Statistical Science
1994, Vol. 9, No. 1, 55–93

Small Area Estimation: An Appraisal

M. Ghosh and J. N. K. Rao

Abstract. Small area estimation is becoming important in survey sampling due to a growing demand for reliable small area statistics from both public and private sectors. It is now widely recognized that direct survey estimates for small areas are likely to yield unacceptably large standard errors due to the smallness of sample sizes in the areas. This makes it necessary to "borrow strength" from related areas to find more accurate estimates for a given area or, simultaneously, for several areas. This has led to the development of alternative methods such as synthetic, sample size dependent, empirical best linear unbiased prediction, empirical Bayes and hierarchical Bayes estimation. The present article is largely an appraisal of some of these methods. The performance of these methods is also evaluated using some synthetic data resembling a business population. Empirical best linear unbiased prediction as well as empirical and hierarchical Bayes, for most purposes, seem to have a distinct advantage over other methods.

Key words and phrases: Borrowing strength, demographic methods, empirical Bayes, empirical best linear unbiased prediction, hierarchical Bayes, synthetic estimation

1. INTRODUCTION

The terms "small area" and "local area" are commonly used to denote a small geographical area, such as a county, a municipality or a census division. They may also describe a "small domain," i.e., a small subpopulation such as a specific age-sex-race group of people within a large geographical area. In this paper, we use these terms interchangeably.

The use of small area statistics originated several centuries ago. Brackstone (1987) mentions the existence of such statistics in 11th century England and 17th century Canada. Many other countries may well have similar early histories. However, these early small area statistics were all based either on a census or on administrative records aiming at complete enumeration.

For the past few decades, sample surveys, for most purposes, have taken the place of complete enumeration or census as a more cost-effective means of obtaining information on wide-ranging topics of interest at frequent intervals over time. Sample survey data certainly can be used to derive reliable estimators of totals and means for large areas or domains. However, the usual direct survey estimators for a small area, based on data only from the sample units in the area, are likely to yield unacceptably large standard errors due to the unduly small size of the sample in the area. Sample sizes for small areas are typically small because the overall sample size in a survey is usually determined to provide specific accuracy at a much higher level of aggregation than that of small areas. Thus, until recently, the use of survey data in developing reliable small area statistics, possibly in conjunction with the census and administrative data, has received very little attention.

Things have changed significantly during the last few years, largely due to a growing demand for reliable small area statistics from both the public and private sectors. These days, in many countries including the United States and Canada, there is "increasing government concern with issues of distribution, equity and disparity" (Brackstone, 1987). For example, there may exist geographical subgroups within a given population that are far below the average in certain respects, and need definite upgrading. Before taking remedial action, there is a need to identify such regions, and accordingly, one must have statistical data at the relevant geographical levels. Small area statistics are also needed

M. Ghosh is Professor, Department of Statistics, University of Florida, Gainesville, Florida 32611-2049. J. N. K. Rao is Professor of Statistics, Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada K1S 5B6.

in the apportionment of government funds, and in regional and city planning. In addition, there are demands from the private sector since the policy-making of many businesses and industries relies on local socio-economic conditions. Thus, the need for small area statistics can arise from diverse sources.

Demands of the type described above could not have been met without significant advances in statistical data processing. Fortunately, with the advent of high-speed computers, fast processing of large data sets made feasible the provision of timely data for small areas. In addition, several powerful statistical methods with sound theoretical foundation have emerged for the analysis of local area data. Such methods "borrow strength" from related or similar small areas through explicit or implicit models that connect the small areas via supplementary data (e.g., census and administrative records). However, these methods are not readily available in a package to the user, and a unified presentation which compares and contrasts the competing methods has not been attempted before.

Earlier reviews on the topic of small area estimation focussed on demographic methods for population estimation in post-censual years. Morrison (1971) covers the pre-1970 period very well, including a bibliography. National Research Council (1980) provides detailed information as well as a critical evaluation of the Census Bureau's procedures for making post-censual estimates of the population and per capita income for local areas. Their document was the report of a panel on small-area estimates of population and income set up by the Committee on National Statistics at the request of the Census Bureau and the Office of Revenue Sharing of the U.S. Department of Treasury. This document also assessed the "levels of accuracy of current estimates in light of the uses made of them and of the effect of potential errors on these uses." Purcell and Kish (1979) review demographic methods as well as statistical methods of estimation for small domains. An excellent review provided by Zidek (1982) introduces a criterion that can be used to evaluate the relative performance of different methods for estimating the populations of local areas. McCullagh and Zidek (1987) elaborate this criterion more fully. Statistics Canada (1987) provides an overview and evaluation of the population estimation methods used in Canada.

Prompted by the growing demand for reliable small area statistics, several symposia and workshops were also organized in recent years, and some of the proceedings have also been published: National Institute on Drug Abuse, Princeton Conference (see National Institute on Drug Abuse, 1979), International Symposium on Small Area Statistics, Ottawa [see Platek et al. (1987) for the invited papers and Platek and Singh (1986) for the contributed papers presented at the symposium]; International Symposium on Small Area Statistics, New Orleans, 1988, organized by the National Center for Health Statistics; Workshop on Small Area Estimates for Military Personnel Planning, Washington, D.C., 1989, organized by the Committee on National Statistics; International Scientific Conference on Small Area Statistics and Survey Designs, Warsaw, Poland, 1992, (see Kalton, Kordos and Platek, 1993). The published proceedings listed above provide an excellent collection of both theoretical and application papers.

Reviews by Rao (1986) and Chaudhuri (1992) cover more recent techniques as well as traditional methods of small area estimation. Schaible (1992) provides an excellent account of small area estimators used in U.S. Federal programs (see NTIS, 1993, for a full report prepared by the Subcommittee on Small Area Estimation of the Federal Committee on Statistical Methodology, Office of Management and Budget).

The present article considerably updates earlier reviews by introducing several recent techniques and evaluating them in the light of practical considerations. Particularly noteworthy among the newer methods are the empirical Bayes (EB), hierarchical Bayes (HB) and empirical best linear unbiased prediction (EBLUP) procedures which have made significant impact on small area estimation during the past decade. Before discussing these methods in the sequel, it might be useful to mention a few important applications of small area estimation methods as motivating examples.

As our first example, we cite the Federal-State Cooperative Program (FSCP) initiated by the U.S. Bureau of the Census in 1967 (see National Research Council, 1980). A basic goal of this program was to provide high-quality, consistent series of county population estimates with comparability from area to area. Forty-nine states (with the exception of Massachusetts) currently participate in this program, and their designated agencies work together with the Census Bureau under this program. In addition to county estimates, several members of the FSCP now produce subcounty estimates as well. The FSCP plays a key role in the Census Bureau's post censual estimation program as the FSCP contacts provide the bureau a variety of data that can be used in making post censual population estimates. Considerable methodological research on small area population estimation is being conducted in the Census Bureau.

Our second example is taken from Fay and Herriot (1979) whose objective was to estimate the per capita income (PCI) for several small places. The U.S. Census Bureau was required to provide the Treasury Department with the PCI estimates and other statistics for state and local governments receiving funds under the General Revenue Sharing Program. These statistics were then used by the Treasury Department to determine allocations to the local governments within the different states by dividing the corresponding state allocations. Initially, the Census Bureau determined the current estimates of PCI by multiplying the 1970 census estimates of PCI in 1969 (based on a 20 percent sample) by ratios of an administrative estimate of PCI in the current year and a similarly derived estimate for 1969. The bureau then confronted the problem that among the approximately 39,000 local government units about 15,000 were for places having fewer than 500 persons in 1970. The sampling errors in the PCI estimates for such small places were large: for a place of 500 persons the coefficient of variation was about 13 percent while it increased to about 30 percent for a place of 100 persons. Consequently, the Bureau initially decided to set aside the census estimates for these small areas and use the corresponding county PCI estimates in their place. This solution proved unsatisfactory, however, in that the census estimates of PCI for a large number of small places differed significantly from the corresponding county estimates, after taking account of the sampling errors. Fay and Herriot (1979) suggest better estimates based on the EB method and present empirical evidence that these have average error smaller than either the census sample estimates or the county averages. The proposed estimate for a small place is a weighted average of the census sample estimate and a "synthetic" estimate obtained by fitting a linear regression equation to the sample estimates of PCI using as independent variables the corresponding county averages, tax-return data for 1969 and data on housing from the 1970 census. The Fay-Herriot method was adopted by the Census Bureau in 1974 to form updated estimates of PCI for small places. Section 4 discusses the Fay-Herriot model and similar models for other purposes, all involving linear regression models with random small area effects.

Our third example refers to the highly debated and controversial issue of adjusting for population undercount in the 1980 U.S. Census. Every tenth year since 1790 a census has been taken to count the U.S. population. The census provides the population count for the whole country as well as for each of the 50 states, 3000 counties and 39,000 civil divisions. These counts are used by the Congress for apportioning funds, amounting to about 100 billion dollars a year during the early 1980s, to the different state and local governments.

It is now widely recognized that complete coverage is impossible. In 1980, vast sums of money and intellectual resources were expended by the U.S. Census Bureau on the reduction of non-coverage. Despite this, there were complaints of undercounts by several major cities and states for their respective areas, and indeed New York State filed a lawsuit against the Census Bureau in 1980 demanding the Bureau to revise its count for that state.

An undercount is the difference between omissions and erroneous inclusions in the census, and it is typically positive. In New York State's law suit against the Census Bureau, E.P. Ericksen and J.B. Kadane, among other statisticians, appeared as the plaintiff's expert witnesses. They proposed using weighted averages of sample estimates and synthetic regression estimates of the 1980 Census undercount, similar to those of Fay and Herriot (1979) for PCI, to arrive at the adjusted population counts of the 50 states and the 16 large cities, including the State of New York and New York City. The sample estimates are obtained from a Post Enumeration Survey. Their general philosophy on the role of adjustment as well as the explicit regression models used for obtaining the regression estimates are documented in Ericksen and Kadane (1985) and Ericksen, Kadane and Tukey (1989). These authors also suggest using the regression equation for areas where no sample data are available. As a historical aside, we may point out here that the regression method for improving local area estimates was first used by Hansen, Hurwitz and Madow (1953, pages 483–486), but its recent popularity owes much to Ericksen (1974).

While the Ericksen-Kadane proposal was applauded by many as the first serious attempt towards adjustment of Census undercount, it has also been vigorously criticized by others (see, e.g., the discussion of Ericksen and Kadane, 1985). In particular, Freedman and Navidi (1986, 1992) criticized them for not validating their model and for not making their assumptions explicit. They also raise several other technical issues, including the effect of large biases and large sampling errors in the sample estimates. Ericksen and Kadane (1987, 1992), Cressie (1989, 1992), Isaki et al. (1987) and others address these difficulties, but clearly further research is needed. Researchers within and outside the U.S. Census Bureau are currently studying various models for census undercount and the properties of the resulting estimators and associated measures of uncertainty using the EBLUP, EB, HB and related approaches.

Our fourth example, taken from Battese, Harter and Fuller (1988), concerns the estimation of areas under corn and soybeans for each of 12 counties in North-Central Iowa using farm-interview data in conjunction with LANDSAT satellite data. Each county was divided into area segments, and the areas under corn and soybeans were ascertained for a sample of segments by interviewing farm operators; the number of sample segments in a county ranged from 1 to 6. Auxiliary data in the form of numbers of pixels (a term used for "picture elements" of about 0.45 hectares) classified as corn and soybeans were also obtained for all the area segments, including the sample segments, in each county using the LANDSAT satellite readings. Battese, Harter and Fuller (1988) employ a "nested error regression" model involving random small area effects and the segment-level data and then obtain the EBLUP estimates of county areas under corn and soybeans using the classical components of variance approach (see Section 5). They also obtain estimates of mean squared error (MSE) of their estimates by taking into account the uncertainty involved in estimating the variance components. Datta and Ghosh (1991) apply the HB approach to these data and show that the two approaches give similar results.

Our final example concerns the estimation of mean wages and salaries of units in a given industry for each census division in a province using gross business income as the only auxiliary variable with known population means (see Särndal and Hidiroglou, 1989). This example will be used in Section 6 to compare and evaluate, under simple random sampling, several competing small area estimators discussed in this paper, treating the census divisions as small areas. We were able to compare the actual errors of the different small area estimators since the true mean wages and salaries for each small area are known.

The outline of the paper is as follows. Section 2 gives a brief account of classical demographic methods for local estimation of population and other characteristics of interest in post-censual years. These methods use current data from administrative registers in conjunction with related data from the latest census. Section 3 provides a discussion of traditional synthetic estimation and related methods under the design-based framework. Two types of small area models that include random area-specific effects are introduced in Section 4. In the first type, only area specific auxiliary data, related to parameters of interest, are available. In the second type of models, element-specific auxiliary data are available for the population elements; and the variable of interest is assumed to be related to these variables through a nested error regression model. We present the EBLUP, EB and HB approaches to small area estimation in Section 5 in the context of basic models given in Section 4. Both point estimation and measurement of uncertainty associated with the estimators are studied. Section 6 compares the performances of several competing small area estimators using sample data drawn from a synthetic population resembling the business population studied by Särndal and Hidiroglou (1989). In Section 7, we focus on special problems that may be encountered in implementing model-based methods for small area estimation. In particular, we give a brief account of model diagnostics for the basic models of Section 4 and of constrained estimation. Various extensions of the basic models are also mentioned in this section. Finally, some concluding remarks are made in Section 8.

The scope of our paper is limited to methods of estimation for small areas; but the development and provision of small area statistics involves many other issues, including those related to sample design and data development, organization and dissemination. Brackstone (1987) gives an excellent account of these issues in the context of Statistics Canada's Small Area Data Program. Singh, Gambino and Mantel (1992) highlight the need for developing an overall strategy that includes planning, designing and estimation stages in the survey process.

2. DEMOGRAPHIC METHODS

As pointed out earlier, demographers have long been using a variety of methods for local estimation of population and other characteristics of interest in post-censual years. Purcell and Kish (1980) categorize these methods under the general heading of Symptomatic Accounting Techniques (SAT). Such techniques utilize current data from administrative registers in conjunction with related data from the latest census. The diverse registration data used in the U.S. include "symptomatic" variables, such as the numbers of births and deaths, of existing and new housing units and of school enrollments whose variations are strongly related to changes in population totals or in its components. The SAT methods studied in the literature include the Vital Rates (VR) method (Bogue, 1950), the composite method (Bogue and Duncan, 1959), the Census Component Method II (CM-II) (U.S. Bureau of the Census, 1966), and the Administrative Records (AR) method (Starsinic, 1974), and the Housing Unit (HU) method (Smith and Lewis, 1980).

