Survey Methodology, June 2006
Vol. 32, No. 1, pp. 97-103
Statistics Canada, Catalogue No. 12-001-XPB

Small Area Estimation Using Area Level Models and Estimated Sampling Variances

Yong You and Beatrice Chapman [1]

Abstract

In small area estimation, area level models such as the Fay–Herriot model (Fay and Herriot 1979) are widely used to obtain efficient model-based estimators for small areas. The sampling error variances are customarily assumed to be known in the model. In this paper we consider the situation where the sampling error variances are estimated individually by direct estimators. A full hierarchical Bayes (HB) model is constructed for the direct survey estimators and the sampling error variances estimators. The Gibbs sampling method is employed to obtain the small area HB estimators. The proposed HB approach automatically takes account of the extra uncertainty of estimating the sampling error variances, especially when the area-specific sample sizes are small. We compare the proposed HB model with the Fay–Herriot model through analysis of two survey data sets. Our results have shown that the proposed HB estimators perform quite well compared to the direct estimates. We also discussed the problem of priors on the variance components.

Key Words: Gibbs sampling; Hierarchical Bayes; Prior sensitivity; Sample size; Variance components.

1. Introduction

Sample surveys, for most purposes, are usually designed to provide reliable direct estimates for total populations and large areas by using area-specific sample data. These direct estimates frequently fail to provide reliable estimates for small areas due to very small sample sizes in the areas. Since small area estimates often have unsuitably large standard errors, to gain precision and reliability it is necessary to "borrow strength" from related areas thus increasing the effective sample size to construct indirect estimates for the small areas (Rao 1999). Explicit model-based methods that use supplementary data such as census and administrative data associated with the small areas in explicit models to link the small areas have been widely used in practice to obtain reliable model-based estimators. There are two broad classifications for these models: area level models and unit level models. Area level models are based on area direct survey estimators and unit level models are based on individual observations in the areas. For overviews and appraisals of models for small area estimation, see Rao (1999, 2003). In this paper we study area level models.

To obtain a basic area level model we assume that the small area parameter of interest θ_i is related to area-specific auxiliary data $x_i = (x_{i1}, \ldots, x_{ip})'$ through a linear model

$$\theta_i = x_i'\beta + v_i, i = 1, \ldots, m, \quad (1)$$

where m is the number of small areas, $\beta = (\beta_1, \ldots, \beta_p)'$ is the $p \times 1$ vector of regression coefficients, and the v_i's are area-specific random effects assumed to be independent and identically distributed (iid) with $\mathrm{E}(v_i) = 0$ and $\mathrm{var}(v_i) = \sigma_v^2$. The assumption of normality may also be included. This model is referred to as a linking model for θ_i.

The basic area level model also assumes that given the area-specific sample size $n_i > 1$, there exists a direct survey estimator y_i (usually design unbiased) for the small area parameter θ_i such that

$$y_i = \theta_i + e_i, \ i = 1, \ldots, m, \quad (2)$$

where the e_i is the sampling error associated with the direct estimator y_i. We also assume are that the e_i's are independent normal random variables with mean $E(e_i \mid \theta_i) = 0$ and sampling variance $\mathrm{var}(e_i \mid \theta_i) = \sigma_i^2$. Combining models (1) and (2) lead to a linear mixed area level model

$$y_i = x_i'\beta + v_i + e_i, \ i = 1, \ldots, m. \quad (3)$$

The well-known Fay–Herriot model (Fay and Herriot 1979) in small area estimation has the form of model (3) with the sampling variance σ_i^2 assumed to be known in the model. This is a very strong assumption. Usually a smoothed estimator of σ_i^2 is used in the model and then treated as known. In this paper, we consider the situation where the sampling variances σ_i^2 are unknown and are estimated by unbiased estimators s_i^2. Following Rivest and Vandal (2002) and Wang and Fuller (2003), we assume that the estimators s_i^2 are independent of the direct survey estimators y_i and s_i^2 has a sampling distribution $d_i s_i^2 \sim \sigma_i^2 \chi_{d_i}^2$, where $d_i = n_i - 1$ and n_i is the sample size for the i^{th} area. For example, suppose we have n_i observations from small area i and these observations are iid $N(\mu_i, \sigma^2)$. Let y_i be the sample mean of the n_i observations. Then $y_i \sim N(\mu_i, \sigma_i^2)$ and $\sigma_i^2 = \sigma^2 / n_i$. Then we can obtain a direct estimator of σ_i^2 as $s_i^2 = \tau_i^2 / n_i$, where τ_i^2 is the sample

1. Yong You and Beatrice Chapman, Household Survey Methods Division, Statistics Canada, Ottawa, K1A 0T6. E-mail: yongyou@statcan.ca.

variance of the n_i observations. Also y_i and s_i^2 are independent and $(n_i - 1)s_i^2 \sim \sigma_i^2 \chi^2_{n_i - 1}$.

