1. Introduction

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

\[ y_i = Y_i + e_i \quad i = 1, \ldots, m \]  

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 \( N(0, \sigma_u^2) \), the \( u_i \) are small area random effects (model errors) distributed i.i.d. \( N(0, \sigma_w^2) \), the \( x_i' \) are \( r \times 1 \) row vectors of regression variables for area \( i \), and \( \beta \) is the corresponding vector of regression parameters.

From (2), letting \( \Sigma = \text{diag}(\sigma_u^2 + v_i) \), \( \beta \) can be estimated by generalized least squares (GLS):

\[ \hat{\beta} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} y \]

with \( \text{Var}(\hat{\beta}) = (X' \Sigma^{-1} X)^{-1} \), where \( y = (y_1, \ldots, y_m)' \), and \( X \) is \( m \times r \) with rows \( x_i' \). Then the best linear unbiased predictors (BLUPs) of the \( Y_i \) can be formed and their error variances obtained from

\[
\hat{Y}_i = h_i y_i + (1 - h_i)x_i' \hat{\beta} \\
\text{Var}(Y_i - \hat{Y}_i) = \sigma_u^2 (1 - h_i) + (1 - h_i)^2 x_i' \text{Var}(\hat{\beta}) x_i
\]

where \( h_i = \sigma_u^2 / (\sigma_u^2 + v_i) \). From (3), the smoothed estimate \( \hat{Y}_i \) is a weighted average of the regression prediction \( x_i' \hat{\beta} \) 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 \( \beta \). Given the \( v_i \), (4) can be augmented to reflect uncertainty about \( \sigma_u^2 \) 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 \( \sigma_u^2 \) 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 \( x_i \) 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 nonfilers 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, \( x_i \) 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 \( \sigma_u^2 \) (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 ($\sigma_u^2$)

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)

$$L(\sigma_u^2) \propto |\Sigma|^{-\frac{1}{2}} |X'\Sigma^{-1}X|^{-\frac{1}{2}} e^{-\frac{1}{2} (y - X\hat{\beta})' \Sigma^{-1} (y - X\hat{\beta})}$$

where $\hat{\beta} = \hat{\beta}(\sigma_u^2)$ is given by GLS. Omitting the term $|X'\Sigma^{-1}X|^{-1/2}$ from (5) gives the concentrated likelihood (for $\sigma_u^2$) maximized by ML. Also, (5) normalized to integrate to 1 is the Bayesian posterior density of $\sigma_u^2$ under the flat prior $p(\beta, \sigma_u^2) \propto$ constant (Berger 1985, p. 192). The corresponding posterior mean of $\sigma_u^2$ 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 $\sigma_u^2$ 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>
<thead>
<tr>
<th>Year</th>
<th>ML</th>
<th>REML</th>
<th>MEL</th>
<th>ML</th>
<th>REML</th>
<th>MEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989</td>
<td>0</td>
<td>0</td>
<td>1.7</td>
<td>4.6</td>
<td>4.9</td>
<td>6.1</td>
</tr>
<tr>
<td>1990</td>
<td>0</td>
<td>0</td>
<td>2.2</td>
<td>1.9</td>
<td>2.5</td>
<td>3.7</td>
</tr>
<tr>
<td>1991</td>
<td>0</td>
<td>0</td>
<td>1.6</td>
<td>0</td>
<td>0</td>
<td>1.6</td>
</tr>
<tr>
<td>1992</td>
<td>0</td>
<td>0</td>
<td>1.6</td>
<td>0</td>
<td>0</td>
<td>1.4</td>
</tr>
<tr>
<td>1993</td>
<td>.4</td>
<td>1.7</td>
<td>3.4</td>
<td>3.3</td>
<td>2.1</td>
<td>3.6</td>
</tr>
</tbody>
</table>

Having $\hat{\sigma}_u^2 = 0$ 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 $\hat{\sigma}_u^2 = 0$ implies $h_i = 0$ for all $i$, (3) implies that each $\tilde{Y}_i$ is just the regression prediction, $X_i\hat{\beta}$; the direct estimates $y_i$ get no weight. Third, $\hat{\sigma}_u^2 = 0$ 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 $\beta$. These results tend to look unreasonable, as will be seen later in Table 2.

Getting $\hat{\sigma}_u^2 = 0$ 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 $\sigma_u^2$, are more reasonable. (These were computed for Table 1 as $\int \sigma_u^2 L(\sigma_u^2) d\sigma_u^2$ where 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 $\sigma_u^2$ for all years.) The reason for the differences between the estimators of $\sigma_u^2$ 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 $(\beta, \sigma_u^2)$. If superscript $(k)$ denotes the $k$th iteration, the update of $v_i$ used was $v_i^{(k)} = v_i^{(0)}(1 - X_i\hat{\beta}^{(k)})/[y_i(1 - y_i)]$, 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 $X_i\hat{\beta}$.

