SSC Annual Meeting, May 2004
Proceedings of the Survey Methods Section

SMALL AREA ESTIMATION USING 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 variances are customarily assumed to be known in these models. In this paper we consider the situation where the sampling variances are estimated individually by direct estimators. A full hierarchical Bayes (HB) model is constructed for the direct survey estimators and the sampling variances estimators. We also consider the empirical best linear unbiased prediction (EBLUP) approach and obtain the EBLUP estimators. We compare the HB and the EBLUP methods through analysis of two survey data sets. Our results have shown that the proposed HB estimators perform better than the EBLUP estimators in terms of taking account of the extra uncertainty of estimating the sampling error variances, especially when the area-specific sample sizes are very small. When the sample sizes are large, both the HB and the EBLUP estimators perform equally better than the direct survey estimators in terms of coefficient of variation (CV) reduction. The proposed HB modeling approach will be applied to Canadian census undercoverage estimation and Canadian Labour Force Survey (LFS) unemployment and employment estimation.

KEY WORDS: EBLUP, Hierarchical Bayes, Sampling variances, Small areas.

RÉSUMÉ

Dans le contexte de l'estimation sur petits domaines, des modèles au niveau du domaine tels que le modèle de Fay-Herriot (Fay et Herriot, 1979) sont souvent utilisés pour obtenir des estimateurs efficaces basés sur un modèle pour les petits domaines. Les variances d'échantillonnage sont habituellement considérées comme étant connues dans ces modèles. Dans cet article, nous considérons la situation où les variances d'échantillonnage sont estimées séparément par des estimateurs directs. Un modèle hiérarchique bayésien (HB) complet est construit pour les estimateurs de sondage directs et les estimateurs de variance d'échantillonnage. Nous considérons également l'approche de la méthode empirique du meilleur prédicteur linéaire sans biais (EBLUP) et obtenons les estimateurs EBLUP. Nous comparons les méthodes HB et EBLUP par l'analyse de jeux de données d'enquête. Nos résultats ont montré que les estimateurs HB proposés performent mieux que les estimateurs EBLUP pour ce qui est de tenir compte de l'incertitude supplémentaire dans l'estimation de la variance des erreurs d'échantillonnage, particulièrement quand la taille de l'échantillon est petite dans le domaine. Quand les tailles d'échantillon sont grandes, les estimateurs HB et EBLUP performent aussi bien que l'estimateur direct pour ce qui est de la réduction des coefficients de variation (CV). L'approche de modélisation HB proposée sera appliquée à l'estimation de la sous-couverture du Recensement du Canada ainsi qu'aux estimations de chômage et d'emploi de l'Enquête sur la population active (EPA) du Canada.

MOTS CLÉS : EBLUP; hiérarchique bayésien; petits domaines; variance d'échantillonnage.

.

1. INTRODUCTION

Sample surveys 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. Explicit model-based methods are usually used to "borrow strength" from related areas to obtain reliable model-based estimators. Area level models are considered in this paper. 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 $x_i = (x_{i1},...,x_{ip})'$ through a linear model

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

[1] Yong You, Household Survey Methods Division, Statistics Canada, Ottawa, K1A 0T6, Canada, yongyou@statcan.ca; Beatrice Chapman, Social Survey Methods Division, Statistics Canada, K1A 0T6, Canada, chapbea@statcan.ca.

where $\beta = (\beta_1,...,\beta_p)'$ is the p × 1 vector of regression coefficients and the v_i's are area-specific iid normal random effects with E(v_i) = 0 and var(v_i) = σ_v^2. This model is referred to as a linking model for θ_i. The basic area level model also assumes that given n_i>1, there exists a direct survey estimator y_i for the small area parameter θ_i such that

$$y_i = \theta_i + e_i,\ i = 1,...,m, \tag{2}$$

where the e_i are the sampling errors associated with the direct estimator y_i. The e_i's are independent normal random variables with mean $E(e_i|\theta_i) = 0$ and sampling variance $\text{var}(e_i|\theta_i) = \sigma_i^2$. Combining models (1) and (2) we have

$$y_i = x_i'\beta + v_i + e_i,\ i = 1,\ldots,m. \tag{3}$$

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. The estimators s_i^2 are independent of the direct survey estimators y_i. Following Wang (2000), Rivest and Vandal (2002) and Wang and Fuller (2003), we also assume that $d_i s_i^2 \sim \sigma_i^2 \chi_{d_i}^2$, 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 an estimator of σ_i^2 as $s_i^2 = s^2 / n_i$, where s^2 is the sample variance of the n_i observations. Also y_i and s_i^2 are independent and $(n_i - 1)s_i^2 \sim \sigma_i^2 \chi_{n_i - 1}^2$.