We are interested in estimating the small area parameters θ_i. Rivest and Vandal (2002) and Wang and Fuller (2003) obtained the empirical best linear unbiased prediction (EBLUP) estimators of θ_i and the associated mean square error (MSE) approximations assuming that m and n_i are relatively large. In this paper, we consider a hierarchical Bayes (HB) approach using the Gibbs sampling method. An advantage of the HB approach is that it is straightforward, and the inferences for parameters θ_i are "exact" unlike the EBLUP approach. The small area parameter θ_i is estimated by its posterior mean and its precision is measured by its posterior variance. The HB approach automatically takes account of the uncertainties associated with unknown parameters in the model. Section 2 presents the HB area level models and related Gibbs sampling inferences. Section 3 presents two survey data analysis and sensitivity analysis. And finally in section 4, we offer some conclusions and future work directions.

2. Hierarchical Bayes Approach

We now present the area level model (3) and the estimated sampling variances s_i^2 in a HB framework as follows:

Model 1

- $y_i \mid \theta_i, \sigma_i^2 \sim \text{ind } N(\theta_i, \sigma_i^2),\ i = 1, \ldots, m;$
- $d_i s_i^2 \mid \sigma_i^2 \sim \text{ind } \sigma_i^2 \chi^2_{d_i},\ d_i = n_i - 1,\ i = 1, \ldots, m;$
- $\theta_i \mid \beta, \sigma_v^2 \sim \text{ind } N(x_i'\beta, \sigma_v^2),\ i = 1, \ldots, m;$
- Priors for the parameters: $\pi(\beta) \propto 1$, $\pi(\sigma_i^2) \sim \text{IG}(a_i, b_i), i = 1, \ldots, m$, $\pi(\sigma_v^2) \sim \text{IG}(a_0, b_0)$, where $a_i, b_i\ (0 \le i \le m)$ are chosen to be very small known constants to reflect vague knowledge on σ_i^2 and σ_v^2. IG denotes the inverse gamma distribution.

In Model 1, the sampling variances σ_i^2 are unknown. In practice however, we may have a simpler model by replacing σ_i^2 by its estimate s_i^2 (here s_i^2 is treated as a constant) and obtain the following model:

Model 2

- $y_i \mid \theta_i \sim \text{ind } N(\theta_i, \sigma_i^2 = s_i^2),\ i = 1, \ldots, m;$
- $\theta_i \mid \beta, \sigma_v^2 \sim \text{ind } N(x_i'\beta, \sigma_v^2),\ i = 1, \ldots, m;$
- Priors: $\pi(\beta) \propto 1$, $\pi(\sigma_v^2) \sim \text{IG}(a_0, b_0)$.

Model 2 is actually the Fay-Herriot model with sampling variances known as s_i^2. If area-specific sample sizes n_i are small, using s_i^2 in Model 2 may lead to underestimation of the MSE under the EBLUP approach or the posterior variance under the HB approach. We are interested in evaluating the effects of using s_i^2 for σ_i^2 in the model. We will obtain the HB estimates of θ_i under both Model 1 and Model 2 and compare the HB estimates through real survey data analysis.

Under the HB approach, we use the posterior mean $E(\theta_i \mid y)$ as a point estimate for θ_i and the posterior variance $V(\theta_i \mid y)$ as a measure of variability, where $y = (y_1, \ldots, y_m)'$. To estimate $E(\theta_i \mid y)$ and $V(\theta_i \mid y)$, we employ the Gibbs sampling method (Gelfand and Smith 1990). From Model 1, we obtain the following full conditional distributions for the Gibbs sampler:

- $[\theta_i \mid y, \beta, \sigma_i^2, \sigma_v^2] \sim N(\gamma_i y_i + (1 - \gamma_i) x_i'\beta, \gamma_i \sigma_i^2)$, where $\gamma_i = \dfrac{\sigma_v^2}{\sigma_v^2 + \sigma_i^2},\ i = 1, \ldots, m;$
- $$[\beta \mid y, \theta, \sigma_i^2, \sigma_v^2] \sim N_p\left(\begin{matrix}\left(\sum_{i=1}^m x_i x_i'\right)^{-1}\left(\sum_{i=1}^m x_i \theta_i\right), \\ \sigma_v^2\left(\sum_{i=1}^m x_i x_i'\right)^{-1}\end{matrix}\right);$$
- $$[\sigma_i^2 \mid y, \theta, \beta, \sigma_v^2] \sim \text{IG}\left(\begin{matrix} a_i + \dfrac{d_i + 1}{2}, b_i \\ + \dfrac{(y_i - \theta_i)^2 + d_i s_i^2}{2}\end{matrix}\right),$$
  where $d_i = n_i - 1,\ i = 1, \ldots, m;$
- $$[\sigma_v^2 \mid y, \theta, \beta, \sigma_i^2] \sim \text{IG}\left(\begin{matrix} a_0 + \dfrac{m}{2}, b_0 \\ + \dfrac{1}{2}\sum_{i=1}^m (\theta_i - x_i'\beta)^2 \end{matrix}\right).$$

