out that the solution yields the constrained empirical Bayes estimates obtained by Cressie (1986), although no Bayes optimality criterion is invoked by the authors.

The multivariate version of (10) is

$$\hat{\theta} = A\tilde{\theta} + b, \tag{11}$$

where A is an $m \times m$ matrix and b is an $m \times 1$ vector. Upon specifying that $E(\hat{\theta}) = E(\theta)$ and $\text{var}(\hat{\theta}) = \text{var}(\theta)$, Cressie (1990b, 1992) obtains a multivariate constrained estimator. In the notation of (1), $\theta = \mu$, $E(\theta) = X\beta$, $\tilde{\theta} = y$, $E(y \mid \theta) = \theta$, $\text{var}(y \mid \theta) = \Sigma$, and $\text{var}(y) = \Sigma + \Gamma$. Then the multivariate constrained estimator for model (1), analogous to Spjøtvoll and Thomsen's, is given by (11), where

$$A = \Gamma^{1/2}(\Sigma + \Gamma)^{-1/2} \tag{12}$$

and

$$b = \{I - \Gamma^{1/2}(\Sigma + \Gamma)^{-1/2}\}X\beta. \tag{13}$$

Notice that $\hat{\theta}$ given by (11), (12) and (13) does not shrink y towards $X\beta$ as far as the Bayes estimator θ^* does (where $A = \Gamma(\Sigma + \Gamma)^{-1}$ and $b = (I - A)X\beta$).

In an elegant paper, Ghosh (1992) derives a multivariate constrained *Bayes* estimator for model (1):

$$\theta^{@} = \{a + (1 - a)\underset{\sim}{1}\underset{\sim}{1}'/m\}\theta^*, \tag{14}$$

where

$$a = \left[\text{trace}\{(I - \underset{\sim}{1}\underset{\sim}{1}'/m)V\} \left(\sum_{i=1}^{m}(\theta_i^* - \bar{\theta}^*)^2\right)^{-1} + 1\right]^{1/2},$$

$$\theta^* = E(\theta \mid y) = \{\Gamma(\Sigma + \Gamma)^{-1}\}y + (I - \Gamma(\Sigma + \Gamma)^{-1}\}X\beta,$$

$$V = \text{var}(\theta \mid y) = \Gamma\{I - \Gamma(\Sigma + \Gamma)^{-1}\}\Gamma.$$

The vector $\theta^{@}$ has the property that it minimizes $E(\sum_{i=1}^{m}(\theta_i - t_i)^2 \mid y)$ with respect to t and subject to conditions that match first and second sample moments of t with those same moments of θ conditional on y. Cressie's proposal given by (11), (12) and (13) does not invoke any optimality conditions and so is likely to be less efficient than Ghosh's estimator (14).

Constrained Bayes estimation for more general models, such as GLMs, is presented by Ghosh (1992), although from an essentially univariate point of view. Our earlier comment, that we do not have flexible ways to model lack of independence in nonlinear, nonnormal models, is equally appropriate here.

Finally, we agree with the authors' comment about the importance of small area estimation in medical geography. A good source for recent research in this area is the May 1993 Supplement Issue of the journal *Medical Care* (Proceedings of the Fourth Biennial Regenstrief Conference, "Methods for Comparing Patterns of Care," October 27–29, 1991). We are working on incorporating spatial variation and dependence into statistical methods for these and other small area estimation problems.

ACKNOWLEDGMENT

Partial support came from the Office of Naval Research Grant N00014-93-1-0001, NSF Grant DMS-92-04521 and the National Security Agency Grant MDA904-92-H-3021.

Comment

D. Holt

The paper by Ghosh and Rao is a valuable summary of recent developments using empirical Bayes and hierarchical Bayes methods for making small area estimates. The need for methods which make provision for local variation while pooling information across areas is well established. The review is a thorough appraisal of the methods and their properties, and the numerical results reinforce earlier results which demonstrate that these methods are preferable to others such as synthetic estimation and sample size dependent estimation.

The value of these approaches is not simply in their ability to provide point estimates for each small area which, on average, have better precision. A very important additional factor is that a measure of precision (MSE) and an estimator of this can be

D. Holt is Professor, Department of Social Statistics, University of Southampton, Southampton S09 5NH, United Kingdom.

developed for each small area separately. This is extremely important since the precision of each small area estimate will depend upon a number of factors including the sample size and the distribution of the area specific covariate values as well as the method of estimation itself. Indeed, no one method of estimation will be necessarily uniformly superior for all small areas; and any choice of estimator will result in a loss of precision for some areas as well as gains for others.

This point leads to the issue of which measures of precision are appropriate and how to present numerical results. Ghosh and Rao present a single point estimate for each small area from a single sample. It is, in effect, a simulation of size one. There are advantages to this approach since we can make direct comparisons between the estimates and true small area means in each case. I will return to this point, but first let us consider the numerical results as presented.

The choice of measures, relative absolute error and squared error for each small area separately, are both natural. The first represents a measure analogous to coefficient of variation and the second represents MSE. However, it is dangerous to summarize these measures into a single average across all small areas without paying some attention to the distribution. In Table 3, for example, one notices that for all four estimators considered the point estimates are less than the true values for 13 of the 16 small areas. Also for each estimator the two measures are extremely variable across the small areas. To consider the sample dependent estimator, for example, Ghosh and Rao comment that in terms of average relative error it is similar to EBLUP and HB but in terms of average squared error it is inferior. However, one may derive from the table that 58% of the ASE for this estimator is derived from the last small area. For each of the estimators, the distribution of relative absolute error and squared error is informative and important.