We are interested in estimating the small area parameters θ_i. Wang (2000), Wang and Fuller (2003), and Rivest and Vandal (2002) obtained the empirical best linear unbiased prediction (EBLUP) estimators of θ_i and the associated mean square error (MSE) approximations. In this paper, we consider both the hierarchical Bayes (HB) approach using the Gibbs sampling method and the EBLUP approach. 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 HB approach will automatically take account of the uncertainties associated with unknown parameters, but it requires the specification of prior distributions on model parameters β, σ_v^2 and σ_i^2. For this we now present the area level model (3) and the estimated sampling variances s_i^2 in a HB framework as follows:

<u>Model 1</u>

- $y_i|\theta_i, \sigma_i^2 \sim ind\ N(\theta_i, \sigma_i^2), i = 1,\ldots,m;$
- $d_i s_i^2|\sigma_i^2 \sim ind\ \sigma_i^2 \chi_{d_i}^2,\ d_i = n_i - 1, i = 1,\ldots,m;$
- $\theta_i|\beta, \sigma_v^2 \sim ind\ N(x_i'\beta, \sigma_v^2), i = 1,\ldots,m;$
- Priors for the parameters: $\pi(\beta) \propto 1$, $\pi(\sigma_i^2) \sim IG(a_i, b_i)$, $i = 1,\ldots,m$, $\pi(\sigma_v^2) \sim IG(a_0, b_0)$, 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 and obtain the following model:

<u>Model 2</u>

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

Model 2 is actually the Fay-Herriot model with sampling variances known as s_i^2. If area-specific sample sizes 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. We will also compare the HB estimates with the EBLUP estimates. Section 2 presents small area estimation using the HB and EBLUP approaches. In section 3, we compare the HB and EBLUP methods through analyzing two real survey data sets. And in section 4 we offer some conclusions and future work directions.

2. SMALL AREA ESTIMATION

2.1 Hierarchical Bayes approach

Under the hierarchical Bayes approach, we use the posterior mean $E(\theta_i|y)$ as a point estimate of θ_i and the posterior variance $V(\theta_i|y)$ as a measure of variability, where $y=(y_1,..,y_m)'$. To estimate $E(\theta_i|y)$ and $V(\theta_i|y)$, we employ the Gibbs sampling method. The full conditional distributions for the Gibbs sampler are given in You and Chapman (2003). For implementations, we use L = 5 parallel runs each with a "burn-in" length of B = 1000 and Gibbs sampling size of G = 5000. The HB estimator of θ_i under Model 1 is thus obtained as

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

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

$$\hat{V}(\theta_i) = (LG)^{-1}\sum_{l=1}^{L}\sum_{g=1}^{G}(\gamma_i^{(\lg)}\sigma_i^{2(\lg)}) + (LG)^{-1}\sum_{l=1}^{L}\sum_{g=1}^{G}(\gamma_i^{(\lg)} y_i + (1-\gamma_i^{(\lg)}) x_i'\beta^{(\lg)})^2$$
$$- \{(LG)^{-1}\sum_{l=1}^{L}\sum_{g=1}^{G}(\gamma_i^{(\lg)} y_i + (1-\gamma_i^{(\lg)}) x_i'\beta^{(\lg)})\}^2, \tag{5}$$

where $\left\{\beta^{(\lg)}, \sigma_i^{2(\lg)}, \sigma_v^{2(\lg)}; g=1,...,G; l=1,...,L\right\}$ is the sample generated from the Gibbs sampler. 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.