It is straight forward to draw samples from these full conditional distributions. For implementations, we use $L = 5$ parallel runs each with a "burn-in" length of $B = 1{,}000$ and Gibbs sampling size of $G = 5{,}000$. The prior parameters a_i, b_i and a_0, b_0 are chosen to 0.0001. The HB estimator of θ_i under Model 1 is thus obtained as

$$\hat{\theta}_i^{\text{HB}} = (LG)^{-1} \sum_{l=1}^{L} \sum_{g=1}^{G} (\gamma_i^{(lg)} y_i + (1 - \gamma_i^{(lg)})\, x_i'\, \beta^{(lg)}), \qquad (4)$$

where $\gamma_i^{(lg)} = \sigma_v^{2(lg)} / (\sigma_v^{2(lg)} + \sigma_i^{2(lg)})$, and the posterior variance of θ_i can be estimated by

$$\begin{aligned}\hat{V}(\theta_i) &= (LG)^{-1} \sum_{l=1}^{L}\sum_{g=1}^{G} (\gamma_i^{(lg)} \sigma_i^{2(lg)}) \\ &\quad + (LG)^{-1} \sum_{l=1}^{L}\sum_{g=1}^{G} (\gamma_i^{(lg)} y_i + (1 - \gamma_i^{(lg)}) x_i' \beta^{(lg)})^2 \\ &\quad - \left\{(LG)^{-1} \sum_{l=1}^{L}\sum_{g=1}^{G} (\gamma_i^{(lg)} y_i + (1 - \gamma_i^{(lg)}) x_i' \beta^{(lg)})\right\}^2, \end{aligned} \qquad (5)$$

where $\{\beta^{(lg)}, \sigma_i^{2(lg)}, \sigma_v^{2(lg)}; g = 1, \ldots, G;\ l = 1, \ldots, L\}$ is the sample generated from the Gibbs sampler. The estimators (4) and (5) are the so-called Rao-Blackwellized HB estimators. The Rao–Blackwellized estimators are more stable in terms of simulation errors as shown, for example, in Gelfand and Smith (1991) and You and Rao (2000).

Now we consider Model 2. The full conditional distributions for the Gibbs sampler under Model 2 are

- $[\theta_i \mid y, \beta, \sigma_v^2] \sim N(\gamma_i y_i + (1-\gamma_i) x_i' \beta, \gamma_i s_i^2)$, where $\gamma_i = \dfrac{\sigma_v^2}{\sigma_v^2 + s_i^2}$, $i = 1, \ldots, m$;
- $$[\beta \mid y, \theta, \sigma_v^2] \sim N_p \left(\begin{array}{c} \left(\sum_{i=1}^m x_i x_i'\right)^{-1} \left(\sum_{i=1}^m x_i \theta_i\right), \\ \sigma_v^2 \left(\sum_{i=1}^m x_i x_i'\right)^{-1} \end{array} \right);$$
- $$[\sigma_v^2 \mid y, \theta, \beta] \sim \mathrm{IG} \left(\begin{array}{c} a_0 + \dfrac{m}{2}, b_0 \\ + \dfrac{1}{2} \sum_{i=1}^m (\theta_i - x_i' \beta)^2 \end{array} \right).$$

Under Model 2, the HB estimator of θ_i and the corresponding posterior variance estimator are given by (4) and (5) respectively with $\sigma_i^{2(lg)}$ replaced by s_i^2. Note that using s_i^2 instead of $\sigma_i^{2(lg)}$ may lead to severe underestimation of the posterior variance of θ_i for some areas with small sample sizes n_i. We will compare the HB estimators and evaluate the effects of using s_i^2 in Model 2 through data analysis in the following section.

3. Data Analysis

3.1 The Data Sets

We consider two interesting data sets in our analysis. The first data set is corn and soybean data with only 8 areas and small sample sizes in each area. The second data set is milk data with 43 areas and relatively large sample sizes in each area. We will compare the HB models and estimates based on these two data sets.