The VR method uses only birth and death data, and these are used as symptomatic variables rather than as components of population change. First, in a given year, say t, the annual number of births,

b_t, and deaths, d_t, are determined for a local area. Next the crude birth and death rates, r_{1t} and r_{2t}, for that local area are estimated by

$$r_{1t} = r_{10}(R_{1t}/R_{10}), \qquad r_{2t} = r_{20}(R_{2t}/R_{20}),$$

where r_{10} and r_{20} respectively denote the crude birth and death rates for the local area in the latest census year ($t = 0$) while $R_{1t}(R_{2t})$ and $R_{10}(R_{20})$ respectively denote the crude birth (death) rates in the current and census years for a larger area containing the local area. The population P_t for the local area at year t is then estimated by

$$P_t = \tfrac{1}{2}(b_t/r_{1t} + d_t/r_{2t}).$$

As pointed out by Marker (1983), the success of the VR method depends heavily on the validity of the assumption that the ratios r_{1t}/r_{10} and r_{2t}/r_{20} for the local area are approximately equal to the corresponding ratios, R_{1t}/R_{10} and R_{2t}/R_{20}, for the larger area. Such an assumption is often questionable, however.

The composite method is an extension of the VR method that sums independently computed age-sex-race specific estimates based on births, deaths and school enrollments (see Zidek, 1982, for details).

The CM-II method takes account of net migration unlike the previous methods. Denoting the net migration in the local area during the period since the last census as m_t, an estimate of P_t is given by

$$P_t = P_0 + b_t - d_t + m_t,$$

where P_0 is the population of the local area in the census year $t = 0$. In the U.S., the net migration is further subdivided into military and civilian migration. The former is readily obtainable from administrative records while the CM-II estimates civilian migration from school enrollments. The AR method, on the other hand, estimates the net migration from records for individuals as opposed to collect units like schools (see Zidek, 1982, for details).

The HU method expresses P_t as

$$P_t = (H_t)(PPH_t) + GQ_t,$$

where H_t is the number of occupied housing units at time t, PPH_t is the average number of persons per housing unit at time t and GQ_t is the number of persons in group quarters at time t. The quantities H_t, PPH_t and GQ_t all need to be estimated. Smith and Lewis (1980) report different methods of estimating these quantities.

As pointed out by Marker (1983), most of the estimation methods mentioned above can be identified as special cases of multiple linear regression. Regression-symptomatic procedures also use multiple linear regression for estimating local area populations utilizing symptomatic variables as independent variables in the regression equation. Two such procedures are the ratio-correlation method and the difference-correlation method. Briefly, the former method is as follows: Let 0, 1 and $t(> 1)$ denote two consecutive census years and the current year, respectively. Also, let $P_{i\alpha}$ and $S_{ij\alpha}$ be the population and the value of the jth symptomatic variable for the ith local area ($i = 1, \dots, m$) in the year $\alpha(= 0, 1, t)$. Further, let $p_{i\alpha} = P_{i\alpha}/\Sigma_i P_{i\alpha}$ and $s_{ij\alpha} = S_{ij\alpha}/\Sigma_i S_{ij\alpha}$ be the corresponding proportions, and write $R'_i = p_{i1}/p_{i0}$, $R_i = p_{it}/p_{i1}$, $r'_{ij} = s_{ij1}/s_{ij0}$ and $r_{ij} = s_{ijt}/s_{ij1}$. Using the data $(R'_i, r'_{i1}, \dots, r'_{ip}; i = 1, \dots, m)$ and multiple regression, we first fit

$$R'_i = \hat{\beta}'_0 + \hat{\beta}'_1 r'_{i1} + \dots + \hat{\beta}'_p r'_{ip}, \tag{2.1}$$

where $\hat{\beta}$s are the estimated regression coefficients that link the change, R'_i, in the population proportions between the two census years to the corresponding changes, r'_{ij}, in the proportions for the symptotmatic variables. Next the changes, R_i, in the post censual period are predicted as

$$\widetilde{R}_i = \hat{\beta}'_0 + \hat{\beta}'_1 r_{i1} + \dots + \hat{\beta}'_p r_{ip},$$

using the known changes, r_{ij}, in the symptomatic proportions in the post censual period and the estimated regression coefficients. Finally, the current population counts, P_{it}, are estimated as

$$\widetilde{P}_{it} = \widetilde{R}_i p_{i1} \left(\sum_i P_{it} \right),$$

where the total current count, $\Sigma_i P_{it}$, is ascertained from other sources. In the difference-correlation method, differences between the proportions at the two pairs of time points, $(0, 1)$ and $(1, t)$, are used rather than their ratios.

The regression-symptomatic procedures described above use the regression coefficients, $\hat{\beta}'_j$, in the last intercensual period, but significant changes in the statistical relationship can lead to errors in the current postcensal estimates. The sample-regression method (Ericksen, 1974) avoids this problem by using sample estimates of R_i to establish the current regression equation. Suppose sample estimates of R_i are available for k out of m local areas, say $\widehat{R}_1, \dots, \widehat{R}_k$. Then one fits the regression equation

$$\widehat{R}_i = \hat{\beta}_0 + \hat{\beta}_1 r_{i1} + \dots + \hat{\beta}_p r_{ip}$$

to the data $(\widehat{R}_i, r_{i1}, \ldots, r_{ip})$ from the k sampled areas, instead of (2.1); and then obtains the sample-regression estimators, $\widehat{R}_{i(\text{reg})}$, for all the areas using the known symptomatic ratios r_{ij} $(i = 1, \ldots, m)$:

$$\widehat{R}_{i(\text{reg})} = \hat{\beta}_0 + \hat{\beta}_1 r_{i1}, + \ldots + \hat{\beta}_p r_{ip}.$$

Using 1970 census data and sample data from the Current Population Survey (CPS), Ericksen (1974) has shown that the reduction of mean error is slight compared to the ratio-correlation method but that of large errors (10% or greater) is more substantial. The success of Ericksen's method depends largely on the size and quality of the samples, the dynamics of the regression relationships and the nature of the variables.

3. SYNTHETIC AND RELATED ESTIMATORS

Gonzalez (1973) describes synthetic estimates as follows: "An unbiased estimate is obtained from a sample survey for a large area; when this estimate is used to derive estimates for subareas under the assumption that the small areas have the same characteristics as the large area, we identify these estimates as synthetic estimates." The National Center for Health Statistics (1968) first used synthetic estimation to calculate state estimates of long and short term physical disabilities from the National Health Interview Survey data. This method is traditionally used for small area estimation, mainly because of its simplicity, applicability to general sampling designs and potential of increased accuracy in estimation by borrowing information from similar small areas. We now give a brief account of synthetic estimation and related methods, under the design-based framework.

3.1 Synthetic Estimation

Suppose the population is partitioned into large domains g for which reliable direct estimators, $\widehat{Y}'_{\cdot g}$, of the totals, $Y_{\cdot g}$, can be calculated from the survey data; the small areas, i, may cut across g so that $Y_{\cdot g} = \Sigma_i Y_{ig}$, where Y_{ig} is the total for cell (i,g). We assume that auxiliary information in the form of totals, X_{ig}, is also available. A synthetic estimator of small area total $Y_i = \Sigma_g Y_{ig}$ is then given by

$$\widehat{Y}_i^S = \sum_g (X_{ig}/X_{\cdot g})\widehat{Y}'_{\cdot g}, \tag{3.1}$$

where $X_{\cdot g} = \Sigma_i X_{ig}$ (Purcell and Linacre, 1976; Ghangurde and Singh, 1977). The estimator (3.1) has the desirable consistency property that $\Sigma_i \widehat{Y}_i^S$ equals the reliable direct estimator $\widehat{Y}' = \Sigma_g \widehat{Y}'_{\cdot g}$ of the population total Y, unlike the original estimator proposed by the National Center for Health Statistics (1968) which uses the ratio $X_{ig}/\Sigma_g X_{ig}$ instead of $X_{ig}/X_{\cdot g}$.

The direct estimator $\widehat{Y}'_{\cdot g}$ used in (3.1) is typically a ratio estimator of the form

$$\widehat{Y}'_{\cdot g} = \left[\left(\sum_{\ell \in s_{\cdot g}} w_\ell y_\ell\right) \Big/ \left(\sum_{\ell \in s_{\cdot g}} w_\ell x_\ell\right)\right] X_{\cdot g} = (\widehat{Y}_{\cdot g}/\widehat{X}_{\cdot g}) X_{\cdot g},$$

where $s_{\cdot g}$ denotes the sample in the large domain g and w_ℓ is the sampling weight attached to the ℓth element. For this choice, the synthetic estimator (3.1) reduces to $\widehat{Y}_i^S = \Sigma_i X_{ig}(\widehat{Y}_{\cdot g}/\widehat{X}_{\cdot g})$.

If $\widehat{Y}'_{\cdot g}$ is approximately design-unbiased, the design-bias of $\widehat{Y}_i^S$ is given by

$$E(\widehat{Y}_i^S) - Y_i \doteq \sum_g X_{ig}(Y_{\cdot g}/X_{\cdot g} - Y_{ig}/X_{ig}),$$

which is not zero unless $Y_{ig}/X_{ig} = Y_{\cdot g}/X_{\cdot g}$ for all g. In the special case where the auxiliary information X_{ig} equals the population count N_{ig}, the latter condition is equivalent to assuming that the small area means $\overline{Y}_{ig}$ in each group g equal the overall group mean, $\overline{Y}_{\cdot g}$. Such an assumption is quite strong, and in fact synthetic estimators for some of the areas can be heavily biased in the design-based framework.

It follows from (3.1) that the design-variance of $\widehat{Y}_i^S$ will be small since it depends only on the variances and covariances of the reliable estimators $\widehat{Y}'_{\cdot g}$. The variance of $\widehat{Y}_i^S$ is readily estimated, but it is more difficult to estimate the MSE of $\widehat{Y}_i^S$. Under the assumption $\text{cov}(\widehat{Y}_i, \widehat{Y}_i^S) \doteq 0$, where $\widehat{Y}_i$ is a direct, unbiased estimator of Y_i, an approximately unbiased estimator of MSE is given by

$$\text{mse}(\widehat{Y}_i^S) = (\widehat{Y}_i^S - \widehat{Y}_i)^2 - v(\widehat{Y}_i). \tag{3.2}$$

Here $v(\widehat{Y}_i)$ is a design-unbiased estimator of variance of $\widehat{Y}_i$. The estimators (3.2), however, are very unstable. Consequently, it is customary to average these estimators over i to get a stable estimator of MSE (Gonzalez, 1973), but such a global measure of uncertainty can be misleading. Note that the assumption $\text{cov}(\widehat{Y}_i, \widehat{Y}_i^S) \doteq 0$ may be realistic in practice since $\widehat{Y}_i^S$ is much less variable than $\widehat{Y}_i$.

Nichol (1977) proposes to add the synthetic estimate, $\widehat{Y}_i^S$, as an additional independent variable in the sample-regression method. This method, called the combined synthetic-regression method, showed improvement, in empirical studies, over both the synthetic and sample-regression estimates.

Chambers and Feeney (1977) and Purcell and Kish (1980) propose structure preserving estimation (SPREE) as a generalization of synthetic estimation in the sense it makes a fuller use of reliable direct estimates. SPREE uses the well-known method of iterative proportional fitting of margins in a multi-way table, where the margins are direct estimates.

3.2 Composite Estimation

A natural way to balance the potential bias of a synthetic estimator against the instability of a direct estimator is to take a weighted average of the two estimators. Such composite estimators may be written as

$$(3.3) \qquad \widehat{Y}_i^C = w_i \widehat{Y}_{1i} + (1 - w_i)\widehat{Y}_{2i},$$

where $\widehat{Y}_{1i}$ is a direct estimator, $\widehat{Y}_{2i}$ is an indirect estimator and w_i is a suitably chosen weight ($0 \le w_i \le 1$). For example, the unbiased estimator $\widehat{Y}_i$ may be chosen as $\widehat{Y}_{1i}$, and the synthetic estimator $\widehat{Y}_i^S$ as $\widehat{Y}_{2i}$. Many of the estimators proposed in the literature, both design-based and model-based, have the form (3.3). Section 5 gives such estimators under realistic small area models that account for area-specific effects. In this subsection, we mainly focus on the determination of weights, w_i, in the design-based framework using $\widehat{Y}_{1i} = \widehat{Y}_i$ and $\widehat{Y}_{2i} = \widehat{Y}_i^S$.

Optimal weights, w_i(opt), may be obtained by minimising the MSE of $\widehat{Y}_i^C$ with respect to w_i assuming $\operatorname{cov}(\widehat{Y}_i, \widehat{Y}_i^S) \doteq 0$:

$$(3.4) \quad w_i(\text{opt}) = \text{MSE}\,(\widehat{Y}_i^S)/[\text{MSE}\,(\widehat{Y}_i^S) + V(\widehat{Y}_i)].$$

The optimal weight (3.4) may be estimated by substituting the estimator mse $(\widehat{Y}_i^S)$ given in (3.2) for the numerator and $(\widehat{Y}_i^S - \widehat{Y}_i)^2$ for the denominator, but the resulting weights can be very unstable. Schaible (1978) proposes an "average" weighting scheme based on several variables to overcome this difficulty, noting that the composite estimator is quite robust to deviations from w_i(opt). Another approach (Purcell and Kish, 1979) uses a common weight, w, and then minimizes the average MSE, i.e., $m^{-1}\Sigma_i$ MSE $(\widehat{Y}_i^C)$, with respect to w. This leads to estimated weight of the form

$$(3.5) \quad \hat{w}(\text{opt}) = 1 - \sum_i v(\widehat{Y}_i) \Big/ \sum_i (\widehat{Y}_i^S - \widehat{Y}_i)^2.$$

If the variances of $\widehat{Y}_i$'s are approximately equal, then we can replace $v(\widehat{Y}_i)$ by the average $\bar{v} = \Sigma_i v(\widehat{Y}_i)/m$ in which case (3.5) reduces to James-Stein type weight:

$$\hat{w}(\text{opt}) = 1 - m\bar{v} \Big/ \sum_i (\widehat{Y}_i^S - \widehat{Y}_i)^2.$$

The choice of a common weight, however, is not reasonable if the individual variances, $V(\widehat{Y}_i)$, vary considerably. Also, the James-Stein estimator can be less efficient than the direct estimator, $\widehat{Y}_i$, for some individual areas if the small areas that are pooled are not "similar" (C.R. Rao and Shinozaki, 1978).