2.2 EBLUP approach

Under the basic area level models (1) and (2), by assuming σ_i^2 and σ_v^2 known in the model, we can obtain the best linear unbiased prediction (BLUP) estimator of θ_i as

$$\tilde{\theta}_i = \gamma_i y_i + (1-\gamma_i) x_i' \tilde{\beta}, \tag{6}$$

where $\gamma_i = \sigma_v^2 / (\sigma_v^2 + \sigma_i^2)$ and

$$\tilde{\beta} = \left[\sum_{i=1}^{m}(\sigma_i^2 + \sigma_v^2)^{-1} x_i x_i'\right]^{-1}\left[\sum_{i=1}^{m}(\sigma_i^2 + \sigma_v^2)^{-1} x_i y_i\right]. \tag{7}$$

To estimate the variance component σ_v^2, we have to assume σ_i^2 to be known in the model. Replacing σ_i^2 by its estimate s_i^2, we use the Fay-Herriot iterative (FHI) method (Fay and Herriot, 1979) to estimate σ_v^2. The FHI method is suggested by Datta, Rao and Smith (2002). The FHI method is as follows: starting with $\sigma_v^{2(0)} = 0$, solve iteratively,

$$\sigma_v^{2(a+1)} = \sigma_v^{2(a)} + \left[h_*'(\sigma_v^{2(a)})\right]^{-1}\left[m - p - h(\sigma_v^{2(a)})\right]$$

constraining $\sigma_v^{2(a+1)} \geq 0$, where $h(\sigma_v^2) = \sum_{i=1}^{m} (y_i - x_i'\tilde{\beta})^2 / (s_i^2 + \sigma_v^2)$ and $h'_*(\sigma_v^2) = -\sum_{i=1}^{m} (y_i - x_i'\tilde{\beta})^2 / (s_i^2 + \sigma_v^2)^2$. $\tilde{\beta}$ is given by (7). Convergence of the iteration is rapid. The asymptotic variance of the FHI estimator σ_v^2 was obtained by Datta, Rao and Smith (2002) as $V(\sigma_v^2) = 2m(\sum_{i=1}^{m} (\sigma_v^2 + s_i^2)^{-1})^{-2}$. Replacing σ_v^2 and σ_i^2 by $\hat{\sigma}_v^2$ and s_i^2 in (6), we obtain the EBLUP estimator of θ_i as $\hat{\theta}_i = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i) x_i'\hat{\beta}$, where $\hat{\gamma}_i = \hat{\sigma}_v^2 / (\hat{\sigma}_v^2 + s_i^2)$. An estimator of the MSE of $\hat{\theta}_i$ is given as $mse(\hat{\theta}_i) = g_{0i} + g_{1i} + g_{2i} + 2g_{3i} + g_{4i}$, where g_{1i}, g_{2i} and g_{3i} are the terms obtained by Prasad and Rao (1990) in the MSE estimation. g_{1i} is the leading term, g_{2i} is due to estimation of β, and g_{3i} is due to estimation of σ_v^2. g_{0i} is an extra term due to estimation of σ_v^2 using the FHI method (Datta, Rao and Smith, 2002). g_{4i} is a new term due to unknown σ_i^2 in the sampling model (1). Wang (2000), Wang and Fuller (2003) and Rivest and Vandal (2002) obtained the g_{4i} term to account for the extra uncertainty associated with the estimation of σ_i^2 by s_i^2. These five terms are $g_{0i} = -(1 - \hat{\gamma}_i)^2 b(\hat{\sigma}_v^2)$, where

$$b(\hat{\sigma}_v^2) = 2\{m \sum_{i=1}^{m} (s_i^2 + \hat{\sigma}_v^2)^{-2} - (\sum_{i=1}^{m} (s_i^2 + \hat{\sigma}_v^2)^{-1})^2\} / \{\sum_{i=1}^{m} (s_i^2 + \hat{\sigma}_v^2)^{-1}\}^3,$$

$$g_{1i} = \hat{\gamma}_i s_i^2, g_{2i} = \hat{\sigma}_v^2 (1 - \hat{\gamma}_i)^2 x_i' (\sum_{i=1}^{m} \hat{\gamma}_i x_i x_i')^{-1} x_i, g_{3i} = s_i^4 (\hat{\sigma}_v^2 + s_i^2)^{-3} V(\hat{\sigma}_v^2), \text{ and } g_{4i} = 4(n_i - 1)^{-1} \hat{\sigma}_v^4 s_i^4 (\hat{\sigma}_v^2 + s_i^2)^{-3}.$$

In the following data analysis, we will evaluate the EBLUP and the MSE estimators and compare them with the HB estimators empirically.