Convergence was effectively achieved in two iterations. For comparison, the right half of Table 1 shows the estimates of $\sigma_u^2$ without updating the $v_i$. The results are different in some cases, though some zero estimates for $\sigma_u^2$ 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 $\hat{\sigma}_u^2 = 0$ 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 $\hat{\sigma}_u^2$ (and corresponding fully updated $v_i$) into (4) for $\hat{\sigma}_u^2$ 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 $\sigma_u^2$. The tables also show two “Bayesian posterior variances” described later.

First consider ML$^1$ and REML$^1$ in 1992, which are the same since $\hat{\sigma}_u^2 = 0$ for both. When $\sigma_u^2 = 0$, (4) reduces to $\text{Var}(Y_i - \bar{Y}) = x_i^* \text{Var} (\hat{\beta}) x_i$, and variation in (4) over states results solely from variations in the regression variables $x_i$. 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, IN, and MS (and 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 $\hat{\sigma}_u^2 = 1.7$ 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 $\sigma_u^2$ (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 $\sigma_u^2$ 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 $\sigma_u^2$ 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 $\sigma_u^2$ 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

$$\text{Var}(Y_i|y) = E[\text{Var}(Y_i|y, \sigma_u^2)] + \text{Var}[E(Y_i|y, \sigma_u^2)]$$

(6)

where the outer expectation and variance on the right hand side are taken over the marginal posterior distribution of $\sigma_u^2$. These were computed by Simpson’s rule in the same manner as the posterior means of $\sigma_u^2$ discussed above. Bayes$^1$ in Tables 2 and 3 denotes $E[\text{Var}(Y_i|y, \sigma_u^2)]$, while Bayes$^2$ denotes $\text{Var}(Y_i|y)$. (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

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

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

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

<table>
<thead>
<tr>
<th>state</th>
<th>$n_i$</th>
<th>$y_i$</th>
<th>$v_i^{(0)}$</th>
<th>ML$^1$</th>
<th>ML$^2$</th>
<th>REML$^1$</th>
<th>REML$^2$</th>
<th>MEL$^1$</th>
<th>Bayes$^1$</th>
<th>Bayes$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>4,639</td>
<td>23.8</td>
<td>2.3</td>
<td>1.5</td>
<td>3.2</td>
<td>1.6</td>
<td>2.2</td>
<td>1.7</td>
<td>1.7</td>
<td>1.7</td>
</tr>
<tr>
<td>NC</td>
<td>2,278</td>
<td>17.0</td>
<td>4.5</td>
<td>1.0</td>
<td>2.4</td>
<td>1.7</td>
<td>2.2</td>
<td>2.2</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>IN</td>
<td>650</td>
<td>10.3</td>
<td>8.5</td>
<td>.8</td>
<td>1.9</td>
<td>1.8</td>
<td>2.2</td>
<td>2.9</td>
<td>2.7</td>
<td>3.0</td>
</tr>
<tr>
<td>MS</td>
<td>747</td>
<td>30.5</td>
<td>13.6</td>
<td>3.2</td>
<td>4.3</td>
<td>4.2</td>
<td>4.5</td>
<td>5.2</td>
<td>5.0</td>
<td>5.1</td>
</tr>
</tbody>
</table>
close to MEL\(^1\), i.e., averaging \( \text{Var}(Y_i|y, \sigma^2_u) \) over the posterior of \( \sigma^2_u \) gives about the same result as evaluating \( \text{Var}(Y_i|y, \sigma^2_u) \) at the posterior mean of \( \sigma^2_u \). Given this, \( \text{Var}[E(Y_i|y, \sigma^2_u)] = \text{Bayes}^2 - \text{Bayes}^1 \) can be thought of as accounting for uncertainty about \( \sigma^2_u \), and as a Bayesian analogue to the term augmenting (4) to account for error in estimating \( \sigma^2_u \). For most states \( \text{Var}[E(Y_i|y, \sigma^2_u)] \) is quite small, so \( \text{Var}(Y_i|y) \) is close to \( E[\text{Var}(Y_i|y, \sigma^2_u)] \), 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 \( \sigma^2_u \)) is \( x_i'\hat{\beta} = 17.7 \), 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 \( (E(Y_i|y, \sigma^2_u) \) as given by (3)) is sensitive to variation in \( \sigma^2_u \). Hence, the posterior variances reflect this. A similar, though less pronounced effect occurs for IN in 1993 when \( x_i'\hat{\beta} = 15.3 \) versus \( y_i = 10.3 \). Such occurrences are rare in this example, and when \( x_i'\hat{\beta} \) and \( y_i \) are close, \( \text{Var}[E(Y_i|y, \sigma^2_u)] \) 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 \( \sigma^2_u \), the Bayesian approach avoids unreasonable results from fixing \( \sigma^2_u \) at a single value near 0, and gives more intuitively plausible results.

References