Simple weights, w_i, that depend only on the domain counts or the domain totals of a covariate x have also been proposed in the literature. For example, Drew, Singh and Choudhry (1982) propose the sample size dependent estimator which uses the weight

$$(3.6) \qquad w_i(D) = \begin{cases} 1, & \text{if } \hat{N}_i \ge \delta N_i, \\ \hat{N}_i/(\delta N_i), & \text{otherwise,} \end{cases}$$

where $\hat{N}_i$ is the direct, unbiased estimator of the known domain population size N_i and δ is subjectively chosen to control the contribution of the synthetic estimator. This estimator with $\delta = 2/3$ and a generalized regression synthetic estimator replacing the ratio synthetic estimator $\widehat{Y}_i^S$ is currently being used in the Canadian Labour Force Survey to produce domain estimates. Särndal and Hidiroglou (1989) propose an alternative estimator which uses the weight

$$(3.7) \qquad w_i(S) = \begin{cases} 1, & \text{if } \hat{N}_i \ge N_i \\ (\hat{N}_i/N_i)^{h-1}, & \text{otherwise,} \end{cases}$$

where h is subjectively chosen. They, however, suggest $h = 2$ as a general-purpose value. Note that the weights (3.6) and (3.7) are identical if one chooses $\delta = 1$ and $h = 2$.

To study the nature of the weights $w_i(D)$ or $w_i(S)$, let us consider the special case of simple random sampling of n elements from a population of N elements. In this case, $\hat{N}_i = N(n_i/n)$, where the random variable n_i is the sample size in ith domain. Taking $\delta = 1$ in (3.6), it now follows that $w_i(D) = w_i(S) = 1$ if n_i is at least as large as the expected sample size $E(n_i) = n(N_i/N)$, that is, the sample size dependent estimators can fail to borrow strength from related domains even when $E(n_i)$ is not large enough to make the direct estimator $\widehat{Y}_i$ reliable. On the other hand, when $\hat{N}_i < N_i$ the weight $w_i(D)$, which equals $w_i(S)$ when $h = 2$, decreases as n_i decreases. As a result, more weight is given to the synthetic component as n_i decreases. Thus, the weights behave well unlike in the case $\hat{N}_i \geq N_i$. Another disadvantage is that the weights do not take account of the size of between area variation relative to within area variation for the characteristic of interest, that is, all characteristics get the same weight irrespective of their differences with respect to between area homogeneity.

Holt, Smith and Tomberlin (1979) obtain a best linear unbiased prediction (BLUP) estimator of Y_i under the following model for the finite population:

$$y_{ig\ell} = \mu_g + e_{ig\ell}, \quad \ell = 1,\dots,N_{ig}; \quad g = 1,\dots,G; \quad i = 1,\dots,m \tag{3.8}$$

where $y_{ig\ell}$ is the y-value of the ℓth unit in the cell (i,g), μ_g's are fixed effects and the errors $e_{ig\ell}$ are uncorrelated with zero means and variances σ_g^2. Further, N_{ig} denotes the number of population elements in the large domain g that belong to the small area i. Suppose n_{ig} elements in a sample of size n fall in cell (i,g), and let $\bar{y}_{ig}$ and $\bar{y}_{\cdot g}$ denote the sample means for (i,g) and g, respectively.

The best linear unbiased estimator of μ_g under (3.8) is $\hat{\mu}_g = \bar{y}_{\cdot g}$ which in turn leads to the BLUP estimator of Y_i given by

$$\widehat{Y}_i^B = \sum_g \widehat{Y}_{ig}^C,$$

where $\widehat{Y}_{ig}^C$ is a composite estimator of the total Y_{ig} giving the weight $w_{ig} = n_{ig}/N_{ig}$ to the direct estimator $\widehat{Y}_{ig} = N_{ig}\bar{y}_{ig}$, and the weight $1 - w_{ig}$ to the synthetic estimator $\widehat{Y}_{ig}^S = N_{ig}\bar{y}_{ig}$. It therefore follows that the BLUP estimator of Y_i tends to the synthetic estimator $\widehat{Y}_{\cdot g}^S = \Sigma_g N_{ig}\bar{y}_{\cdot g}$ if the sampling fraction n_{ig}/N_{ig} is negligible for all g, irrespective of the size of between area variation relative to within area variation. This limitation of model (3.8) can be avoided by using more realistic models that include random area-specific effects. We consider such models in Section 4, and we obtain small area estimators under these models in Section 5 using a general EB or a variance components approach as well as a HB procedure.

4. SMALL AREA MODELS

We now consider small area models that include random area-specific effects. Two types of models have been proposed in the literature. In the first type, only area-specific auxiliary data $\mathbf{x}_i = (x_{i1},\dots,x_{ip})^T$ are available and the parameters of interest, θ_i, are assumed to be related to $\mathbf{x}_i$. In particular, we assume that

$$\theta_i = \mathbf{x}_i^T\beta + v_i z_i, \qquad i = 1,\dots,m, \tag{4.1}$$

where the z_i's are known positive constants, β is the vector of regression parameters and the v_i's are independent and identically distributed (iid) random variables with

$$E(v_i) = 0, \qquad V(v_i) = \sigma_v^2.$$

In addition, normality of the random effects v_i is often assumed. In the second type of models, element-specific auxiliary data $\mathbf{x}_{ij} = (x_{ij1},\dots,x_{ijp})^T$ are available for the population elements, and the variable of interest, y_{ij}, is assumed to be related to $\mathbf{x}_{ij}$ through a nested error regression model:

$$y_{ij} = \mathbf{x}_{ij}^T\beta + v_i + e_{ij}, \quad j = 1,\dots,N_i; \qquad i = 1,\dots,m. \tag{4.2}$$

Here $e_{ij} = \tilde{e}_{ij}k_{ij}$ and the $\tilde{e}_{ij}$'s are iid random variables, independent of the v_i's, with

$$E(\tilde{e}_{ij}) = 0, \qquad V(\tilde{e}_{ij}) = \sigma^2,$$

the k_{ij}'s being known constants and N_i the number of elements in the ith area. In addition, normality of the v_i's and $\tilde{e}_{ij}$'s is often assumed. The parameters of inferential interest here are the small area totals Y_i or the means $\bar{Y}_i = Y_i/N_i$.

For making inferences about the θ_i's under model (4.1), we assume that direct estimators, $\hat{\theta}_i$, are available and that

$$\hat{\theta}_i = \theta_i + e_i, \qquad i = 1,\dots,m \tag{4.3}$$

where the e_i's are sampling errors, $E(e_i|\theta_i) = 0$ and $V(e_i|\theta_i) = \psi_i$, that is, the estimators $\hat{\theta}_i$ are design-unbiased. It is also customary to assume that the sampling variances, ψ_i, are known. These assumptions may be quite restrictive in some applications. For example, in the case of adjustment for census underenumeration, the estimates $\hat{\theta}_i$ obtained from a post-enumeration survey (PES) could be seriously biased, as noted by Freedman and Navidi (1986). Simlarly, if θ_i is a nonlinear function of the small area total Y_i and the sample size, n_i is small, then $\hat{\theta}_i$ may be seriously biased even if the direct estimator of Y_i is unbiased. We also assume normality of the $\hat{\theta}_i$'s, but this may not be as restrictive as the normality of the random effects v_i, due to the central limit theorem's effect on the $\hat{\theta}_i$'s.

Combining (4.3) and (4.1), we obtain the model

$$\hat{\theta}_i = \mathbf{x}_i^T\beta + v_i z_i + e_i, \qquad i = 1,\dots,m \tag{4.4}$$

which is a special case of the general mixed linear model. Note that (4.4) involves design-induced random variables, e_i, as well as model-based random variables v_i.

Turning to the nested error regression model (4.2), we assume that a sample of size n_i is taken from the ith area and that selection bias is absent; that is, the sample values also obey the assumed model. The latter is satisfied under simple random sampling. It may also be noted that model (4.2) may not be appropriate under more complex sampling designs, such as stratified multistage sampling, since the design features are not incorporated. However, it is possible to extend this model to account for such features (see Section 7).

Writing model (4.2) in matrix form as

$$\mathbf{y}_i^P = \mathbf{X}_i^P\beta + v_i\mathbf{1}_i^P + \mathbf{e}_i^P, \tag{4.5}$$

where $\mathbf{X}_i^P$ is $N_i \times p$, $\mathbf{y}_i^P, \mathbf{e}_i^P$ and $\mathbf{1}_i^P$ are $N_i \times 1$ and $\mathbf{1}_i^P = (1,\dots,1)^T$, we can partition (4.5) as

$$\mathbf{y}_i^P = \begin{bmatrix}\mathbf{y}_i\\ \mathbf{y}_i^*\end{bmatrix} = \begin{bmatrix}\mathbf{X}_i\\ \mathbf{X}_i^*\end{bmatrix}\beta + v_i\begin{bmatrix}\mathbf{1}_i\\ \mathbf{1}_i^*\end{bmatrix} + \begin{bmatrix}\mathbf{e}_i\\ \mathbf{e}_i^*\end{bmatrix}, \tag{4.6}$$

where the superscript $*$ denotes the nonsampled elements. Now, writing the mean $\overline{Y}_i$ as

$$\overline{Y}_i = f_i\bar{y}_i + (1 - f_i)\bar{y}_i^*, \tag{4.7}$$

with $f_i = n_i/N_i$ and $\bar{y}_i$, $\bar{y}_i^*$ denoting the means for sampled and nonsampled elements respectively, we may view estimation of $\overline{Y}_i$ as equivalent to prediction of $\bar{y}_i^*$ given the data $\{\mathbf{y}_i\}$ and $\{\mathbf{X}_i\}$.

Various extensions of models (4.4) and (4.6), as well as models for binary and Poisson data, have been proposed in the literature. Some of these extensions will be briefly discussed in Section 7.

In the examples given in the Introduction, the models considered are special cases of (4.4) or (4.6). In Example 3, Ericksen and Kadane (1985, 1987) use model (4.4) with $z_i = 1$ and assume σ_v^2 to be known. Here $\hat{\theta}_i$ is a PES estimate of census undercount $\theta_i = \{(T_i - C_i)/T_i\}100$, where T_i is the true (unknown) count and C_i is the census count in the ith area. Cressie (1992) uses (4.4) with $z_i = C_i^{-1/2}$, where $\hat{\theta}_i$ is a PES estimate of the adjustment factor $\theta_i = T_i/C_i$. In Example 2, Fay and Herriot (1979) use (4.4) with $z_i = 1$, where $\hat{\theta}_i$ is a direct estimator of $\theta_i = \log P_i$ and P_i is the average percapita income (PCI) in the ith area. Further, $\mathbf{x}_i^T\beta = \beta_0 + \beta_1 x_i$ with x_i denoting the associated county value of log (PCI) from the 1970 census. In Example 4, Battese, Harter and Fuller (1988) use model (4.6) with $k_{ij} = 1$ and $\mathbf{x}_{ij}^T\beta = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}$, where y_{ij}, x_{1ij} and x_{2ij} respectively denote the number of hectares of corn (or soybeans), the number of pixels classified as corn and the number of pixels classified as soybeans in the jth area segment of the ith county. A suitable model for our final example is also a special case of (4.6) with $\mathbf{x}_i^T\beta = \beta_0 + \beta_1 x_{ij}$ and $k_{ij} = x_{ij}^{1/2}$, where y_{ij} and x_{ij} respectively denote the total wages and salaries and gross business income for the jth firm in the ith area (census division).

5. EBLUP, EB AND HB APPROACHES

We now present the EBLUP, EB and HB approaches to small area estimation in the context of models (4.4) and (4.6). Both point estimation and measurement of uncertainty associated with the estimators will be studied.

5.1 EBLUP (Variance Components) Approach

As noted in Section 4, most small area models are special cases of a general mixed linear model involving fixed and random effects, and small area parameters can be expressed as linear combinations of these effects. Henderson (1950) derives BLUP estimators of such parameters in the classical frequentist framework. These estimators minimize the mean squared error among the class of linear unbiased estimators and do not depend on normality, similar to the best linear unbiased estimators (BLUEs) of fixed parameters. Robinson (1991) gives an excellent account of BLUP theory and examples of its application.

Under model (4.4), the BLUP estimator of $\theta_i = \mathbf{x}_i^T\beta + v_i z_i$ simplifies to a weighted average of the direct estimator $\hat{\theta}_i$ and the regression-synthetic estimator $\mathbf{x}_i^T\tilde{\beta}$:

$$\tilde{\theta}_i^H = \gamma_i\hat{\theta}_i + (1 - \gamma_i)\mathbf{x}_i^T\tilde{\beta}, \tag{5.1}$$

where the superscript H stands for Henderson,

$$\tilde{\beta} = \left[\sum_i \mathbf{x}_i\mathbf{x}_i^T/(\sigma_v^2 z_i^2 + \psi_i)\right]^{-1} \cdot \left[\sum_{i=1}^m \mathbf{x}_i\hat{\theta}_i/(\sigma_v^2 z_i^2 + \psi_i)\right] \tag{5.2}$$

is the BLUE estimator of β and

$$\gamma_i = \sigma_v^2 z_i^2/(\sigma_v^2 z_i^2 + \psi_i).$$

The weight γ_i measures the uncertainty in modelling the $\theta_i s$, namely, $\sigma_v^2 z_i^2$ relative to the total variance $\sigma_v^2 z_i^2 + \psi_i$. Thus, the BLUP estimator takes proper account of between area variation relative to the precision of the direct estimator. It is valid for general sampling designs since we are modelling only the θ_i's and not the individual elements in the population. It is also design consistent since $\gamma_i \to 1$ as the sampling variance $\psi_i \to 0$.

The mean squared error (MSE) of $\tilde{\theta}_i^H$ under model (4.4) may be written as

$$M_{1i}(\sigma_v^2) = E(\tilde{\theta}_i^H - \theta_i)^2 = g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2),$$

where

$$g_{1i}(\sigma_v^2) = \sigma_v^2 z_i^2 \psi_i (\sigma_v^2 z_i^2 + \psi_i)^{-1} = \gamma_i \psi_i$$

and

$$g_{2i}(\sigma_v^2) = (1-\gamma_i)^2 \mathbf{x}_i^T \left[\sum_i \mathbf{x}_i \mathbf{x}_i^T / (\sigma_v^2 z_i^2 + \psi_i)\right]^{-1} \mathbf{x}_i.$$

The first term $g_{1i}(\sigma_v^2)$ is of order $O(1)$ while the second term $g_{2i}(\sigma_v^2)$, due to estimating β, is of order $O(m^{-1})$ for large m.