3. DATA ANALYSIS

3.1 Data Description

In data analysis, we consider two data sets. One is corn and soybean data and another one is milk 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 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. Because the sample sizes are so small, these sample standard errors cannot be trusted to approximate the true standard errors. The data is given in You and Chapman (2003). 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). Following Arora and Lahiri (1997), we use $x_i'\beta = \beta_j$ if $i \in j$-th major area, a collection of similar publication areas. Arora and Lahiri (1997) used eight major areas. In this paper, we will use 4 groups as auxiliary variables for illustration purpose only. The 4 major areas are 1-7, 8-14, 15-25 and 26-43.

3.2 Comparison of Estimates

Corn and soybean data

First we consider the effect of σ_i^2 using the HB approach. We considered the HB estimates $\hat{\theta}_i^{HB}$ and the associated standard errors (SDs) and CVs. 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. 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.

When $\sigma_i^2 = s_i^2$ for the corn and soybean data, both the HB and the EBLUP approaches yield comparable results, with only a slight apparent advantage in terms of efficiency with the use of the EBLUP estimation. Of the 16 cases from the 2 data sets combined, EBLUP is more efficient 8 times, HB and EBLUP give similar results in 5 cases and HB outperforms EBLUP in the remaining 3 cases. When σ_i^2 is unknown, the EBLUP appears to outperform HB in 13 cases, yield similar results in 2 cases and only once does the HB show slightly smaller CV than the EBLUP. The EBLUP clearly has lower CVs than the HB when σ_i^2 is unknown. The EBLUP approach seems more efficient than the HB approach. The disparity between the two approaches is more evident when σ_i^2 is unknown in the model. Our result indicates that in the crop data example the EBLUP approach may lead to severe underestimation of the true MSE especially when the sampling variances σ_i^2 are unknown. With the small sample sizes, the g_4 term used in the MSE estimator does not effectively capture all the variability of estimating σ_i^2 by s_i^2 whereas, the HB approach leads to exact and complete inferences in the sense that the posterior variances automatically take into account the variability associated with the unknown parameters in the model.

Milk data

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. The HB estimation approach is thus an improvement on the direct survey estimates. Both the HB and EBLUP approaches lead to almost the same point estimates and CVs. With the sample sizes being larger in the milk data both approaches perform very similarly regardless of the handling of σ_i^2, leading to an average of about 25% CV reduction. Thus both the HB and EBLUP approaches uniformly improve efficiency on the direct survey estimates in this example.

For more detailed tables and results for the crop and milk data, we refer to You and Chapman (2003).

4. CONCLUDING REMARKS

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 known and where they are estimated by unbiased estimators s_i^2, using the HB and the EBLUP approaches. The full HB approach with the Gibbs sampling method automatically takes into account the extra uncertainty associated with the estimation of σ_i^2. The EBLUP approach has an extra term added to the MSE estimator to account for this uncertainty. When the area-specific sample sizes are small, our data analysis has shown that the EBLUP approach tends to underestimate the MSE when the sampling variances are estimated by s_i^2. Thus when both the area-specific sample sizes and the number of small areas are small, it is better to use the proposed complete HB approach. When the area-specific sample sizes and the number of small areas are relatively large, both the HB and the EBLUP approaches perform equally well and both lead to smaller CVs compared to the direct survey estimates. 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 (2002). 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 Labour Force Survey (LFS) unemployment rate estimation (You, Rao and Gambino, 2003).

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., Rao, J.N.K. and Smith, D.D. (2002). On measures of uncertainty of small area estimators in the Fay-Herriot model. Technical Report, University of Georgia, Athens.

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.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators, *Journal of the American Statistical Association*, 85, 163 -171.

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

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. (2000). Topics in Small Area Estimation with Applications to the National Resources Inventory, Ph.D. dissertation, Iowa State University.

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 Chapman, B. (2003). Small area estimation using area level models and estimated sampling variances. Statistics Canada Methodology Branch working paper, HSMD-2003-005E.

You, Y. and Rao, J. N. K. (2002). Small area estimation using unmatched sampling and linking models. *The Canadian Journal of Statistics, Vol.30, No. 1*, 3-15.

You, Y., Rao, J.N.K. and Dick, P. (2002). Benchmarking hierarchical Bayes small area estimators with application in census undercoverage estimation. *Statistical Society of Canada 2002 Proceedings of the Survey Methods Section [CD-ROM]*, 81-86.

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.