When one considers the distribution of performance measures for each small area, then the reader cannot separate systematic performance from random error since the results represent a simulation of size one. Is it the case, for example, that in small area 4 the tiny deviations for the ratio synthetic and sample size dependent estimators and the much larger deviations for EBLUP and HB reflect a true difference in performance or is this random fluctuation? Would it not have been better to produce measures which were based upon a set of repeated simulations and which could have included an average bias, average relative absolute error and mean squared error for each small area separately? If this had been done then comparisons could have been made between estimators for each small area separately (e.g., comparison of average bias, MSE, etc.). The distribution of these comparative measures and their overall summary could then have been considered.

This rather simple comment raises a rather fundamental issue about the framework for measures of performance and how numerical simulations should be designed. The measures of precision (e.g., MSE) given in Section 5 of Ghosh and Rao's paper are essentially model based. To some extent assumptions of normality are required but the authors comment about the robustness of the methods. However the properties within the model framework are conditional on the values of the auxiliary variable (x_{ij}) and the sample size achieved in each small area (n_i). Within the predictive framework many analysts would prefer measures of precision which condition on the achieved sample in this way. Survey practitioners, on the other hand, and anyone considering the choice of estimation method in advance of the survey being analyzed will want to understand the properties of estimators across of range of circumstances. This creates a dilemma for the theoretician who wishes to demonstrate the comparative properties of alternative estimation methods, using simulations.

Should Ghosh and Rao:

(a) Fix the sample values of n_i, $\{x_{ij}\}$ and a single randomly generated value of the small area effect ν_i and carry out repeated simulations to obtain the properties of each estimator for each small area?
(b) Fix the sample values of n_i and simulate repeated sample selections from the population of each small area?
(c) Draw repeated random samples from the whole population without restriction?

The model based MSE given in Section 5 will be constant under (a) but not under (b) or (c). An analyst might be more interested in (a) but would want to be assured that the results did not depend on the particular choice of the sample configuration. Many survey practitioners would lean towards (b) or (c). Perhaps the practical solution is to draw several samples under (b) or (c); and then for each one selected, carry out repeated simulations under (a). By presenting the results from one simulation, Ghosh and Rao effectively avoid all of these issues.

Finally, to turn to a separate issue, the models described in Section 4 provide for local differences in the small area means by introducing a random effect, ν_i, for each small area. This is a random term which is the same for all units in the small area and essentially introduces a random effect for the intercept of the linear model. This approach may be extended and the two model frameworks for equations (4.1) and (4.2) essentially integrated. Equation (4.5) may be generalized to allow all (or any) of the regression coefficients including the intercept to be random. Furthermore, small area level variables (z_i) may be used to explain some of the between small area variation:

$$y_i = x_i\beta_{1i} + e_i,$$
$$\beta_{1i} = z_i\gamma + \nu_i;$$

X_i is the $N_i \times (p+1)$ matrix of unit level covariates (including an intercept) and z_i is the $(p+1)\times q$ matrix of small area level variables. Here γ is the vector of length q of fixed coefficients and $\nu_i = (\nu_{i0}, \ldots, \nu_{ip})^T$ is a vector of length $p+1$ of random effects for the ith small area. In the general form the ν_i are independent between small areas but may have a joint distribution within each small area with $E(\nu_i) = 0$ and $V(\nu_i) = \Omega$:

$$\Omega = \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \cdots & \sigma_{0p} \\ \sigma_{10} & \sigma_1^2 & \cdots & \sigma_{1p} \\ \vdots & & & \\ \sigma_{p0} & \sigma_{p1} & \cdots & \sigma_p^2 \end{bmatrix}$$

A special case is when the random effects are uncorrelated so that Ω is diagonal.

The use of area level variables, Z_i, to help explain the between area variation should help when the sample size in a small area is small. Also this more general model effectively integrates the use of unit level and area level covariates into a single model. Holt and Moura (1993) provide point estimates and expressions for MSE following the framework of Prasad and Rao (1990).

The use of extra random effects for the regression coefficients gives greater flexibility. If the unit level covariate is a set of dummy variables signifying group membership, for example, then this approach will allow a set of correlated and heteroscedastic random effects for the group means in each small area rather than a single random effect for all subjects.

The introduction of a random effect for the regression coefficient of a continuous covariate is likely to have more impact when the individual covariate values x_{ij} are variable within each small area. Judging by the values displayed in Table 2 where the values of x_{ij} vary greatly, it is possible that a more general model would provide even greater gains in precision for the empirical example which Ghosh and Rao consider.

Comment

Wesley L. Schaible and Robert J. Casady

Professors Ghosh and Rao have provided us with an excellent, comprehensive review of indirect estimation methods which have been suggested for the production of estimates for small areas and other domains. They make a timely contribution by reviewing and comparing a number of new methods which have recently appeared in the literature as well as updating previous work on some of the more established approaches. Demographic methods, synthetic and related estimators, empirical Bayes estimators, hierarchical Bayes estimators and empirical best linear unbiased prediction methods are thoroughly discussed; evidence that the Bayes and empirical prediction methods have advantages over the others is presented. Special problems in the application of small area estimation methods are also addressed. This is an extremely important issue and additional discussion would have been desirable. In our comments, we will expand on this subject by discussing some of the characteristics of indirect estimators and some specific practical problems associated with their use. In addition, we will attempt to state in general terms what we believe to be the fundamental problem associated with the application of small area estimation methodology.

Very generally speaking, applications of indirect estimation methods fall into one of three categories:

1. An indirect estimator is used to estimate a population parameter;
2. an indirect procedure is used to modify a direct estimator of a population parameter (e.g., a direct estimator that incorporates indirectly estimated post-stratification controls or seasonal

Wesley L. Schaible is Associate Commissioner and Robert J. Casady is Senior Mathematical Statistician, Office of Research and Evaluation, Bureau of Labor Statistics, 2 Massachusetts Avenue N.E., Washington, DC 20212.