Corn and Soybean Data: The corn and soybean data comes from the U.S. Department of Agriculture and was first studied by Battese, Harter and Fuller (1988). The data contains reported crop hectares and LANDSAT satellite data for corn and soybeans in sample segments of 12 Iowa counties. The reported number of hectares for each crop comprise the direct survey estimates. Used as auxiliary data are the population means of number of pixels of a given crop per segment. The sample sizes are small for these areas, ranging from 1–5. For our purposes only the counties with a sample size of 3 and greater are used (8 areas meet the criteria). Therefore of the included counties the sample sizes range from 3–5. The original data is unit level data. In order to have area level data the sample mean and the sample standard error are calculated for each county. The sample standard errors for the corn and soybean data are quite large in general (yielding some CVs in the 0.3–0.4 range and one CV of 0.532) but by chance there are also some small values in some instances (for corn data, Franklin has standard error 5.704 and CV 0.036). Because the sample sizes are so small, these sample standard errors cannot be trusted to approximate the true standard errors. Table 1 presents the modified area level data for corn and soybeans from the unit level data of Battese *et al.* (1988).

Table 1
Modified Crop Area Level Data, from Battese, Harter and Fuller (1988)

		Corn			Soybeans		
County	n_i	y_i	SD	CV	y_i	SD	CV
Franklin	3	158.623	5.704	0.036	52.473	16.425	0.313
Pocahontas	3	102.523	43.406	0.423	118.697	50.290	0.424
Winnebago	3	112.773	30.547	0.271	88.573	10.453	0.118
Wright	3	144.297	53.999	0.374	97.800	52.034	0.532
Webster	4	117.595	21.298	0.181	112.980	23.531	0.208
Hancock	5	109.382	15.661	0.143	117.478	17.209	0.146
Kossuth	5	110.252	12.112	0.110	117.844	20.954	0.178
Hardin	5	120.054	36.807	0.307	101.834	26.790	0.263

Milk Data: The milk data, used in an article by Arora and Lahiri (1997), comes from the U.S. Bureau of Labor Statistics. The estimated values are the average expenditure on fresh milk for the year 1989. There is data for 43 areas with sample sizes ranging from 95 to 633. The CVs range from 0.074 to 0.341 over the 43 areas. A more detailed description of the data can be found in Arora and Lahiri (1997). For completeness, we give the data in Table 2. Following Arora and Lahiri (1997), we use $x_i' \beta = \beta_j$ if $i \in j^{\text{th}}$ major area, a collection of similar publication areas. Arora and Lahiri (1997) used eight major areas. Since this division of the eight major areas is not given in their paper, after noting trends in the data we used the Fay-Herriot model to test two new divisions of 6 and 4 major areas that combine similar survey estimates. These major areas produced large CV reduction in general. Where the 6 groups had yielded an average CV reduction of about 20% the 4 groups gave approximately an average 25% CV reduction over the direct estimates. Comparison of the point estimates and CVs have shown that the 4 major areas perform better than the 6 major areas. The 4 major areas are 1–7, 8–14, 15–25 and 26–43. In this paper, we will use these 4 groups as auxiliary variables for illustration purpose only.

Table 2
Milk Data, from Arora and Lahiri (1997)

Small Area	n_i	y_i	SD	CV
1	191	1.099	0.163	0.148
2	633	1.075	0.080	0.074
3	597	1.105	0.083	0.075
4	221	0.628	0.109	0.174
5	195	0.753	0.119	0.158
6	191	0.981	0.141	0.144
7	183	1.257	0.202	0.161
8	188	1.095	0.127	0.116
9	204	1.405	0.168	0.120
10	188	1.356	0.178	0.131
11	149	0.615	0.100	0.163
12	290	1.460	0.201	0.138
13	250	1.338	0.148	0.111
14	194	0.854	0.143	0.167
15	184	1.176	0.149	0.127
16	193	1.111	0.145	0.131
17	218	1.257	0.135	0.107
18	266	1.430	0.172	0.120
19	214	1.278	0.137	0.107
20	213	1.292	0.163	0.126
21	196	1.002	0.125	0.125
22	95	1.183	0.247	0.209
23	195	1.044	0.140	0.134
24	187	1.267	0.171	0.135
25	479	1.193	0.106	0.089
26	230	0.791	0.121	0.153
27	186	0.795	0.121	0.152
28	199	0.759	0.259	0.341
29	238	0.796	0.106	0.133
30	207	0.565	0.089	0.158
31	165	0.886	0.225	0.254
32	153	0.952	0.205	0.215
33	210	0.807	0.119	0.147
34	383	0.582	0.067	0.115
35	255	0.684	0.106	0.155
36	226	0.787	0.126	0.160
37	224	0.440	0.092	0.209
38	212	0.759	0.132	0.174
39	211	0.770	0.100	0.130
40	179	0.800	0.113	0.141
41	312	0.756	0.083	0.110
42	241	0.865	0.121	0.140
43	205	0.640	0.129	0.202