The BLUP estimator (5.1) depends on the variance component σ_v^2 which is unknown in practical applications. However, various methods of estimating variance components in a general mixed linear model are available, including the method of fitting constants or moments, maximum likelihood (ML) and restricted maximum likelihood (REML). Cressie (1992) gives a succinct account of these methods in the context of model (4.4). All these methods yield asymptotically consistent estimators under realistic regularity conditions.

Replacing σ_v^2 with an asymptotically consistent estimator $\hat{\sigma}_v^2$, we obtain a two-stage estimator, $\hat{\theta}_i^H$, which is referred to as the empirical BLUP or EBLUP estimator (Harville, 1991), in analogy with the EB estimator. It remains unbiased provided (i) the distributions of v_i and e_i are both symmetric (not necessarily normal); (ii) $\hat{\sigma}_v^2$ is an even function of $\hat{\theta}_i$'s and remains invariant when $\hat{\theta}_i$ is changed to $\hat{\theta}_i - \mathbf{x}_i^T \mathbf{a}$ for all $\mathbf{a}$ (Kackar and Harville, 1984). Standard methods of estimating variance components all satisfy (ii). We may also point out that the MSE of the EBLUP estimator appears to be insensitive to the choice of the estimator $\hat{\sigma}_v^2$.

If normality of the errors v_i also holds, then we can write the MSE of $\hat{\theta}_i^H$ as

$$M_{2i}(\sigma_v^2) = M_{1i}(\sigma_v^2) + E(\hat{\theta}_i^H - \tilde{\theta}_i^H)^2, \tag{5.3}$$

see Kackar and Harville (1984). It follows from (5.3) that the MSE of $\hat{\theta}_i^H$ is always larger than that of the BLUP estimator $\tilde{\theta}_i^H$. The second term of (5.3) is not tractable, unlike the first term $M_{1i}(\sigma_v^2)$; but it can be approximated for large m (Kackar and Harville, 1984; Prasad and Rao, 1990; Cressie, 1992). We have, for large m,

$$E(\hat{\theta}_i^H - \tilde{\theta}_i^H)^2 \doteq g_{3i}(\sigma_v^2) \tag{5.4}$$

where

$$g_{3i}(\sigma_v^2) = \psi_i^2 z_i^4 (\sigma_v^2 z_i^2 + \psi_i)^{-3} \overline{V}(\hat{\sigma}_v^2),$$

and the neglected terms in the approximation (5.4) are of lower order than $O(m^{-1})$. Here $\overline{V}(\hat{\sigma}_v^2)$ denotes the asymptotic variance of $\hat{\sigma}_v^2$; Cressie (1992) gives the asymptotic variance formulae for ML and REML estimators. It is customary to ignore the uncertainty in $\hat{\sigma}_v^2$ and use $M_{1i}(\hat{\sigma}_v^2) = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2)$ as an estimator of MSE of $\hat{\theta}_i^H$, but this procedure could lead to severe underestimation of the true MSE. A correct, approximately unbiased estimator of MSE $(\hat{\theta}_i^H)$ is given by

$$\operatorname{mse}(\hat{\theta}_i^H) = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{3i}(\hat{\sigma}_v^2), \tag{5.5}$$

(see Prasad and Rao, 1990). The bias of (5.5) is of lower order than m^{-1}.

Noting that $E[\Sigma_i (y_i - \mathbf{x}_i^T \tilde{\beta})^2 / (\sigma_v^2 z_i + \psi_i)] = m - p$, a method of moments estimator $\hat{\sigma}_v^2$ can be obtained by solving iteratively

$$\sum_{i=1}^m (y_i - x_i^T \tilde{\beta})^2 / (\sigma_v^2 z_i + \psi_i) = m - p$$

in conjunction with (5.2) and letting $\hat{\sigma}_v^2 = 0$ when no positive solution exists (Fay and Herriot, 1979). This method does not require normality, unlike the ML and REML. Alternatively, a simple moment estimator is given by $\hat{\sigma}_v^2 = \max(\tilde{\sigma}_v^2, 0)$, where

$$\tilde{\sigma}_v^2 = (t-p)^{-1} \left[\sum_i \frac{1}{z_i^2} (y_i - \mathbf{x}_i^T \beta^*)^2 - \sum_i \frac{\psi_i}{z_i^2} \left\{ 1 - \mathbf{x}_i^T \left(\sum_i \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \mathbf{x}_i \right\} \right] \tag{5.6}$$

and $\beta^* = (\Sigma_i \mathbf{x}_i \mathbf{x}_i^T)^{-1} (\Sigma_i \mathbf{x}_i \hat{\theta}_i)$ is the ordinary least squares estimator of β. The estimator $\tilde{\sigma}_v^2$ is unbiased for σ_v^2 and under normality,

$$\overline{V}(\hat{\sigma}_v^2) = \overline{V}(\tilde{\sigma}_v^2) = 2t^{-2} \sum_i (\sigma_v^2 + \psi_i / z_i^2)^2$$

(see Prasad and Rao, 1990 for the case $z_i = 1$).

Lahiri and Rao (1992) show that the estimator of MSE, (5.5), using the moment estimator (5.6), is also valid under moderate nonnormality of the random effects, v_i. Thus, inference based on $\hat{\theta}_i^H$ and mse($\hat{\theta}_i^H$) is robust to nonnormality of the random effects.

We next turn to the nested error regression model (4.6). The BLUP estimator of $\overline{Y}_i$ in this case is obtained as follows: (i) using the model $\mathbf{y}_i = \mathbf{X}_i + v_i \mathbf{1}_{n_i} + \mathbf{e}_i$ for the sampled elements, obtain the BLUP estimator of $\overline{\mathbf{X}}_i^{*T}\beta + v_i$, where $\overline{\mathbf{X}}_i^*$ is the mean for nonsampled elements; (ii) substitute this estimator for $\bar{y}_i^*$ in (4.7). Thus the BLUP estimator of $\overline{Y}_i$ is given by

$$\tilde{\overline{Y}}_i^H = f_i\bar{y}_i + (1 - f_i)\left[\overline{X}_i^{*T}\tilde{\beta} + \gamma_i(\bar{y}_{iw} - \bar{\mathbf{x}}_{iw}^T\tilde{\beta}\right], \tag{5.7}$$

where $\tilde{\beta}$ is the BLUE of β,

$$\gamma_i = \sigma_v^2(\sigma_v^2 + \sigma^2/w_{i\cdot})^{-1}$$

with $w_{i\cdot} = \Sigma_{j=1}^{n_i} w_{ij}$ and $w_{ij} = k_{ij}^{-2}$, and $\bar{y}_{iw}$ and $\bar{\mathbf{x}}_{iw}$ are the weighted means with weights w_{ij} (see Prasad and Rao, 1990, and Stukel, 1991). The BLUE $\tilde{\beta}$ is readily obtained by applying ordinary least squares to the transformed data $\{(y_{ij} - \gamma_i\bar{y}_{iw})/k_{ij}, (\mathbf{x}_{ij} - \gamma_i\bar{\mathbf{x}}_{iw})/k_{ij}\}$ (see Stukel, 1991, and Fuller and Battese, 1973). If the sample fraction f_i is negligible, we can write $\tilde{\overline{Y}}_i^H$ as a composite estimator of the form

$$\tilde{\overline{Y}}_i^H \doteq \gamma_i[\bar{y}_{iw} + (\overline{\mathbf{X}}_i - \bar{\mathbf{x}}_{iw})^T\tilde{\beta}] + (1 - \gamma_i)\overline{\mathbf{X}}_i^T\tilde{\beta}, \tag{5.8}$$

where $\overline{\mathbf{X}}_i$ is the ith area population mean of $\mathbf{x}_{ij}$'s. It follows from (5.8) that the BLUP estimator is a weighted average of the "survey regression" estimator $\bar{y}_{iw} + (\overline{\mathbf{X}}_i - \bar{\mathbf{x}}_{iw})^T\tilde{\beta}$ and the regression synthetic estimator $\overline{\mathbf{X}}_i^T\tilde{\beta}$. If $k_{ij} = 1$ for all (ij), then the survey regression estimator is approximately design-unbiased for $\overline{Y}_i$ under simple random sampling even if n_i is small. In the case of general k_{ij}'s, it is model-unbiased conditional on the realized local effect v_i, unlike the BLUP estimator which is conditionally biased.

An empirical BLUP estimator, $\hat{\overline{Y}}_i^H$, is obtained from (5.7) by replacing (σ_v^2, σ^2) with asymptotically consistent estimators $(\hat{\sigma}_v^2, \hat{\sigma}^2)$. Further, assuming normality of the errors an approximately unbiased estimator of MSE ($\hat{\overline{Y}}_i^H$), similar to (5.5) under model (4.4), is given by

$$\text{mse}(\hat{\overline{Y}}_i^H) = (1 - f_i)^2\Big[g_{1i}(\hat{\sigma}_v^2, \hat{\sigma}^2) + g_{2i}(\hat{\sigma}_v^2, \hat{\sigma}^2) + 2g_{3i}(\hat{\sigma}_v^2, \hat{\sigma}^2)\Big]. \tag{5.9}$$

Here

$$g_{1i}(\sigma_v^2, \sigma^2) = \gamma_i(\sigma^2/w_{i\cdot}) + (1 - f_i)^2 N_i^{-2}\mathbf{k}_i^{*T}\mathbf{k}_i^*$$

with $\mathbf{k}_i^*$ denoting the vector of k_{ij}'s for nonsampled units in ith area, and

$$g_{2i}(\sigma_v^2, \sigma^2) = (\bar{\mathbf{x}}_i^* - \gamma_i\bar{\mathbf{x}}_{iw})^T\mathbf{A}^{-1}(\bar{\mathbf{x}}_i^* - \gamma_i\bar{\mathbf{x}}_{iw})\sigma^2$$

with

$$\mathbf{A} = \sum_{i=1}^m\left[\sum_{j=1}^{n_i} w_{ij}\mathbf{x}_{ij}\mathbf{x}_{ij}^T - \gamma_i w_{i\cdot}\bar{\mathbf{x}}_{iw}\bar{\mathbf{x}}_{iw}^T\right].$$

Further,

$$g_{3i}(\sigma_v^2, \sigma^2) = w_{i\cdot}^{-2}(\sigma_v^2 + \sigma^2/w_{i\cdot})^{-3}\Big[\sigma^2\overline{V}(\hat{\sigma}_v^2) + \sigma_v^2\overline{V}(\hat{\sigma}_e^2) - 2\sigma^2\sigma_v^2\overline{\text{cov}}(\hat{\sigma}_v^2, \hat{\sigma}_e^2)\Big],$$

where $\overline{\text{cov}}$ denotes the asymptotic covariance (see Stukel, 1991 and Prasad and Rao, 1990).

For the ML and REML methods, the asymptotic covariance matrix of $(\hat{\sigma}_v^2, \hat{\sigma}^2)$ can be obtained from general theory (see, e.g., Cressie, 1992). Stukel (1991) and Fuller and Battese (1973) use the method of fitting constants which involves two ordinary least square fittings: first, we calculate the residual sum of squares, SSE(1), with ν_1 degrees of freedom by regressing through the origin the y-deviations $k_{ij}^{-1}(y_{ij} - \bar{y}_{iw})$ on the nonzero x-deviations $k_{ij}^{-1}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_{iw})$ for these areas with $n_i > 1$. Second, we calculate the residual sum of squares SSE(2) by regressing y_{ij}/k_{ij} on $\mathbf{x}_{ij}/k_{ij}$. Then $\hat{\sigma}^2 = \nu_1^{-1}$ SSE(1) and $\hat{\sigma}_v^2 = \max(\tilde{\sigma}_v^2, 0)$ with

$$\tilde{\sigma}_v^2 = \eta_{**}^{-1}[SSE(2) - (n - p)\hat{\sigma}^2],$$

where

$$\eta^* = \sum_i w_{i\cdot}(1 - w_{i\cdot}\bar{\mathbf{x}}_{iw}^T\mathbf{A}_1^{-1}\bar{\mathbf{x}}_{iw})$$

with

$$\mathbf{A}_1 = \sum_i\sum_j w_{ij}\mathbf{x}_{ij}\mathbf{x}_{ij}^T.$$

The Appendix gives the variances and covariance of $\hat{\sigma}^2$ and $\hat{\sigma}_v^2$.

Again, ignoring the uncertainty in $\hat{\sigma}_v^2$ and $\hat{\sigma}^2$ and using $M_{1i}(\hat{\sigma}_v^2, \hat{\sigma}^2) = g_{1i}(\hat{\sigma}_v^2, \hat{\sigma}^2)$ as an estimator of MSE ($\hat{\overline{Y}}_i^H$) could lead to severe underestimation of the true MSE.

Limited simulation results (Prasad and Rao, 1990; Datta and Ghosh, 1991 and Hulting and

Harville, 1991) indicate that the estimator of MSE, mse $(\hat{\bar{Y}}_i^H)$, given by (5.9), performs well even for moderate m (as small as 15), provided σ_v^2/σ^2 is not close to zero.

5.2 EB Approach

In the EB approach, the posterior distribution of the parameters of interest given the data is first obtained, assuming that the model parameters are known. The model parameters are estimated from the marginal distribution of the data, and inferences are then based on the estimated posterior distribution. Morris (1983) gives an excellent account of the EB approach and significant applications.

Under model (4.4) with normal errors, the posterior distribution of θ_i given $\hat{\theta}_i, \beta$ and σ_v^2 is normal with mean θ_i^B and variance $g_{1i}(\sigma_v^2) = \gamma_i\psi_i$, where

$$\theta_i^B = E(\theta_i|\hat{\theta}_i, \beta, \sigma_v^2) = \gamma_i\hat{\theta}_i + (1-\gamma_i)\mathbf{x}_i^T\beta.$$

Under quadratic loss, θ_i^B is the Bayes estimator of θ_i. Noting that the $\hat{\theta}_i \sim N(\mathbf{x}_i^T\beta,\ \sigma_v^2 z_i^2 + \psi_i)$ are marginally independent, we can obtain the estimators $\hat{\sigma}_v^2$ and $\hat{\beta}$ as before using ML, REML or the method of moments. The estimated posterior distribution is $N(\hat{\theta}_i^{EB}, g_{1i}(\hat{\sigma}_v^2))$, where $\hat{\theta}_i^{EB}$ is identical to the EBLUP estimator $\hat{\theta}_i^H$. A naive EB approach uses $\hat{\theta}_i^{EB}$ as the estimator of θ_i and measures its uncertainty by the estimated posterior variance

$$V(\theta_i|\hat{\theta}_i, \hat{\beta}, \hat{\sigma}_v^2) = g_{1i}(\hat{\sigma}_v^2). \tag{5.10}$$

This can lead to severe underestimation of the true posterior variance $V(\theta_i|\hat{\boldsymbol{\theta}})$ (under a prior distribution on β and σ_v^2), although $\hat{\theta}_i^{EB} = E(\theta_i|\hat{\theta}_i, \hat{\beta}, \hat{\sigma}_v^2)$ is approximately equal to the true posterior mean $E(\theta_i|\hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)^T$.

