville, David A. (1977), "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems (with discussion)," *Journal of the American Statistical Association*, **72**, 320-340.

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

[10] 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*, **78**, 47-59.