3.2 Analysis of Results

Corn and Soybean Data: First we consider the effect of our treatment of σ_i^2 using the HB approach. Table 3 presents the HB estimates $\hat{\theta}_i^{\text{HB}}$ and the associated standard errors (SDs) and CVs for the small area corn and soybean data sets. The SD is the square root of the posterior variance. Under Model 1 (σ_i^2 unknown), the SDs and CVs are consistently larger than the corresponding SDs and CVs under Model 2 ($\sigma_i^2 = s_i^2$ known). The increased SDs and CVs of Model 1 are expected since this model takes into account the added variability of estimating σ_i^2. On average there is about 20% increase in SDs and CVs (this calculation excludes Franklin for corn data). The results support the fact that letting $\sigma_i^2 = s_i^2$, the known direct estimate of σ_i^2, leads to underestimation of the SD and CV of $\hat{\theta}_i$. Inspection of small areas Franklin and Webster for the corn data and county Winnebago for the soybean data establish in some cases where the sampling errors by chance are quite small this under estimation is severe.

Comparison of the HB estimates under Model 1 and Model 2 to the direct estimates can be made using the CVs in Table 1 and Table 3. Under Model 2 the HB estimates have smaller CVs than the direct estimates in 6 of the 8 counties for the corn data and similarly for the soybean data, 6 out of 8 counties. Of the remaining 2 counties for each crop, the CVs under Model 2 are the same as the direct survey CVs or only slightly larger. Estimators from Model 2 therefore seem to have gained efficiency compared to the direct survey estimators. Now examining the HB estimates under Model 1 and the direct survey estimates lead to mixed results for the corn and soybean data sets. Model 1 accounts for the added uncertainty of estimating the sampling variances and so in only 4 of the 8 counties the HB estimates show improvements in efficiency for the corn data. For the soybean data 5 out of 8 counties demonstrate the HB estimates as improvements on the direct survey CVs. For the remaining counties the direct estimates exhibit lower CVs and even substantially lower CVs in some cases. For the corn data, counties Franklin and Webster have CV increases with Model 1 of more than 0.09 and 0.12 respectively. As well for the soybean data, county Winnebago has a CV increase of almost 0.10 from the direct survey estimate, using Model 1. Areas where the direct estimates demonstrate smaller CVs compared to the HB estimates include a number of those areas where the CVs are by chance atypically small. So the increased model-based CVs may reflect more appropriate CVs for those areas. Of the 7 cases where the direct CVs are smaller compared to the HB CVs under Model 1, the 3 cases noted above have severe differences and the remaining 4 instances show only slight reduction in efficiency with use of Model 1. Since direct survey estimates quite often have unacceptably large CVs and yet still by chance may have CVs grossly and inexplicably small, HB estimation under Model 1 may be more reliable and reasonable by taking into consideration the uncertainty of estimating σ_i^2.

Milk data: Table 4 contains the HB estimates for the milk data. As expected, over the 43 areas the treatment of σ_i^2 as known or unknown shows negligible differences in terms of point estimates, SDs and CVs due to the large sample sizes in the 43 areas. Therefore the substitution of $\sigma_i^2 = s_i^2$ in the model is reasonable when the area-specific sample sizes are large, as clearly shown in this example. Also the HB estimates give reduced SDs and CVs when compared to the direct survey estimates in Table 2. As would be expected, the HB estimation approach is thus an improvement on the direct survey estimates.

Table 3
Comparison of HB Estimates for Crop Data

County	σ_i^2 known ($\sigma_i^2 = s_i^2$) $\hat{\theta}_i^{HB}$	SD	CV	σ_i^2 unknown $\hat{\theta}_i^{HB}$	SD	CV
			Corn			
Franklin	155.788	6.061	0.039	142.862	18.408	0.129
Pocahontas	100.813	28.297	0.281	91.560	32.420	0.356
Winnebago	115.337	28.406	0.246	113.130	35.207	0.311
Wright	131.630	28.345	0.215	123.547	30.764	0.250
Webster	109.030	20.634	0.189	97.856	29.834	0.307
Hancock	121.682	15.656	0.129	123.478	17.857	0.145
Kossuth	115.710	11.180	0.097	114.910	12.510	0.109
Hardin	135.626	23.228	0.171	135.178	23.804	0.176
			Soybean			
Franklin	75.375	16.272	0.216	88.186	21.067	0.239
Pocahontas	116.943	27.031	0.231	109.052	30.098	0.276
Winnebago	87.525	10.304	0.118	88.053	18.854	0.214
Wright	104.184	23.671	0.227	105.825	24.497	0.232
Webster	115.510	20.789	0.180	109.455	25.801	0.236
Hancock	101.368	15.741	0.155	102.876	17.311	0.169
Kossuth	102.388	14.948	0.146	101.862	15.019	0.148
Hardin	87.455	17.774	0.203	93.397	20.251	0.217