The above point is better understood when one writes

$$E(\theta_i|\hat{\boldsymbol{\theta}}) = E_{\beta,\sigma_v^2}[E(\theta_i|\hat{\theta}_i, \beta, \sigma_v^2)]$$

and

$$\begin{aligned} V(\theta_i|\hat{\boldsymbol{\theta}}) &= E_{\beta,\sigma_v^2}[V(\theta_i|\hat{\theta}_i, \beta, \sigma_v^2)] \\ &\quad + V_{\beta,\sigma_v^2}[E(\theta_i|\hat{\theta}_i, \beta, \sigma_v^2)], \end{aligned} \tag{5.11}$$

where E_{β,σ_v^2} and V_{β,σ_v^2} respectively denote the expectation and variance with respect to the posterior distribution of β and σ_v^2 given the data $\hat{\boldsymbol{\theta}}$. It follows from (5.11) that (5.10) is a good approximation only to the first variance term on the right side of (5.11), but the second variance term is ignored in the naive EB approach, that is, it fails to take account of the uncertainty about the parameters β and σ_v^2. Note that the form of the prior distribution on β and σ^2 is not specified in the EB approach, unlike in the HB approach (Section 5.3).

Two methods of accounting for the underestimation of true posterior variance have been proposed in the literature. The first method is based on the bootstrap (Laird and Louis, 1987), while the second method uses an asymptotic approximation to the posterior variance $V(\theta_i|\hat{\boldsymbol{\theta}})$ irrespective of the form of the prior on β and σ_v^2 (Kass and Steffey, 1989). In the bootstrap method, a large number, B, of independent bootstrap samples $\{\theta_1^*(b), \ldots, \theta_m^*(b);\ b = 1, \ldots, B\}$ are first drawn, where $\theta_i^*(b)$ is drawn from the estimated marginal distribution $N(\mathbf{x}_i^T\hat{\beta}, \hat{\sigma}_v^2 z_i^2 + \psi_i)$. Estimates $\beta^*(b)$ and $\sigma_v^{*2}(b)$ are then computed from the bootstrap data $\{\theta_i^*(b), \mathbf{x}_i,\ i = 1, \ldots, m\}$ for each b. The EB bootstrap estimator of θ_i is given by

$$\begin{aligned} \theta_i^*(\cdot) &= \frac{1}{B}\sum_{b=1}^{B} E[\theta_i|\theta^*(b), \beta^*(b), \sigma_v^{*2}(b)] \\ &= \frac{1}{B}\sum_{b=1}^{B} \theta_i^{*EB}(b), \end{aligned}$$

and its uncertainty is measured by

$$\begin{aligned} V_i^* &= \frac{1}{B}\sum_{b=1}^{B} V[\theta_i|\theta_i^*(b), \beta^*(b), \sigma_v^{*2}(b)] \\ &\quad + \frac{1}{B-1}\sum_{b=1}^{B}[\theta_i^{*EB}(b) - \theta_i^{*EB}(\cdot)]^2. \end{aligned} \tag{5.12}$$

The second term on the right side of (5.12) accounts for the underestimation. The EB bootstrap method looks promising, but further studies on its frequentist performance are needed.

In the Kass-Steffey method, $\hat{\theta}_i^{EB}$ is taken as the estimator of θ_i, but a positive correction term is added to the estimated posterior variance $V(\theta_i|\hat{\theta}_i, \hat{\beta}, \hat{\sigma}_v^2)$ to account for the underestimation. This term depends on the observed information matrix and the partial derivatives of θ_i^B, evaluated at the ML estimates $\hat{\beta}$ and $\hat{\sigma}_v^2$. This method also looks promising, but its frequentist properties remain to be investigated. (Steffey and Kass, 1991 conjecture that the MSE of EB estimator is approximately equal to their approximation to the posterior variance.) Kass and Steffey (1989) also give an improved second-order approximation to the true posterior variance, $V(\theta_i|\hat{\boldsymbol{\theta}})$.

Turning to the nested error regression model (4.6), the estimated posterior distribution of $\bar{Y}_i$ given the data $\mathbf{y}$ is normal with mean equal to the EBLUP $\hat{\bar{Y}}_i^H$ and variance equal to $(1-f_i)^2 g_{1i}(\hat{\sigma}_v^2, \hat{\sigma}^2)$ which is a severe underestimate of the true posterior variance $V(\overline{Y}_i|\mathbf{y})$. Again, the bootstrap and Kass-Steffey methods can be applied to account for the underestimation.

If one wishes to view the EB approach in the frequentist framework, a prior distribution on β and σ_v^2 cannot be entertained. In this case, MSE is a natural measure of uncertainty and any differences between the EB and EBLUP approaches disappear under the normality assumption. It may also be noted that the EB estimator can be justified without the normality assumption, similar to the EBLUP, using the "posterior linearity" property (Ghosh and Lahiri, 1987; Ericson, 1969).

5.3 HB Approach

In the HB approach, a prior distribution on the model parameters is specified and the posterior distribution of the parameters of interest is then obtained. Inferences are based on the posterior distribution; in particular, a parameter of interest is estimated by its posterior mean and its precision is measured by its posterior variance. The HB approach is straightforward and clear–cut but computationally intensive, often involving high dimensional integration. Recent advances in computational aspects of the HB approach, such as Gibbs sampling (cf. Gelfand and Smith, 1990) and importance sampling, however, seem to overcome the computational difficulties to a large extent. If the solution involves only one or two dimensional integration, it is often easier to perform direct numerical integration than to use Gibbs sampling or any other Monte Carlo numerical integration method. Datta and Ghosh (1991) apply the HB approach to estimation of small area means, $\overline{Y}_i$, under general mixed linear models, and also discuss the computational aspects.

We now illustrate the HB approach under our models (4.4) and (4.6), assuming noninformative priors on β and the variance components σ_v^2 and σ^2. The HB approach, however, can incorporate prior information on these parameters through informative priors.

Under model (4.4), we first obtain the posterior distribution of θ_i given $\hat{\boldsymbol{\theta}}$ and σ_v^2, by assuming that β has a uniform distribution over R^p to reflect absence of prior information on β. Straightforward calculations show that it is normal with mean equal to the BLUP estimator $\tilde{\theta}_i^H$ and variance equal to $M_{1i}(\sigma_v^2)$, the MSE of $\tilde{\theta}_i^H$, that is, $E(\theta_i|\hat{\boldsymbol{\theta}}, \sigma_v^2) = \tilde{\theta}_i^H$ and $V(\theta_i|\hat{\boldsymbol{\theta}}, \sigma_v^2) = \text{MSE}\ (\tilde{\theta}_i^H)$. Hence, when σ_v^2 is assumed to be known, the HB and BLUP approaches lead to identical inferences.

To take account of the uncertainty about σ_v^2, we need to calculate the posterior distribution of σ_v^2 given $\hat{\boldsymbol{\theta}}$ under a suitable prior on σ_v^2. The posterior mean and variance of θ_i are then given by

$$E(\theta_i|\hat{\boldsymbol{\theta}}) \equiv E_{\sigma_v^2}(\tilde{\theta}_i^H) \tag{5.13}$$

and

$$V(\theta_i|\hat{\boldsymbol{\theta}}) = E_{\sigma_v^2}[M_{1i}(\sigma_v^2)] + V_{\sigma_v^2}(\tilde{\theta}_i^H), \tag{5.14}$$

where $E_{\sigma_v^2}$ and $V_{\sigma_v^2}$ respectively denote the expectation and variance with respect to the posterior distribution of σ_v^2 given $\hat{\boldsymbol{\theta}}$. Numerical evaluation of (5.13) and (5.14) involves one dimensional integration. Ghosh (1992) obtains the posterior distribution, $f(\sigma_v^2|\hat{\boldsymbol{\theta}})$, assuming that σ_v^2 has a uniform distribution over $(0, \infty)$ to reflect the absence of prior information about σ_v^2, and that σ_v^2 and β are independently distributed. It is given by

$$f(\sigma_v^2|\hat{\boldsymbol{\theta}}) = (\sigma_v^2)^{-\frac{(m-p)}{2}} \left\{\prod_1^m \gamma_i^{1/2}\right\} \left|\sum_i \gamma_i \mathbf{x}_i \mathbf{x}_i^T\right|^{-\frac{1}{2}} \cdot \exp\left[-\frac{1}{2}Q_a(\hat{\boldsymbol{\theta}})\right],$$

where

$$Q_a(\hat{\boldsymbol{\theta}}) = (\sigma_v^2)^{-1}\left[\sum_i \gamma_i \hat{\theta}_i^2 - \left(\sum_i \gamma_i \hat{\theta}_i \mathbf{x}_i\right)^T \cdot \left(\sum_i \gamma_i \mathbf{x}_i \mathbf{x}_i^T\right)^{-1} \left(\sum_i \gamma_i \hat{\theta}_i \mathbf{x}_i\right)\right].$$

We next turn to the nested error regression model (4.6). We first obtain the posterior distribution of $\overline{Y}_i$ given $\mathbf{y}$, σ_v^2 and σ^2, by assuming that β has uniform distribution over R^p. Straightforward calculations show that it is normal with mean equal to the BLUP estimator $\tilde{\overline{Y}}_i^H$ and variance equal to MSE $(\tilde{\overline{Y}}_i^H) = M_{1i}(\sigma_v^2, \sigma^2)$, that is, $E(\overline{Y}_i|\mathbf{y}, \sigma_v^2, \sigma^2) = \tilde{\overline{Y}}_i^H$ and $V(\overline{Y}_i|\mathbf{y}, \sigma_v^2, \sigma^2) = \text{MSE}(\tilde{\overline{Y}}_i^H)$. Hence, when both σ_v^2 and σ^2 are assumed to be known, the HB and BLUP approaches lead to identical inferences.

To take account of the uncertainty about σ_v^2 and σ^2, Datta and Ghosh (1991) further assume $\beta, (\sigma^2)^{-1}$ and $(\sigma_v^2)^{-1} = (\sigma^2)^{-1}\lambda$ to be independently distributed with $(\sigma^2)^{-1} \sim$ gamma $((1/2)a_0, (1/2)g_0)$ and $(\sigma^2)^{-1}\lambda \sim$ gamma $((1/2)a_1, (1/2)g_1)$, where $a_0 \geq 0$, $g_0 \geq 0$, $a_1 > 0$, $g_1 \geq 0$ and $\lambda = \sigma^2/\sigma_v^2$. Here gamma (α, β) denotes the gamma random variable with pdf $f(z) = \exp(-\alpha z)\alpha^\beta z^{\beta-1}/\Gamma(\beta)$, $z > 0$. Datta and Ghosh (1991) obtain closed form expressions for $E(\overline{Y}_i|\mathbf{y}, \lambda)$

and $V(\overline{Y}_i|\mathbf{y},\lambda)$ by showing that $f(\mathbf{y}^*|\mathbf{y},\lambda)$ is a multivariate t-distribution. They also derive the posterior distribution of λ given $\mathbf{y}$, but it has a complex structure making it necessary to perform one-dimensional numerical integration to get $E(\overline{Y}_i|\mathbf{y})$ and $V(\overline{Y}_i|\mathbf{y})$ using the following relationships:

$$E(\overline{Y}_i|\mathbf{y}) = E_\lambda[E(\overline{Y}_i|\mathbf{y},\lambda)]$$

and

$$V(\overline{Y}_i|\mathbf{y}) = E_\lambda[V(\overline{Y}_i|\mathbf{y},\lambda)] + V_\lambda[E(\overline{Y}_i|\mathbf{y},\lambda)],$$

where E_λ and V_λ respectively denote the expectation and variance under the posterior distribution of λ given the data $\mathbf{y}$.

Datta and Ghosh (1991) compare the HB, EB and EBLUP approaches using the data for our example 4 and letting $a_0 = a_1 = 0.005$ and $g_0 = g_1 = 0$ to reflect the absence of prior information on σ_v^2 and σ^2. As one might expect, the three estimates were close to each other as point predictors of small area (county) means; the EB estimate was obtained by replacing λ with the method-of-fitting constants estimate $\hat{\lambda}$ in $E(\overline{Y}_i|\mathbf{y},\lambda)$. The naive variance estimate, $V(\overline{Y}_i|\mathbf{y},\hat{\lambda}) = (s_i^{EB})^2$ associated with the EB estimate $E(\overline{Y}_i|\mathbf{y},\hat{\lambda})$, was always found to be smaller than the true posterior variance, $V(\overline{Y}_i|\mathbf{y}) = (s_i^{HB})^2$, associated with the HB estimate $\hat{\overline{Y}}_i^{HB} = E(\overline{Y}_i|\mathbf{y})$; for one county, s_i^{EB} was about 10% smaller than s_i^{HB}. Note that the customary naive EB variance estimate, $V(\overline{Y}_i|\mathbf{y},\hat{\beta},\hat{\sigma}_v^2,\hat{\sigma}^2)$, will lead to much more severe underestimation than $V(\overline{Y}_i|\mathbf{y},\hat{\lambda})$ since the latter takes account of the uncertainty about β and σ^2. The estimated MSE, mse $(\hat{\overline{Y}}_i^H) = (s_i^H)^2$, associated with the EBLUP estimate, $\hat{\overline{Y}}_i^H$, was found to be similar to the HB variance estimate. Our example in Section 6 also gives similar results. Datta and Ghosh (1991) have also conducted a simulation study on the frequentist properties of the HB and EBLUP methods using the Battese, Harter and Fuller (1988) model. Their findings indicate that the simulated MSEs for the HB estimator are very close to those for the EBLUP estimator while the coverage probabilities based on $\hat{\overline{Y}}_i^{HB} \pm (1.96)s_i^{HB}$ turn out to be slightly bigger than those based on $\hat{\overline{Y}}_i^H \pm (1.96)s^H$, both being close to nominal confidence level of 95%. Hulting and Harville (1991) obtain similar results in another simulation study using the Battese, Harter and Fuller (1988) model and varying the variance ratio σ_v^2/σ^2. However, they find the HB method produces different and more sensible answers than the EBLUP procedure if the estimate for σ_v^2/σ^2 is zero or close to zero.

