and found that, as expected, there are high positive correlations between yields in neighboring states and between agronomically related crops that are grown in overlapping regions. This research, however, was conducted at the state rather than county level; it is still an open question whether similar relations will be useful at the county level.

Another area for research is in using the historical data on crop production in current county estimates. A natural way to use this information would be in a Bayesian setting such as the hierarchical Bayes estimates described in Section 5.3 of Ghosh and Rao's paper. Indeed, it seems surprising that a noninformative prior would be used in small-area estimation problems involving census data or data from continuing surveys; there is certainly a wealth of information on which to base an informative prior.

Finally, I would like to mention a success story in research in the production of county estimates. Ghosh and Rao describe the experimental research of Battese, Harter and Fuller (1988) on county estimation of crop production using satellite data. This year, for the first time, Arkansas is using satellite data to aid in production of crop acreage estimates as part of their county estimates program. Over the next few years, other states are expected to begin using such data to aid in the production of their crop acreage estimates.

ACKNOWLEDGMENT

This research was supported in part by the USDA/NASS under Cooperative Agreements 58-3AEU-9-80040 and 43-3AEU-3-80083. The author takes sole responsibility for the work.

Comment

Ib Thomsen

It takes talent and hard work to provide an overview and evaluation of a rapidly evolving subject like small area estimation. In my opinion the authors have succeeded in doing this, and I want to congratulate them with a very useful review. In many statistical offices, substantial methodological work is being done to find suitable estimators for small areas. People involved in such work will be grateful to Ghosh and Rao for their present contribution.

Below I shall communicate some experiences gained when developing and using small area estimates within Statistics Norway. But first a few comments to the example given in Section 6 of the paper. In this example a synthetic population is constructed by fitting a nested error regression model to a business population. For this synthetic population, the EBLUB (or EB) and the HB estimators are shown to produce small area estimators which are superior to the ratio-synthetic and a sample-size dependent estimator. As pointed out by the authors, this demonstrates the advantages of using EBLUB or HB estimators when the model fits the data well. A question remains concerning the robustness of these estimators as compared to the simpler sample-size dependent estimator. A column in Table 3 showing the small area means of the real business population could have thrown some light on the robustness of the estimators studied in the paper.

At Statistics Norway, small area estimators have been used for some years now (Laake, 1978). In the beginning we concentrated on synthetic estimators, but more recently composite estimators are being used. In what follows some of our experiences concerning the feasibility of the EB estimator are presented.

I shall look at a very simple situation in which θ_i, $(i = 1, \dots, T)$ is a small area parameter, and $\overline{X}_i$, $(i = 1, \dots, T)$ is a direct estimator such that

$$E(\overline{X}_i|\theta_i) = \theta_i \quad i = 1, \dots, T.$$

The parameters $\theta_1, \theta_2, \dots, \theta_T$ are considered realizations of a random variable with unknown distribution $G(\cdot)$. The mean μ and variance σ^2 are assumed to be known or that estimates are available. For a set of small areas, unbiased estimators $\overline{X}_1, \dots, \overline{X}_T$ are available with conditional distributions equal to the binomial.

When $G(\theta)$ is unknown, empirical Bayes estimators generally employ $(\overline{X}_1, \dots, \overline{X}_T)$ to estimate $E(\theta|\overline{X}_1, \dots, \overline{X}_T)$. However, for many distribution, $E(\theta|\overline{X}_1, \dots, \overline{X}_T)$ cannot be consistently estimated un-

Ib Thomsen is Director of Research, Statistics Norway, and Professor of Statistics, University of Oslo, P.B. 8131 Dep, 0033 Oslo, Norway.

less other assumptions are made. Therefore, one often restricts attention to linear estimators, $c = a\overline{X} + b$. Within this class, the estimator which minimizes the mean squared error depends only upon the first two prior moments, both of which can often be estimated with $(\overline{X}_1, \ldots, \overline{X}_T)$. The optimal linear estimator is often the same as the unrestricted Bayes estimator derived under a conjugate prior (Rao, 1976). When the conditional distribution of $\overline{X}_i$ is binomial, the optimal linear estimator is a composite estimator,

$$c_i = W_i\overline{X}_i + (1 - W_i)\mu ,$$

where

$$W_i = \sigma^2\{(1 - 1/n_i)\sigma^2 + \mu(1 - \mu)/n_i\}^{-1}$$

and n_i denotes the number of observations from small area i (Spjøtvoll and Thomsen, 1987). With these weights we have that

$$(1)\quad E\left\{(1/T)\sum_{i=1}^{T}(c_i - \mu)^2\right\} = \sigma^2(1/T)\sum_{i=1}^{T} W_i \leq \sigma^2.$$

It follows that the variation between the small area estimators can be much smaller than the prior known variance. I have often observed this phenomenon in practice; a consequence is usually that the range of the small area estimators is much smaller than expected. (Expectations are based on information outside the sample.) In practice the parameter σ^2 is often of great importance in itself. As said in the introduction, "Increasing concern with issues of distribution, equity and disparity (Brackstone, 1987)." To me, this means that the disparity between the small area is important and should be easily read from a table presenting small-area estimators. As mentioned by Ghosh and Rao, there are composite estimators which have the same expectation and variance as the prior distribution, one of which is simply to use $\{W_i\}^{1/2}$ instead of W_i as weights in the composite estimator.

When area-specific auxiliary information is available and a model like (4.1) in the paper is used, I have often observed a similar "overshrinkage" as under the simpler model above. An inequality similar to (1) can be found under model (4.1), but now σ^2 denotes the variance of the residual in equation (4.1). Again $\{W_i\}^{1/2}$ can be used to avoid "overshrinkage".

Due to the often observed "overshrinkage" and the fact that our models seem too complicated to many of our users of small-area estimators, I have often found it very difficult to make them use the optimal estimators presented in the paper. On the other hand, a number of sample-size dependent estimators are more easily "sold" to the user and therefore more used up until now.

In Statistics Norway a number of administrative registers are available and used to construct small-area estimators. In many cases it is natural to use nested error regression models. However, progress in this area has been slow due to difficulties concerning model diagnostics for linear models involving random effects. I therefore find Section 7.1 particularly interesting and shall use this section intensively in our further hunt for feasible small area estimates.

Rejoinder

M. Ghosh and J. N. K. Rao

We thank the discussants for their insightful comments as well as for providing various extensions of the models and the methods reviewed in our paper. These expert commentaries have brought out many diverse issues and concerns related to small area estimation, particularly on the model-based methods.

Several discussants emphasised the importance of model diagnostics in the context of small area estimation. We agree wholeheartedly with the discussants on this issue. As noted in Section 7.1 of our article, the literature on this topic is not extensive, unlike standard regression diagnostics. We hope that future research on small area estimation will give greater emphasis to model validation issues.

A second concern expressed by some of the discussants is that the composite estimators typically used for small area estimation may "overshrink" towards a synthetic estimator. Thomsen, in his discussion, suggests that a larger weight should be given to the direct estimator. We agree with his suggestion but are hesitant to recommend blanket use of the weight $W_i^{1/2}$, instead of W_i, to the direct estimator ($0 < W_i < 1$). We believe that the weight should be determined adaptively meeting certain optimality criteria as in Louis (1984) and Ghosh (1992). Cressie and Kaiser, in their discussion, address con-