Table 4
Comparison of HB Estimates for Milk Data

Small area	σ_i^2 known ($\sigma_i^2 = s_i^2$) $\hat{\theta}_i^{HB}$	SD	CV	σ_i^2 unknown $\hat{\theta}_i^{HB}$	SD	CV
1	1.020	0.113	0.111	1.021	0.111	0.109
2	1.045	0.072	0.069	1.045	0.071	0.068
3	1.065	0.073	0.069	1.065	0.074	0.069
4	0.767	0.095	0.124	0.770	0.096	0.125
5	0.849	0.096	0.113	0.852	0.096	0.113
6	0.975	0.103	0.106	0.975	0.102	0.105
7	1.058	0.125	0.118	1.055	0.125	0.118
8	1.097	0.099	0.090	1.096	0.099	0.090
9	1.219	0.121	0.099	1.215	0.121	0.100
10	1.192	0.122	0.102	1.190	0.122	0.102
11	0.793	0.094	0.119	0.799	0.097	0.122
12	1.213	0.131	0.108	1.209	0.130	0.107
13	1.206	0.112	0.093	1.203	0.112	0.093
14	0.984	0.107	0.109	0.987	0.107	0.109
15	1.187	0.105	0.088	1.187	0.104	0.087
16	1.156	0.104	0.090	1.156	0.102	0.089
17	1.225	0.101	0.083	1.225	0.100	0.081
18	1.284	0.115	0.089	1.281	0.113	0.088
19	1.234	0.101	0.082	1.235	0.100	0.081
20	1.233	0.110	0.089	1.233	0.110	0.089
21	1.092	0.097	0.089	1.095	0.098	0.089
22	1.192	0.128	0.107	1.193	0.127	0.106
23	1.122	0.103	0.092	1.125	0.103	0.091
24	1.221	0.113	0.092	1.220	0.111	0.091
25	1.193	0.086	0.072	1.193	0.086	0.072
26	0.761	0.091	0.120	0.762	0.091	0.120
27	0.763	0.092	0.120	0.762	0.091	0.119
28	0.734	0.125	0.170	0.732	0.123	0.169
29	0.768	0.085	0.110	0.767	0.085	0.110
30	0.615	0.076	0.124	0.618	0.076	0.123
31	0.769	0.122	0.158	0.767	0.120	0.156
32	0.795	0.119	0.150	0.792	0.118	0.148
33	0.771	0.091	0.118	0.770	0.090	0.117
34	0.612	0.060	0.099	0.613	0.062	0.100
35	0.701	0.085	0.121	0.701	0.084	0.120
36	0.757	0.094	0.123	0.759	0.093	0.123
37	0.534	0.080	0.150	0.538	0.081	0.151
38	0.744	0.096	0.129	0.743	0.095	0.128
39	0.754	0.082	0.108	0.753	0.082	0.108
40	0.768	0.088	0.115	0.768	0.088	0.115
41	0.747	0.071	0.095	0.747	0.070	0.094
42	0.801	0.093	0.116	0.800	0.092	0.116
43	0.682	0.094	0.139	0.682	0.094	0.138

3.3 Priors and Sensitivity Analysis

In Model 1, the sampling variances σ_i^2 are assumed to be independent with inverse gamma prior distribution $\text{IG}(a_i, b_i)$, and the model variance σ_v^2 also has inverse gamma prior distribution $\text{IG}(a_0, b_0)$, where a_i, b_i $(0 \le i \le m)$ are chosen to be very small known constants to reflect vague knowledge on σ_i^2 and σ_v^2. So we have used proper priors to avoid the problem of any improper posteriors. One may consider using flat priors for σ_i^2 and σ_v^2, *i.e.*, $\pi(\sigma_i^2) \propto 1$, and $\pi(\sigma_v^2) \propto 1$, similar to the flat prior on β. With the flat priors on σ_i^2 and σ_v^2, the full conditional distributions for σ_i^2 and σ_v^2 are given as

$$[\sigma_i^2 \mid y, \theta, \beta, \sigma_v^2] \sim \text{IG}\left(\frac{d_i - 1}{2}, \frac{(y_i - \theta_i)^2 + d_i s_i^2}{2}\right),$$

and

$$[\sigma_v^2 \mid y, \theta, \beta, \sigma_i^2] \sim \text{IG}\left(\frac{m-2}{2}, \frac{1}{2}\sum_{i=1}^{m}(\theta_i - x_i'\beta)^2\right).$$