The HB approach looks promising, but we need to study its robustness to choice of prior distributions on the model parameters.

6. EXAMPLE

Several of the proposed small area estimators are now compared on the basis of their squared errors and relative errors from the true small area means $\overline{Y}_i$. For this purpose, we first constructed a synthetic population of pairs (y_{ij}, x_{ij}) resembling the business population studied by Särndal and Hidiroglou (1989) where the census divisions are small areas, y_{ij} denotes wages and salaries of jth firm in the ith census division and x_{ij} the corresponding gross business income. To generate the synthetic population, we fitted the nested error regression model (4.6) with $\mathbf{x}_{ij}^T\beta = \beta_0 + \beta_1 x_{ij}$ and $k_{ij} = x_{ij}^{1/2}$ to a real population to estimate β_0 and β_1 and the variance components σ_v^2 and σ^2. The resulting synthetic model is given by

$$\begin{aligned} y_{ij} &= -2.47 + 0.20x_{ij} + v_i + e_{ij}, \\ j &= 1,\ldots,N_i,\ i = 1,\ldots,m, \\ v_i &\overset{\text{i.i.d}}{\sim} N(0, 22.14), \\ e_{ij} &\overset{\text{i.i.d}}{\sim} N(0, 0.47x_{ij}). \end{aligned} \tag{6.1}$$

We then used model (6.1) in conjunction with the population x_{ij}-values to generate a synthetic population of pairs (y_{ij}, x_{ij}) with $m = 16$ small areas. Table 1 reports the small area population sizes, N_i, and the small area means $(\overline{Y}_i, \overline{X}_i)$ for this synthetic population of size $N = 114$. A simple random sample of size $n = 38$ was drawn from the synthetic population. The resulting small area sample sizes, n_i, and sample data (y_{ij}, x_{ij}) are reported in Table 2. Note that direct estimators cannot be implemented for areas 1, 4 and 13 since $n_i = 0$ for these areas. We have, therefore, confined ourselves to the following indirect estimators valid for all $n_i \geq 0$:

(i) Ratio-synthetic estimator: $\hat{\overline{Y}}_i^{RS} = (\bar{y}/\bar{x})\overline{X}_i$, where $(\bar{y}, \bar{x})$ are the overall sample means.

(ii) Sample-size dependent estimator:

$$\hat{\overline{Y}}_i^{SD} = \begin{cases} \hat{\overline{Y}}_i^{\mathrm{REG}} = \bar{y}_i + (\bar{y}/\bar{x})(\overline{X}_i - \bar{x}_i), & \text{if } w_i \geq W_i, \\ \frac{w_i}{W_i}(\hat{\overline{Y}}_i^{\mathrm{REG}}) + \left(1 - \frac{w_i}{W_i}\right)\hat{\overline{Y}}_i^{RS}, & \text{if } w_i < W_i, \end{cases}$$

where $\hat{\overline{Y}}_i^{\mathrm{REG}}$ is a "survey regression" estimator, $(\bar{y}_i, \bar{x}_i)$ are the sample means, $w_i = n_i/n$ and $W_i = N_i/N$. This estimator corresponds to the weight (3.6) with $\delta = 1$ or the weight

TABLE 1
Small area sizes, N_i, and means $(\bar{Y}_i, \bar{X}_i)$ for a synthetic population ($N = 114$)

Area No.	N_i	$\bar{X}_i$	$\bar{Y}_i$	Area No.	N_i	$\bar{X}_i$	$\bar{Y}_i$
1	1	137.70	24.22	9	27	97.58	15.56
2	6	100.84	20.43	10	5	76.04	5.88
3	4	47.72	5.48	11	12	90.15	15.20
4	1	45.64	6.55	12	7	86.24	13.40
5	8	108.53	20.55	13	4	164.28	26.06
6	6	65.68	14.85	14	6	164.70	22.44
7	6	116.34	21.46	15	13	83.86	9.40
8	6	92.74	13.40	16	2	134.49	29.49

TABLE 2
Data from a simple random sample drawn from a synthetic population ($n = 38, N = 114$)

Area No.	n_i	x_{ij}	y_{ij}	Area No.	n_i	x_{ij}	x_{ij}
1	0	—	—	9	10	333.24	47.62
						80.91	5.27
2	3	33,70	5.90			43.65	6.97
		47.19	13.22			29.29	−0.19
		75.21	17.44			102.66	15.94
						109.34	19.84
3	1	36.43	2.54			30.56	2.57
						127.96	24.61
4	0	—	—			190.34	35.41
						52.16	2.54
5	1	28.82	3.61	10	1	45.91	−6.34
				11	2	43.03	8.83
6	2	30.60	11.48			190.12	27.31
		129.69	21.45	12	1	47.39	1.70
				13	0	—	—
7	4	200.60	46.96	14	3	35.66	−0.80
		113.92	15.57			40.23	2.75
		74.33	8.66			111.23	10.87
		53.00	11.90	15	6	51.61	−3.20
						67.46	12.47
8	3	95.43	11.76			190.97	21.77
		35.75	−0.69			35.11	2.92
		39.08	21.46			25.09	−5.46
						73.51	7.35
				16	1	229.32	53.83

(3.7) with $h = 2$. We have not included the optimal composite estimator due to difficulties in estimating the optimal weight (3.4).

(iii) EBLUP (or EB) estimator $\hat{\bar{Y}}_i^H$ under model (4.6) with $\mathbf{x}_{ij}^T\beta = \beta_0+\beta_1 x_{ij}$ and $k_{ij} = x_{ij}^{1/2}$, where σ_v^2 and σ^2 are estimated by the method of fitting constants.

(iv) HB estimator $\hat{\bar{Y}}_i^{HB}$ under model (4.6) as in (iii), using Datta-Ghosh's diffuse priors with $a_0 = 0, g_0 = 0, a_1 = 0.05$ and $g_1 = 0$.

Using the sample data (y_{ij}, x_{ij}) and the known small area population means $\bar{X}_i$ we computed the above four estimates along with their average relative errors

$$\mathrm{ARE} = \frac{1}{m}\sum_{i=1}^{m} |\mathrm{est.} - \bar{Y}_i|/\bar{Y}_i$$

and average squared errors

$$\mathrm{ASE} = \frac{1}{m}\sum_{i=1}^{m} (\mathrm{est.} - \bar{Y}_i)^2.$$

These values are reported in Table 3. We also calculated the standard error, s_i^H, of EBLUP estimator using (5.9) and the posterior standard deviation

TABLE 3

Small area estimates and their (%) average relative errors and average squared roots; standard error (S.E.) of EBLUP and HB estimators

Area No.	n_i	$\bar{Y}_i$	RS	SD	EBLUP	HB	S.E. EBLUP	S.E. HB
1	0	24.22	19.79	19.79	22.16	22.16	7.40	8.29
2	3	20.43	14.90	19.20	20.47	20.18	2.20	2.47
3	1	5.48	6.86	5.34	4.85	4.87	2.62	2.60
4	0	6.55	6.56	6.56	4.97	4.94	5.40	5.99
5	1	20.55	15.60	15.52	17.98	17.81	3.10	3.17
6	2	14.85	9.44	14.39	13.99	13.47	2.07	2.40
7	4	21.46	16.72	21.62	21.31	21.22	1.59	1.74
8	3	13.40	13.33	11.22	11.44	11.58	1.86	2.00
9	10	15.56	14.02	14.27	13.95	13.98	1.14	1.22
10	1	5.88	10.93	6.27	3.30	3.96	3.06	3.63
11	2	15.20	12.96	13.29	14.66	14.44	2.61	2.57
12	1	13.40	12.11	11.17	9.97	10.17	3.14	3.14
13	0	26.06	23.61	23.61	27.13	27.13	5.52	6.13
14	3	22.44	23.67	18.98	24.05	24.22	3.10	3.48
15	6	9.40	12.05	7.40	8.24	8.43	1.32	1.50
16	1	29.49	19.33	40.20	30.31	30.24	2.58	2.87
Av. Rel. Error%:			17.85	12.40	11.74	11.23		
Av. Sq. Error:			22.10	12.38	2.84	2.69		

RS=ratio synthetic estimator; SD=sample-size dependent estimator; EBLUP=EBLUP or EB estimator; HB=HB estimator.

(standard error), s_i^{HB}, of HB estimator using one-dimensional numerical integration. These values are also reported in Table 3.

The following observations on the relative performances of small area estimates may be drawn from Table 3: (1) EBLUP and HB estimators give similar values over small areas, and their average relative errors (%) are 11.74 and 11.23 and squared errors are 2.84 and 2.69 respectively. Asymptotically (as $m \to \infty$), the two estimators are identical, and the observed differences are due to moderate $m(= 16)$ and the method of estimating σ_v^2 and σ^2 (REML or ML would give slightly different EBLUP values). (2) Standard error values for EBLUP and HB estimators are also similar. This is in agreement with the empirical results of Datta and Ghosh (1991) and Hulting and Harville (1991). (3) Under the criterion of average squared error, EBLUP and HB estimators perform much better than the ratio-synthetic and sample-size dependent estimators: 2.84 for EBLUP vs. 12.38 for sample-size dependent (SD) and 22.10 for ratio-synthetic (RS). (4) Under the criterion of average relative error (%), however, EBLUP and HB estimates are not much better than the sample-size dependent estimator: 11.74 for EBLUP versus 12.40 for SD. However, both perform much better than the ratio-synthetic estimator with % ARE = 17.85.

It may be noted that EBLUP, EB and HB estimators are optimal under squared error loss and cease to be so under relative error loss. This is due to the fact that the Bayes estimators under relative error loss can often differ quite significantly from those under squared error loss. This nonoptimality carries over to EBLUP estimator which usually mimics closely the Bayes estimators. The above observations could perhaps explain why in our example the Bayes and EBLUP estimator did not improve significantly over the SD estimator under relative error.

All in all, our results in Table 3 clearly demonstrate the advantages of using the EBLUP or HB estimator and associated standard error when the assumed random effects model fits the data well. (Note that we simulated the data from an assumed model.) It is important, therefore, to examine the aptness of the assumed model using suitable diagnostic tools; Section 7.1 gives a brief account of diagnostics for models (4.4) and (4.6).

7. SPECIAL PROBLEMS

In this section we focus on special problems that may be encountered in implementing model-based methods for small area estimation. We also discuss some extensions of our basic models (4.4) and (4.6).

7.1 Model Diagnostics

Model-based methods rely on careful checking of the assumed models in order to find suitable models that fit the data well. Model diagnostics, therefore, play an important role. However, the literature on diagnostics for mixed linear models involving random effects is not extensive, unlike standard regression diagnostics. Only recently have some useful diagnostic tools been proposed. See, for example, Battese, Harter and Fuller (1988); Beckman, Nachtsheim and Cook (1987); Calvin and Sedransk (1991); Christensen, Pearson and Johnson (1992); Cressie (1992); Dempster and Ryan (1985) and Lange and Ryan (1989).

We first consider the Fay-Herriot type model (4.4), where only area-specific covariates are used. When the model is correct, the standardized residuals $r_i = (\hat{\sigma}_v^2 z_i^2 + \psi_i)^{-(1/2)} (\hat{\theta}_i - \mathbf{x}_i^T \hat{\beta})$, $i = 1, \ldots, m$ are approximately iid $N(0,1)$ for large m where $\hat{\beta}$ is the BLUE estimator (5.2) with σ_v^2 replaced by $\hat{\sigma}_v^2$. We can, therefore, use a $q-q$ plot of r_i against $\Phi^{-1}[F_m(r_i)]$, where $\Phi(r)$ and $F_m(r)$ are the standard normal and empirical cdfs, respectively. A primary goal of this plot is to check the normality of the random effects v_i since the sampling errors e_i are approximately normal due to the central limit theorem effect. Dempster and Ryan (1985) note that the above $q-q$ plot may be inefficient for this purpose since it gives equal weight to each observation, even though the $\hat{\theta}_i$s differ in the amount of information contained about the v_is. They propose a weighted $q-q$ plot which uses a weighted empirical cdf $F_m^*(r) = \Sigma_i I(r - r_i) W_i / \Sigma_i W_i$ in place of $F_m(r)$, where $I(t) = 1$ for $t \geq 0$ and 0 otherwise, and $W_i = (\hat{\sigma}_v^2 + z_i^{-2} \psi_i)^{-1}$ in our case. This plot is more sensitive to departures from normality than the unweighed plot since it assigns greater weight to those observations for which $\hat{\sigma}_v^2$ account for a larger part of the total variance $\hat{\sigma}_v^2 + z_i^{-2}\psi_i$.

We next turn to the nested errors regression model (4.6), where the y_{ij}'s are correlated for each i. In this case, the transformed residuals $r_{ij} = k_{ij}^{-1} (y_{ij} - \hat{\gamma}_i \bar{y}_{iw}) - k_{ij}^{-1} (\mathbf{x}_{ij} - \hat{\gamma}_i \bar{\mathbf{x}}_{iw})^T \hat{\beta}$ are approximately uncorrelated with equal variances σ^2. Therefore, traditional regression diagnostics may be applied to the r_{ij}s, but the transformation can mask the effect of individual errors e_{ij}. On the other hand, standardized BLUP residuals $k_{ij}^{-1}(y_{ij} - \mathbf{x}_{ij}^T \hat{\beta} - \hat{v}_i)/\hat{\sigma}$ may be used to study the effect of individual units (ij) on the model, provided they are not strongly correlated. Lange and Ryan (1989) propose methods for checking the normality assumption on the random effects v_i using the BLUP estimates $\hat{v}_i$.

Christensen, Pearson and Johnson (1992) develop case-deletion diagnostics for detecting influential observations in mixed linear models. Their methods can be applied to model (4.6) as well as to more complex small area models.

