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.

adjustment procedures); and

3. an indirect estimator is used to estimate the variance of an estimator (e.g., a generalized variance function).

Essentially all of the small area estimation literature focuses on the first category of applications; the authors' review and our comments will do likewise.

The authors refer to the Federal Committee on Statistical Methodology report, "Indirect Estimators in Federal Programs." This report focuses on applications of indirect estimators and provides an interesting supplement to the paper under discussion. Some of the characteristics of indirect estimators and practical problems associated with their application, which are summarized in this report, are mentioned below:

- A domain and time specific model is implicitly assumed to be true when analyses among domains and over time are conducted. From a best linear unbiased prediction point of view, a domain and time specific model leads to a best linear unbiased direct estimator and also defines a family of indirect models which allow strength to be borrowed from other domains and/or time periods. The direct estimator is unbiased, not only under the domain and time specific model, but also under each of the corresponding indirect models. However, the best linear unbiased indirect estimators associated with the indirect models are not unbiased under the original domain and time specific model. This indirect estimator bias under the more plausible domain and time specific model adds to the uneasiness associated with the use of indirect estimators. It is the primary reason that indirect estimators are generally considered only when resources prohibit the use of direct estimators of adequate reliability.
- The variance of an indirect estimator will be smaller than that of the corresponding direct estimator since the indirect estimator not only incorporates observations of the variable of interest from the domain and time of concern but also from other domains and/or time periods.
- If the stochastic model underlying an indirect estimator is a satisfactory representation of reality, then the mean square error of the indirect estimator will likely be smaller than that of the corresponding direct estimator. However, many indirect estimators require strong model assumptions that may not be satisfied in most applications. If this is the case, then the mean square error of the indirect estimator may in fact be larger than the variance of the direct estimator. Although estimation of variances and (more importantly) mean square errors of indirect estimators has received attention, the estimation of a meaningful measure of error for a single small area remains a problem.
- Usually the task at hand is to produce estimates for a number of small areas simultaneously. There is considerable empirical evidence suggesting that the size of an error of an indirect estimator depends on the relationship of the area population value and population values of the other areas from which strength is borrowed. For example, the error in an indirect estimate for a small area with a very large population value is likely to be relatively large and negative, so that the estimate is closer to the population values of small areas that are not so large. This characteristic is not displayed to the same extent by all indirect estimators, and, as discussed by the authors, constrained estimators have been recently suggested to help address this problem.

The authors discuss the extremely important problem of model evaluation and suggest model diagnostics to help in the search for models that fit the data well. Until recently, model diagnostics have not played a major role in the evaluation of indirect estimators. Even though this approach is not free from dangers such as overfitting, especially when data sets are small, practitioners should make more use of these tools in estimator evaluation. Most government survey systems are designed to collect data and produce estimates periodically, yet the potential for continuing, routine estimator evaluations has not been fully explored. Problems of overfitting and small data sets associated with model diagnostics can be at least partially overcome by continuing evaluations.

We now turn to what we believe to be the general fundamental problem associated with the application of indirect estimation methods. A truly plausible model would depend on domain and time specific parameters, but indirect estimators are associated with models that contain one or more parameters that do not vary either over domains, time or both. In addition, in most practical applications, the statistician is pragmatically forced to settle for a stochastic model determined by the ancillary variables which are available. Models based on such expediency instill little confidence in either the producers or consumers of the estimates. Consequently, everyone concerned is usually convinced that the estimation process produces biased estimates; and, invariably, an empirical study is mounted to evaluate the average mean squared error (or some other appropriate loss function) across the range of small areas. Such studies depend on "target values" for the parameter of interest for each small area, and generally accepted values of these target values are rarely, if ever, available (if they were, then there would be no need for indirect estimates). Thus, evaluation studies tend to produce conflicting and ambiguous results and leave all concerned less than completely satisfied. A good case in point are the many problems associated with use of a synthetic estimator to adjuste for state population undercounts in the 1990 census.

Having emphasized some of the problems associated with applications of indirect estimators, we should also mention the obvious fact that these estimation methods provide practitioners with many useful tools. Challenging research issues concerning the estimation of meaningful measures of error remain; without such measures, we must be cautious regarding inferences and actions based on these estimators. Nevertheless, in many applications, these methods provide us with an attractive alternative to the use of high variance direct estimates or, in some cases, no estimates at all.

Comment

Avinash C. Singh

The review paper of Ghosh and Rao fills a very important gap by giving a comprehensive and coherent picture of various developments in small area estimation over the last twenty years. This area is fascinating for at least three reasons: (1) there is a great demand for small area statistics by both government and private sectors for purposes of planning and policy analysis; (2) the small area problem provides a fertile ground for theoretical and applied research; and (3) the problem has attracted the attention of both Bayesians and frequentists because both approaches arise naturally and often seem to give similar results.

The main theme of my discussion is to compare and contrast the Bayesian and frequentist solutions to the problem of small area estimation. Why is it that for this problem the two approaches to statistical inference seem to converge in many practical examples including the one considered by Ghosh and Rao; that is, they provide similar results for both point estimates and the corresponding measures of uncertainty? Can we make some general statements about the similarity between the two approaches for small area estimation? How do their frequentist properties compare? Questions about the frequentist properties of some empirical Bayes methods are also raised by Ghosh and Rao in Section 5.2. Although the task of making exact comparisons is a difficult one, it is possible to make asymptotic comparisons for large m—the number of small areas. This will be the focus of my discussion.

Avinash C. Singh is Senior Methodologist, Methods Development and Analysis Section, Social Survey Methods Division, Statistics Canada, Ottawa K1A 0T6. He is also Adjunct Research Professor, Department of Mathematics and Statistics, Carleton University, Ottawa K1S 5B6.

1. MODEL REFORMULATION

As discussed in the review paper of Robinson (1991), understanding of procedures for estimating fixed and random effects helps to bridge the apparent gulf between the Bayesian and frequentist schools of thought. The present discussion will also strengthen this point. First, it will be convenient for our purposes to reformulate the model with fixed and random effects for small area estimation. Now, the general mixed linear model is given by

$$y = X\beta + Z\nu + \epsilon \tag{1}$$

where y is the n-vector of element-level data; X and Z are known matrices of orders $n \times p$ and $n \times m$, respectively, with rank $(X) = p$; β is a p-vector of fixed effects; ν is a m-vector of small area specific random effects and ϵ is a n-vector of random errors independent of ν such that $\nu \sim$ WS $(0, G)$, $\epsilon \sim$ WS $(0, R)$. The abbreviation "WS" stands for "wide sense"; that is, the distribution is specified only up to the first two moments. The covariance matrices G and R depend on some parameters λ called variance components. For the reformulation of (1), we will regard the fixed effects β as random with mean 0 and covariance matrix $\sigma_\beta^2 I$ where $\sigma_\beta^2 \to \infty$. Thus, the limiting prior distribution of β is uniform (improper) which is commonly assumed in the Bayesian approach. The reformulation is useful for computational convenience as well as for making connections