The implementation of the Gibbs sampler under the flat priors is also straightforward. However, the flat priors on σ_i^2 and σ_v^2 may lead to improper posteriors if the sample sizes and the number of small areas are small. In order to see the problem on σ_i^2 more clearly, we can study the Model 1 in two steps. First, we can obtain the posterior of σ_i^2 given its direct estimate s_i^2 as

$$\begin{aligned}\pi(\sigma_i^2 \mid s_i^2) &\propto f(s_i^2 \mid \sigma_i^2) \cdot \pi(\sigma_i^2)\\ &\propto (\sigma_i^2)^{-d_i/2} \cdot \exp\{-\sigma_i^{-2} d_i s_i^2 / 2\} \cdot \pi(\sigma_i^2).\end{aligned}$$

By assuming a flat prior $\pi(\sigma_i^2) \propto 1$, we can obtain

$$\pi(\sigma_i^2 \mid s_i^2) \sim \text{IG}\left(\frac{d_i}{2} - 1, \frac{d_i s_i^2}{2}\right),$$

provided that $d_i > 2$, or $n_i > 3$. Then we can use this proper IG posterior $\pi(\sigma_i^2 \mid s_i^2)$ as an informative prior for σ_i^2 in the sampling model $y_i \mid \theta_i, \sigma_i^2 \sim \text{ind } N(\theta_i, \sigma_i^2)$. This will ensure to have proper posterior inference. For the modified corn and soybean data, using flat priors on σ_i^2 will lead to improper posterior due to the small sample sizes $(n_i = 3)$ for some areas. Thus, proper inverse gamma priors are used in the data analysis to ensure that all the posteriors are proper, as commonly used in the HB small area estimation in practice (*e.g.*, Arora and Lahiri 1997; Datta, Lahiri, Maiti and Lu 1999; You and Rao 2000; Rao 2003). Hence we do not face the problem of some posteriors being improper, since correct HB inference should be based on proper posteriors. Under Model 2 with the sampling variance known as $\sigma_i^2 = s_i^2$, using a flat prior $\pi(\sigma_v^2) \propto 1$ for σ_v^2, the posterior of σ_v^2 will be proper provided that

$m > p+2$, where m is the number of small areas and p is the size of regression parameters β (Rao 2003, page 238). Since the number of small areas is usually relatively large, this condition is generally satisfied in practice.

For the sensitivity analysis of vague proper priors, we can test the sensitivity of the posterior estimates to the choice of prior parameters $a_i, b_i (0 \le i \le m)$. Under Model 1, we set $a_i = b_i$ at four different values, *i.e.*, 0.0001, 0.001, 0.01 and 0.1. Table 5 presents the estimated posterior means for the corn and soybean data, and Table 6 presents the corresponding CVs.

Table 5
Comparison of Posterior Mean Estimates for Crop Data

	IG (a_i, b_i), $a_i = b_i$			
County	0.0001	0.001	0.01	0.1
		Corn		
Franklin	142.862	142.593	143.155	144.311
Pocahontas	91.560	91.912	91.422	91.974
Winnebago	113.130	113.068	121.578	114.430
Wright	123.547	124.170	125.103	125.351
Webster	97.856	98.231	99.132	98.511
Hancock	123.478	123.858	124.395	124.138
Kossuth	114.910	115.281	115.316	115.528
Hardin	135.178	134.157	135.223	136.001
		Soybean		
Franklin	88.186	89.368	89.145	89.513
Pocahontas	109.052	109.571	107.745	108.176
Winnebago	88.053	87.478	86.267	87.302
Wright	105.825	106.712	105.142	104.676
Webster	109.455	108.392	109.835	110.252
Hancock	102.876	103.413	102.240	101.808
Kossuth	101.862	101.159	101.379	100.808
Hardin	93.397	94.713	93.576	94.767

Table 6
Comparison of Posterior CVs for Crop Data

	IG (a_i, b_i), $a_i = b_i$			
County	0.0001	0.001	0.01	0.1
		Corn		
Franklin	0.129	0.124	0.128	0.125
Pocahontas	0.356	0.351	0.347	0.341
Winnebago	0.311	0.314	0.321	0.324
Wright	0.250	0.246	0.235	0.236
Webster	0.307	0.292	0.285	0.280
Hancock	0.145	0.148	0.148	0.142
Kossuth	0.109	0.110	0.107	0.104
Hardin	0.176	0.173	0.178	0.168
		Soybean		
Franklin	0.239	0.233	0.231	0.227
Pocahontas	0.276	0.281	0.271	0.296
Winnebago	0.214	0.193	0.196	0.198
Wright	0.232	0.223	0.231	0.226
Webster	0.236	0.231	0.237	0.228
Hancock	0.169	0.165	0.168	0.161
Kossuth	0.148	0.145	0.142	0.135
Hardin	0.217	0.215	0.213	0.213