7.2 Constrained Estimation

Direct survey estimates are often adequate at an aggregate (or large area) level in terms of precision. For example, Battese, Harter and Fuller (1988), in their application, find that the direct regression estimator of the mean crop area for the 12 counties together has adequate precision. It is, therefore, sometimes desirable to modify the individual small area estimators so that a properly weighted sum of these estimators equals the model-free, direct estimator at the aggregate level. The modified estimators will be somewhat less efficient than the original, optimal estimators, but they avoid possible aggregation bias by ensuring consistency with the direct estimator. One simple way to achieve consistency is to make a ratio adjustment, for example, the EBLUP estimator $\hat{Y}_i^H$ of a total Y_i is modified to

$$\hat{Y}_i^H \text{ (mod)} = \left(\hat{Y}_i^H \Big/ \sum_i \hat{Y}_i^H \right) \hat{Y}, \tag{7.1}$$

where $\hat{Y}$ is a direct estimator of the aggregate population total $Y = \Sigma_i Y_i$. Battese, Harter and Fuller (1988) and Pfeffermann and Barnard (1991) propose alternative estimators involving estimated variances and covariances of the optimal estimators $\hat{Y}_i^H$.

The previous sections focused on simultaneous estimation of small area means or totals, but in some applications the main objective is to produce an ensemble of parameter estimates whose histogram is in some sense close to the histogram of small area parameters. Spjøtvoll and Thomsen (1987), for example, were interested in finding how 100 municipalities in Norway were distributed according to proportion of persons not in the labor force. They propose constrained *EB* estimators whose variation matched the variation of the small area population means. By comparing with the actual distribution in their example, they show that the *EB* estimators are biased toward the prior mean compared to the constrained *EB* estimators. Constrained estimators reduce shrinking towards the synthetic component; for example, in (5.1) the weight $1 - \gamma_i$, attached to the synthetic component, is reduced to $1 - \gamma_i^{1/2}$. Following Louis (1984), Ghosh (1992) develops a general theory of constrained HB estimation. Ghosh obtains constrained HB estimates by matching the first two moments of the histogram of the estimates, and the posterior expectations of the first two moments of the histogram of the parameters and minimizing, subject to these conditions, the posterior expectation of the Euclidean distance between the estimates and the parameters. Lahiri (1990) obtains similar results in the context of small area estimation, assuming "posterior linearity," thus avoiding distributional assumptions. Constrained Bayes estimates are suitable for subgroup analysis where the problem is not only to estimate the different components of a parameter vector but also to identify the parameters that are above or below a specified cut-off point. It should be noted that synthetic estimates are inappropriate for this purpose.

The optimal estimators (i.e., EBLUP, EB and HB estimators) may perform well overall but poorly for particular small areas that are not consistent with the assumed model on small area effects. To avoid this problem, Efron and Morris (1972) and Fay and Harriot (1979) suggest a straightforward compromise that consists of restricting the amount by which the optimal estimator differs from the direct estimator by some multiple of the standard error of the direct estimator. For example, a compromise estimator corresponding to the HB estimator $\hat{\theta}_i^{HB}$, under a normal prior on the θ_i's, is given by

$$\tilde{\theta}_i^{HB} = \begin{cases} \hat{\theta}_i^{HB}, & \text{if } \hat{\theta}_i - c\psi_i^{1/2} \le \hat{\theta}_i^{HB} \le \hat{\theta}_i + c\psi_i^{1/2} \\ \hat{\theta}_i - c\psi_i^{1/2}, & \text{if } \hat{\theta}_i^{HB} < \hat{\theta}_i - c\psi_i^{1/2} \\ \hat{\theta}_i + c\psi_i^{1/2}, & \text{if } \hat{\theta}_i^{HB} > \hat{\theta}_i + c\psi_i^{1/2}, \end{cases}$$

where $c > 0$ is a suitable chosen constant, say $c = 1$. A limitation of the compromise estimators is that no reliable measures of their precision are available.

7.3 Extensions

Various extensions of the basic models (4.4) and (4.6) have been studied in the literature. Due to space limitation, we can only mention some of these extensions.

Datta et al. (1992) extend the aggregate-level model (4.4) to the case of correlated sampling errors with a known covariance matrix and develop HB and EB estimators and associated measures of precision. In their application to adjustment of census undercount, the sampling covariance matrix is block diagonal. Cressie (1990a) introduces spatial dependence among the random effects v_i, in the context of adjustment for census undercount. Fay (1987) and Ghosh, Datta and Fay (1991) extend model (4.4) to multiple characteristics and perform hierarchical and empirical multivariate Bayes analysis, assuming that the sampling covariance matrix of $\hat{\boldsymbol{\theta}}_i$, the vector of direct estimators for ith area, is known for each i. In their application to estimation of median income for four-person families by state, $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2})^T$ with θ_{i1} = population median income of four-person families in state i and $\theta_{i2} = \frac{3}{4}$ (population median income of three-person families in state i) $+\frac{1}{4}$ (median income of four-person families in state i). By taking advantage of the strong correlation between the direct estimators $\hat{\theta}_{i1}$ and $\hat{\theta}_{i2}$, they were able to obtain improved estimators of θ_{i1}.

Many surveys are repeated in time with partial replacement of the sample elements, for example, the monthly U.S. Current Population Survey and the Canadian Labor Force Survey. For such repeated surveys considerable gain in efficiency can be achieved by borrowing strength across both small areas and time. Cronkite (1987) developed regression synthetic estimators using pooled cross-sectional time series data and applied them to estimate substate area employment and unemployment using the Current Population Survey monthly survey estimates as dependent variable and counts from the Unemployment Insurance System and Census variables as independent variables. Rao and Yu (1992) propose an extension of model (4.4) to time series and cross-sectional data. Their model is of the form

$$\text{(7.2)} \qquad \hat{\theta}_{it} = \theta_{it} + e_{it}, \qquad t = 1, \dots, T,$$

$$\text{(7.3)} \qquad \theta_{it} = \mathbf{x}_{it}'\beta + v_i + u_{it}, \qquad i = 1, \dots, m,$$

where $\hat{\theta}_{it}$ is the direct estimator for small area i at time t, the e_{it}'s are sampling errors with a known block diagonal covariance matrix Ψ = block diag $(\Psi_1, \dots, \Psi_m)$, $\mathbf{x}_{it}$ is a vector of covariates and $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$. Further, the u_{it}'s are assumed to follow a first order autoregressive process for each i, i.e., $u_{it} = \rho u_{i,t-1} + \epsilon_{it}, |\rho| < 1$ with $\epsilon_{it} \overset{\text{iid}}{\sim} N(0, \sigma^2)$. They obtain the EBLUP and HB estimators and their standard errors under (7.2) and (7.3).

Models of the form (7.3) have been extensively used in the econometric literature, ignoring sampling errors (see, e.g., Anderson and Hsiao, 1981; Judge, 1985, Chapter 13). Choudhry and Rao (1989) treat the composite error $w_{it} = e_{it} + u_{it}$ as a first order autoregressive process and obtain the EBLUP estimator of $\mathbf{x}_{it}^T\beta + v_i$. A drawback of their method is that the area by time specific effect u_{it} is ignored in modelling the θ_{it}'s.

Pfeffermann and Burck (1990) investigate more general models on the θ_{it}'s, but they assume modeling of sampling errors across time. They obtain EBLUP estimators of small area means using the Kalman filter. Singh and Mantel (1991) consider arbitrary covariance structures on sampling errors and propose recursive composite estimators using the Kalman filter. These estimators are not optimal but appear to be quite efficient relative to the corresponding EBLUP estimators.

Turning to extension of the nested error regression model (4.6), Fuller and Harter (1987) propose a multivariate nested error regression model and obtain EBLUP estimators and associated standard errors. Stukel (1991) studies two-fold nested error regression models, and obtains EBLUP estimators and associated standard errors. Such models are appropriate for two-stage sampling within small areas. Kleffe and Rao (1992) extend model (4.6) to the case of random error variances, σ_i^2, and obtain EBLUP estimator and associated standard errors in the special case of $\mathbf{x}_{ij} = 1$.

MacGibbon and Tomberlin (1989) and Malec, Sedransk and Tompkins (1991) study logistic regression models with random area-specific effects. Such models are appropriate for binary response variables when element-specific covariates are available. MacGibbon and Tomberlin (1989) obtain EB estimators of small area proportions and associated standard errors, but they ignore the uncertainty about the prior parameters. Farrell, MacGibbon and Tomberlin (1992) apply the bootstrap method of Laird and Louis (1987) to account for the underestimation of true posterior variance. Malec, Sedransk and Tompkins (1991) obtain HB estimators and associated standard errors using Gibbs sampling and apply their method to data from the U.S. National Health Interview Survey to produce estimates of proportions for individual states.

EB and HB methods have also been used for estimating regional mortality and disease rates (see, e.g., Marshall, 1991). In these applications, the observed small area counts, y_i, are assumed to be independent Poisson with conditional mean $E(y_i|\theta_i) = n_i\theta_i$, where θ_i and n_i respectively denote the true rate and number exposed in the ith area. Further, the θ_is are assumed to be random with a specified distribution (e.g., a gamma distribution with unknown scale and shape parameters). The EB or HB estimators are shrinkage estimators in the sense that the crude rate y_i/n_i is shrunk towards an overall regional rate, ignoring the spatial aspect of the problem. Marshall (1991) proposes "local" shrinkage estimators obtained by shrinking the crude rate towards a local neighbourhood rate. Such estimators are practically appealing and further work on their statistical properties is desirable.

De Souza (1992) studies joint mortality rates of two cancer sites over several geographical areas and obtains asymptotic approximations to posterior means and variances using the general first order approximations given by Kass and Steffey (1989). The bivariate model leads to improved estimators for each site compared to the estimators based on univariate models.

8. CONCLUSION

In this article, we have used the term "small area" to denote any local geographical area that is small or to describe any small subgroup of a population such as a specific age-sex-race group of people within a large geographical area. Sample sizes for small areas are typically small because the overall sample size in a survey is usually determined to provide desired accuracy at a much higher level of aggregation. As a result, the usual direct estimators of a small area mean are unlikely to give acceptable reliability; and it becomes necessary to "borrow strength" from related areas to find more accurate estimators for a given area or, simultaneously, for several areas. Considerable attention has been given to such indirect estimators in recent years.

We have attempted to provide an appraisal of indirect estimation covering both traditional design-based methods and newer model-based approaches to small area estimation. Traditional methods covered here include demographic techniques for local estimation of population and other characteristics of interest in post-censal years, and synthetic and sample size dependent estimation. Model-based methods studied here include EBLUP, EB and HB estimation. Two types of basic small area models that include random area-specific effects are used to describe these methods. In the first type of models, only area-specific auxiliary data are available for the population elements while in the second type element-specific auxiliary data are available for the population elements.

We have emphasized the importance of obtaining accurate measures of uncertainty associated with the model-based estimators. To this end, an approximately unbiased estimator of MSE of the EBLUP estimator is given as well as two methods of approximating the true posterior variance, irrespective of the form of the prior distribution on the model parameters. The latter approximations may be used as measures of uncertainty associated with the EB estimator. In the HB approach, a prior distribution on the model parameters is specified and the resulting posterior variance is used as a measure of uncertainty associated with the HB estimator (posterior mean). We have also mentioned several applications of the model-based methods.

We have also considered special problems that may be encountered in implementing model-based methods for small area estimation; in particular, model diagnostics for small area models, constrained estimation, "local" shrinkage, spatial modelling and borrowing strength across both small areas and time. We anticipate quite a bit of future research on these topics.

Caution should be exercised in using or recommending indirect estimators since they are based on implicit or explicit models that connect the small areas, unlike the direct estimators. As noted by Schaible (1992): "Indirect estimators should be considered when better alternatives are not available, but only with appropriate caution and in conjunction with substantial research and evaluation efforts. Both producers and users must not forget that, even after such efforts, indirect estimates may not be adequate for the intended purpose." (Also see Kalton, 1987.)

Finally, we should emphasize the need for developing an overall program that covers issues relating to sample design and data evolvement, organization and dissemination, in addition to those pertaining to methods of estimation for small areas.

APPENDIX

Variances and Covariance of $\hat{\sigma}_v^2$ and $\hat{\sigma}^2$

Let $\hat{\sigma}_v^2$ and $\hat{\sigma}^2$ be the estimators of σ_v^2 and σ^2 obtained from the method of fitting constants. Then

$$V(\hat{\sigma}^2) = 2\nu_1^{-1}\sigma^4$$

$$V(\hat{\sigma}_v^2) \doteq 2\eta_*^{-2}\Big[\nu_1^{-1}(n-p-\nu_1)(n-p)\sigma^4 + \eta_{**}\sigma_v^4 + 2\eta_*\sigma^2\sigma_v^2\Big]$$

with

$$\eta_{**} = \sum w_{i\cdot}^2\left(1 - w_{i\cdot}\bar{\mathbf{x}}_{iw}^T\mathbf{A}_1^{-1}\bar{\mathbf{x}}_{iw}\right) + \operatorname{tr}\left(\mathbf{A}_1^{-1}\sum w_{i\cdot}^2\bar{\mathbf{x}}_{iw}\bar{\mathbf{x}}_{iw}^T\right)^2$$

and

$$\operatorname{cov}(\hat{\sigma}_v^2, \hat{\sigma}_e^2) \doteq -2\eta_*^{-1}\nu_1^{-1}(n-p-\nu_1)\sigma^4.$$

(See Stukel, 1991.)

ACKNOWLEDGMENTS

The authors would like to thank the former editor, J. V. Zidek, and a referee for several constructive suggestions. M. Ghosh was supported by NSF Grants DMS-89-01334 and SES-92-01210. J. N. K. Rao's research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

ANDERSON, T. W. and HSIAO, C. (1981). Formulation and estimation of dynamic models using panel data. *J. Econometrics* **18** 67–82.

BATTESE, G. E., HARTER, R. M. and FULLER, W. A. (1988). An error-component model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **83** 28–36.

BECKMAN, R. J., NACHTSHEIM, C. J. and COOK, R. D. (1987). Diagnostics for mixed-model analysis of variance. *Technometrics* **29** 413–426.

BOGUE, D. J. (1950). A technique for making extensive postcensal estimates. *J. Amer. Statist. Assoc.* **45** 149–163.

BOGUE, D. J. and DUNCAN, B. D. (1959). A composite method of estimating post censal population of small areas by age, sex and colour. Vital Statistics–Special Report 47, No. 6, National Office of Vital Statistics, Washington, DC.

BRACKSTONE, G. J. (1987). Small area data: policy issues and technical challenges. In *Small Area Statistics* (R. Platek, J. N. K. Rao, C. E. Särndal and M. P. Singh eds.) 3–20. Wiley, New York.

CALVIN, J. A. and SEDRANSK, J. (1991). Bayesian and frequentist predictive inference for the patterns of care studies. *J. Amer. Statist. Assoc.* **86** 36–48.

