SSC Annual Meeting, July 1995
Proceedings of the Survey Methods Section

DISCUSSION OF PAPERS BY ARORA/LAHIRI AND RIVEST

J.N.K. Rao[1]

The two papers presented in this session make important contributions to small area estimation. I will first discuss the paper "On the superiority of the Bayesian method over the BLUP in small area estimation problems" by Vipin Arora and P. Lahiri and then turn to the paper "A composite estimator for provincial undercoverage in the Canadian Census" by Louis-Paul Rivest.

Arora and Lahiri study a realistic small area model (Model 1) due to Kleffe and Rao (1992) that relaxes the customary assumption of constant error variance. This model is not covered by the general mixed linear model with fixed variance components under which the BLUP estimator and the Bayes estimator are identical, assuming the hyperparameters are known. Under model 1, Arora and Lahiri compare the BLUP estimator of Kleffe and Rao with the Bayes estimator and show the superiority of the Bayes estimator with respect to MSE. This result is quite surprising, but a careful comparison shows that the Bayes estimator cannot be calculated without actually specifying the prior density of the error variances σ_i. On the other hand, the BLUP requires only the specification of the first two moments of σ_i. Thus the BLUP is obtained under weaker assumptions. It should also be noted that the Bayes estimator is not linear under model 1.

In the second part of the paper, Arora and Lahiri extend model 1 to incorporate sampling weights and perform a Hierarchical Bayes (HB) analysis. In particular, they obtain the posterior mean and variance using Gibbs sampling. The results are then applied to data from the U.S. Consumer Expenditure Survey. The posterior variance is shown to be smaller than the estimated variance of the direct estimator Y_i for all the 43 small areas. (It appears to me that Y_i should be $\Sigma_l W_{il} Z_{il} / \Sigma_l W_{il}$ and not $\Sigma_l W_{il} Z_{il}$ as stated in the paper). The estimated variance, $(n_i - 1)^{-1} S_i^2 K_i$, of Y_i is based on assumption 1 of model 2 which does not reflect the design effect due to clustering etc., and assumption 1 is used in the HB analysis to get the posterior variance. It would be useful to study the sensitivity of the posterior variance to deviations from assumption 1 of model 2.

Arora and Lahiri also conduct an evaluation study by treating the sample data as a pseudo-population and considering eight 12.5% samples from this pseudo-population. Under the criterion of average squared relative deviation (ASRD) they show the superiority of the HB estimator over the direct estimator for all eight samples, and over the empirical BLUP (EBLUP) estimator for six out of the eight samples. This evaluation study is very useful because the ASRD criterion is not model-dependent. A cross-validation may also be useful because the eight samples together constitute the pseudo-population.

Rivest addresses the important problem of adjustment for provincial undercoverage in the 1991 Canadian Census. Undercoverage estimates $\hat{U}_i$ for the 10 provinces are obtained from the Reverse Record Check (RRC), and the estimates $\hat{U}_i$ may be biased. Rivest proposes a composite estimator $\hat{V}_i = (1-\alpha)(\sum_l \hat{U}_l)\ p_i + \alpha \hat{U}_i$, where $p_i = Y_i / \sum_i Y_i$ is the provincial share i and Y_i is the census count for the i-th province. This composite estimator uses the same weight, α, for all the provinces unlike the empirical Bayes (EB) estimator, studied by P. Dick (*Survey Methodology*, 1995, pp. 45-54), which utilizes auxiliary information and local (provincial) effects in modelling the undercount.

Under the criterion of weighted MSE (WMSE), Rivest shows that the expected WMSE with respect to the distribution of $\hat{U}$, generated by the RRC design, is minimized by the same value of α for both totals T_i and shares $T_i / \sum_i T_i$, where T_i is the unknown true provincial total. This optimal α-value takes account of the bias, b_{ui}, in $\hat{U}_i$. The optimal α is then estimated (see eq. (2)), assuming that the bias b_{ui} and the sampling variance σ_i^2 are known. Using the estimated α in WMSE, the expected WMSE is evaluated and an approximately unbiased estimator of WMSE is obtained.

The composite estimator $\hat{V}_i$ with estimated α is then compared to the full adjustment $\hat{U}_i$, using the estimated WMSE as a measure of efficiency and assuming no bias ($b_{ui} = 0$). Rivest makes an important point that estimating α is costly when the undercoverage rate in a large province is markedly different from the national rate. The numerical results from the 1991 Census study confirm this observation: efficiency of the composite estimator dropped from 1.17 to 1.03 when accounting for the estimation of α. This is attributed to the larger undercoverage rate

---

[1] J.N.K. Rao, Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada, K1S 5B6.

(3.64%) in Ontario than the national undercoverage rate (2.84%). Rivest also makes efficiency comparisons in the presence of bias and concludes that the composite estimator $\hat{V}_i$ gains in efficiency over the full adjustment $\hat{U}_i$ when bias is positively correlated with undercoverage.

Rivest makes no specific recommendation, but the small gain in efficiency (3%) under the restrictive assumption of known σ_i^2 does not seem to justify the use of composite estimator $\hat{V}_i$ over the full adjustment $\hat{U}_i$. The efficiency gain could be even smaller if one accounts for the estimation of σ_i^2. Also, the overall efficiency criterion, EWMSE, is less relevant to a given province than a province-specific efficiency criterion. The EB and Hierarchial Bayes (HB) methods have an advantage in providing province-specific measures of uncertainty by utilizing auxiliary information and random provincial effects. Of course, these methods are based on assumed random effects models unlike the composite estimator of Rivest based on EWMSE, but such models can be validated through model diagnostics, etc.