It is clear from Table 5 and Table 6 that the posterior estimates and the corresponding CVs are about the same and stable, which indicates that the HB estimates are not sensitive to the choice of vague proper priors. For the milk data, the HB estimates are very stable to these proper vague priors (results are not provided here). Since the milk data has large sample sizes, flat priors on variance components can also be used to analyze the milk data under Model 1. We thus obtained the HB estimates based on the flat priors and compared them with the HB estimates based on the vague IG priors. These HB estimates are almost identical and stable with relative difference ranging from 0.07% to 2.23%, an average value of 0.69% over 43 areas, which indicates that the posterior estimates of small area means based on Model 1 are very stable and not sensitive to the choice of flat priors or vague IG priors, provided that the sample sizes and number of small areas are relatively large.

4. Conclusion and Future Work

In this paper we have studied the well-known Fay-Herriot model with the situations where σ_i^2, the sampling error variances, are assumed unknown and where they are estimated by unbiased estimators s_i^2, using the HB approach. The full HB approach with the Gibbs sampling method automatically takes into account the extra uncertainty associated with the estimation of σ_i^2. We applied the HB approach in two survey data analysis. Our results have shown that the proposed HB approach under Model 1 works quite well no matter the area-specific sample sizes are small or large. For future work, the proposed HB modeling approach can be extended to the general area level models studied by You and Rao (2002). Application of the new HB modeling approach includes the census undercoverage estimation as in You, Rao and Dick (2004). Under Model 1, the HB estimators of the sampling variances σ_i^2 can be obtained. These HB estimators of σ_i^2 can be used as alternative smoothed estimators for σ_i^2 in the sampling models. Application and evaluation of the HB estimators of the sampling variances include the census undercoverage estimation and the Canadian Labour Force Survey (LFS) unemployment rate estimation (You, Rao and Gambino 2003). We also plan to compare the HB approach with the EBLUP approach as studied by Rivest and Vandal (2002) and Wang and Fuller (2003).

Acknowledgements

The authors would like to thank two referees, an Associate Editor, the Deputy Editor and the Editor, Dr. M.P. Singh, for their constructive comments and suggestions. The authors also would like to thank J.N.K. Rao of Carleton University for his useful suggestion and Jack Gambino and Eric Rancourt of Statistics Canada for their comments on the early version of the paper. This work was supported by Statistics Canada Methodology Branch Research Block Fund.

References

Arora, V., and Lahiri, P. (1997). On the superiority of the Bayesian method over the BLUP in small area estimation problems. *Statistica Sinica*, 7, 1053-1063.

Battese, G.E., Harter, R.M. and Fuller, W.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.

Datta, G.S., Lahiri, P., Maiti, T. and Lu, K.L. (1999) Hierarchical Bayes estimation of unemployment rates for the states of the U.S. *Journal of the American Statistical Association*, 94, 1074-1082.

Fay, R.E., and Herriot, R.A. (1979). Estimation of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 268-277.

Gelfand, A.E., and Smith, A.F.M. (1990). Sample-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 972-985.

Gelfand, A.E., and Smith, A.F.M. (1991). Gibbs sampling for marginal posterior expectations. *Communications In Statisitcs – Theory and Methods*, 20, 1747-1766.

Rao, J.N.K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, 25, 175-186.

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

Rivest, L.P., and Vandal, N. (2002). Mean squared error estimation for small areas when the small area variances are estimated. *Proceedings of the International Conference on Recent Advances in Survey Sampling*, July 10-13, 2002, Ottawa, Canada.

Wang, J., and Fuller, W.A. (2003). The mean square error of small area predictors constructed with estimated area variances. *Journal of the American Statistical Association*, 98, 716-723.

You, Y., and Rao, J.N.K. (2000). Hierarchical Bayes estimation of small area means using multi-level models. *Survey Methodology*, 26, 173-181.

You, Y., and Rao, J.N.K. (2002). Small area estimation using unmatched sampling and linking models. *The Canadian Journal of Statistics*, 30, 1, 3-15.

You, Y., Rao, J.N.K. and Dick, P. (2004). Benchmarking hierarchical Bayes small area estimators in the Canadian census undercoverage estimation. *Statistics in Transition*, 6, 631-640.

You, Y., Rao, J.N.K. and Gambino, J. (2003). Model-based unemployment rate estimation for the Canadian Labour Force Survey: A hierarchical Bayes approach. *Survey Methodology*, 29, 25-32.