CHAMBERS, R. L. and FEENEY, G. A. (1977). Log linear models for small area estimation. Unpublished paper, Australian Bureau of Statistics.

CHAUDHURI, A. (1992). Small domain statistics: a review. Technical report ASC/92/2, Indian Statistical Institute, Calcutta.

CHOUDHRY, G. H. and RAO, J. N. K. (1989). Small area estimation using models that combine time series and cross-sectional data. In *Symposium 89—Analysis of Data in Time—Proceedings* (A. C. Singh and P. Whitridge, eds.) 67–74. Statistics Canada, Ottawa.

CHRISTENSEN, R., PEARSON, L. M. and JOHNSON, W. (1992). Case deletion diagnostics for mixed models. *Technometrics* **34** 38–45.

CRESSIE, N. (1989). Empirical Bayes estimation of undercount in the decennial census. *J. Amer. Statist. Assoc.* **84** 1033–1044.

CRESSIE, N. (1990a). Small area prediction of undercount using the general linear model. In *Symposium 90—Measurement and Improvement of Data Quality—Proceedings* 93–105. Statistics Canada, Ottawa.

CRESSIE, N. (1992). REML estimation in empirical Bayes smoothing of census undercount. *Survey Methodology* **18** 75–94.

CRONKITE, F. R. (1987). Use of regression techniques for developing state and area employment and unemployment estimates. In *Small Area Statistics* (R. Platek, J. N. K. Rao, C. E. Särndal and M. P. Singh, eds.) 160–174. Wiley, New York.

DATTA, G. S. and GHOSH, M. (1991). Bayesian prediction in linear models: Applications to small area estimation. *Ann. Statist.* **19** 1748–1770.

DATTA, G. S., GHOSH, M., HUANG, E. T., ISAKI, C. T., SCHULTZ, L. K. and TSAY, J. H. (1992). Hierarchical and empirical Bayes methods for adjustment of census undercount: the 1980 Missouri dress rehearsal data. *Survey Methodology* **18** 95–108.

DEMPSTER, A. P. and RYAN, L. M. (1985). Weighted normal plots. *J. Amer. Statist. Assoc.* **80** 845–850.

DESOUZA, C. M. (1992). An approximate bivariate Bayesian method for analysing small frequencies. *Biometrics* **48** 1113–1130.

DREW, D., SINGH, M. P. and CHOUDHRY, G. H. (1982). Evaluation of small area estimation techniques for the Canadian Labour Force Survey. *Survey Methodology* **8** 17–47.

EFRON, B. and MORRIS, C. (1972). Limiting the risk of Bayes and empirical Bayes estimates—Part II: the empirical Bayes case. *J. Amer. Statist. Assoc.* **67** 130–139.

ERICKSEN, E. P. (1974). A regression method for estimating populations of local areas. *J. Amer. Statist. Assoc.* **69** 867–875.

ERICKSEN, E. P. and KADANE, J. B. (1985). Estimating the population in a census year (with discussion). *J. Amer. Statist. Assoc.* **80** 98–131.

Ericksen, E. P. and Kadane, J. B. (1987). Sensitivity analysis of local estimates of undercount in the 1980 U.S. Census. In *Small Area Statistics* (R. Platek, J. N. K. Rao, C .E. Särndal and M. P. Singh, eds.) 23–45. Wiley, New York.

Ericksen, E. P. and Kadane, J. B. (1992). Comment on "Should we have adjusted the U.S. Census of 1980?," by D. A. Freedman and W. C. Navidi. *Survey Methodology* **18** 52–58.

Ericksen, E. P., Kadane, J. B. and Tukey, J. W. (1989). Adjusting the 1981 census of population and housing. *J. Amer. Statist. Assoc.* **84** 927–944.

Ericson, W. A. (1969). A note on the posterior mean. *J. Roy. Statist. Soc. Ser. B* **31** 332–334.

Farrell, P. J., MacGibbon, B. and Tomberlin, T.J. (1992). An evaluation of bootstrap techniques for correcting empirical Bayes interval estimates. Unpublished manuscript, Dept. Statistics and Actuarial Science, Univ. Waterloo.

Fay, R. E. (1987). Application of multivariate regression to small domain estimation. In *Small Area Statistics* (R. Platek, J. N. K. Rao, C. E. Särndal and M. P. Singh, eds.) 91–102. Wiley, New York.

Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.* **74** 269–277.

Freedman, D. A. and Navidi, W. C. (1986). Regression models for adjusting the 1980 Census (with discussion). *Statist. Sci.* **1** 1–39.

Freedman, D. A. and Navidi, W. C. (1992). Should we have adjusted the U.S. Census of 1980? (with discussion). *Survey Methodology* **18** 3–74.

Fuller, W. A. and Battese, G. E. (1973). Transformations for estimation of linear models with nested error structure. *J. Amer. Statist. Assoc.* **68** 626–632.

Fuller, W. A. and Harter, R. M. (1987). The multivariate components of variance model for small area estimation. In *Small Area Statistics* (R. Platek, J. N. K. Rao, C. E. Särndal and M. P. Singh, eds.) 103–123. Wiley, New York.

Gelfand, A. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409.

Ghangurde, D. D. and Singh, M. P. (1977). Synthetic estimators in periodic households surveys. *Survey Methodology* **3** 152–181.

Ghosh, M. (1992a). Hierarchical and empirical Bayes multivariate estimation. In *Current Issues in Statistical Inference: Essays in Honor of D. Basu* (M. Ghosh and P. K. Pathak, eds.) 151–177. IMS, Hayward, CA.

Ghosh, M. (1992b). Constrained Bayes estimation with applications. *J. Amer. Statist. Assoc.* **87** 533–540.

Ghosh, M., Datta, G. S. and Fay, R. E. (1991). Hierarchical and empirical multivariate Bayes analysis in small area estimation. In *Proceeding of the Bureau of the Census Annual Research Conference* 63–79. Bureau of the Census, Washington, DC.

Ghosh, M. and Lahiri, P. (1987). Robust empirical Bayes estimation of means from stratified samples. *J. Amer. Statist. Assoc.* **82** 1153–1162.

Gonzalez, M. E. (1973). Use and evaluation of synthetic estimators. In *Proceedings of the Social Statistics Section* 33–36. Amer. Statist. Assoc., Washington, DC.

Hansen, M., Hurwitz, W. N. and Madow, W. G. (1953). *Sample Survey Methods and Theory*, **1**. Wiley, New York.

Harville, D. A. (1991). Comment on, "That BLUP is a good thing: The estimation of random effects," by G. K. Robinson. *Statist. Sci.* **6** 35–39.

Henderson, C. R. (1950). Estimation of genetic parameters (abstract). *Ann. Math. Statist.* **21** 309–310.

Holt, D., Smith, T. M. F. and Tomberlin, T. J. (1979). A model-based approach to estimation for small subgroups of a population. *J. Amer. Statist. Assoc.* **74** 405–410.

Hulting, F. L. and Harville, D. A. (1991). Some Bayesian and non-Bayesian procedures for the analysis of comparative experiments and for small area estimation: computational aspects, frequentist properties, and relationships. *J. Amer. Statist. Assoc.* **86** 557–568.

Isaki, C. T., Schultz, L. K., Smith, P. J. and Diffendal, D. J. (1987). Small area estimation research for census undercount—progress report. In *Small Area Statistics* (R. Platek, J. N. K. Rao, C.E. Särndal and M.P. Singh, eds.) 219–238. Wiley, New York.

Judge, G. G. (1985). *The Theory and Practice of Econometrics*, 2nd ed. Wiley, New York.

Kackar, R. N. and Harville, D. A. (1984). Approximations for standard errors of estimators for fixed and random effects in mixed models. *J. Amer. Statist. Assoc.* **79** 853–862.

Kalton, G. (1987). Discussion. In *Small Area Statistics* (R. Platek, J. N. K. Rao, C. E. Särndal and M. P. Singh, eds.) 264-266. Wiley, New York.

Kalton, G., Kordos, J. and Platek, R. (1993). *Small Area Statistics and Survey Designs Vol. I: Invited Papers; Vol. II: Contributed Papers and Panel Discussion.* Central Statistical Office, Warsaw.

Kass, R. E. and Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *J. Amer. Statist. Assoc.* **84** 717–726.

Kleffe, J. and Rao, J. N. K. (1992). Estimation of mean square error of empirical best linear unbiased predictors under a random error variance linear model. *J. Multivariate Anal.* **43** 1–15.

Lahiri, P. (1990). "Adjusted" Bayes and empirical Bayes estimation in finite population sampling. *Sankhyā Ser. B* **52** 50–66.

Lahiri, P. and Rao, J. N. K. (1992). Robust estimation of mean square error of small area estimators. Unpublished manuscript.

Laird, N. M. and Louis, T. A. (1987). Empirical Bayes confidence intervals based on bootstrap samples. *J. Amer. Statist. Assoc.* **82** 739–750.

Lange, N. and Ryan, L. (1989). Assessing normality in random effects models. *Ann. Statist.* **17** 624–642.

Louis T. (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods. *J. Amer. Statist. Assoc.* **79** 393–398.

MacGibbon, B. and Tomberlin, T.J. (1989). Small area estimation of proportions via empirical Bayes techniques. *Survey Methodology* **15** 237–252.

Malec, D., Sedransk, J. and Tompkins, L. (1991). Bayesian predictive inference for small areas for binary variables in the National Health Interview Survey. Unpublished manuscript.

Marker, D. A. (1983). Organization of small area estimators. *Proceedings of Survey Research Methods Section* 409–414. Amer. Statist. Assoc., Washington, DC.

Marshall, R. J. (1991). Mapping disease and mortality rates using empirical Bayes estimators. *J. Roy. Statist. Soc. Ser. C* **40** 283–294.

McCullagh, P. and Zidek, J. (1987). Regression methods and performance criteria for small area population estimation. In *Small Area Statistics* (R. Platek, J. N. K. Rao, C. E. Särndal and M. P. Singh, eds.) 62–74. Wiley, New York.

Morris, C. (1983). Parametric empirical Bayes inference: Theory and applications (with discussions). *J. Amer. Statist. Assoc.* **78** 47–65.

Morrison, P. (1971). Demographic information for cities: a manual for estimating and projecting local population characteristics. RAND report R-618-HUD.

National Center for Health Statistics (1968). *Synthetic State Estimates of Disability*. P.H.S. Publication 1759. U.S. Government Printing Office, Washington, DC.

National Institute on Drug Abuse (1979). *Synthetic Estimates* for *Small Areas* (research monograph 24). U.S. Government Printing Office, Washington, DC.

NATIONAL RESEARCH COUNCIL (1980). *Panel on Small-Area Estimates of Population and Income. Estimating Population and Income of Small Areas*. National Academy Press, Washington, DC.

NICHOL, S. (1977). A regression approach to small area estimation. Unpublished manuscript, Australian Bureau of Statistics, Canberra, Australia.

NTIS (1963). Indirect estimators in federal programs. Statistical policy working paper 21, prepared by the Subcommittee on Small Area Estimation, Federal Committee on Statistical Methodology, Office of Management and Budget, Washington, DC.

PFEFFERMANN, D. and BURCK, L. (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methodology* **16** 217–237.

PFEFFERMANN, D. and BARNARD, C. (1991). Some new estimators for small area means with applications to the assessment of farmland values. *Journal of Business and Economic Statistics* **9** 73–84.

PLATEK, R. and SINGH, M. P. (1986). *Small Area Statistics: Contributed Papers*. Laboratory for Research in Statistics and Probability, Carleton Univ.

PLATEK, R., RAO, J. N. K., SÄRNDAL, C. E. and SINGH, M. P. (1987). *Small Area Statistics*. Wiley, New York.

PRASAD, N. G. N. and RAO, J. N. K. (1990). The estimation of mean squared errors of small-area estimators. *J. Amer. Statist. Assoc.* **85** 163–171.

PURCELL, N. J. and LINACRE, S. (1976). Techniques for the estimation of small area characteristics. Unpublished manuscript.

PURCELL, N. J. and KISH, L. (1979). Estimation for small domain. *Biometrics* **35** 365–384.

PURCELL, N. J. and KISH, L. (1980). Postcensal estimates for local areas (or domains). *Internat. Statist. Rev.* **48** 3–18.

RAO, C. R. and SHINOZAKI, N. (1978). Precision of individual estimates in simultaneous estimation of parameters. *Biometrika* **65** 23–30.

RAO, J. N. K. (1986). Synthetic estimators, SPREE and best model based predictors. In *Proceedings of the Conference on Survey Research Methods in Agriculture* 1–16. U.S. Dept. Agriculture, Washington, DC.

RAO, J. N. K. and YU, M. (1992). Small area estimation by combining time series and cross-sectional data. In *Proceedings of the Survey Research Methods Section* 1–19. Amer. Statist. Assoc., Alexandria, VA.

ROBINSON, G. K. (1991). That BLUP is a good thing: the estimation of random effects (with discussion). *Statist. Sci.* **6** 15–51.

SÄRNDAL, C. E. and HIDIROGLOU, M. A. (1989). Small domain estimation: a conditional analysis. *J. Amer. Statist. Assoc.* **84** 266–275.

SCHAIBLE, W. L. (1978). Choosing weights for composite estimators for small area statistics. In *Proceedings of the Survey Research Methods Section* 741–746. Amer. Statist. Assoc., Washington, DC.

SCHAIBLE, W. L. (1992). Use of small area statistics in U.S. Federal Programs. In *Small Area Statistics and Survey Designs* (G. Kalton, J. Kordos and R. Platek, eds.) **1** 95–114. Central Statistical Office, Warsaw.

SINGH, A. C. and MANTEL, H. J. (1991). State space composite estimation for small areas. In *Symposium 91—Spatial Issues in Statistics—Proceedings* 17–25. Statistics Canada, Ottawa.

SINGH, M. P., GAMBINO, J. and MANTEL, H. (1992). Issues and options in the provision of small area data. In *Small Area Statistics and Survey Designs* (G. Kalton, J. Kordos and R. Platek, eds.) **1** 37–75. Central Statistical Office, Warsaw.

SMITH, S. K. and LEWIS, B. B. (1980). Some new techniques for applying the housing unit method of local population estimation. *Demography* **17** 323–340.

SPJØTVOLL, E. and THOMSEN, I. (1987). Application of some empirical Bayes methods to small area statistics. *Bulletin of the International Statistical Institute* **2** 435–449.

STARSINIC, D. E. (1974). Development of population estimates for revenue sharing areas. Census Tract Papers, Ser. GE40, No. 10 U.S. Government Printing Office, Washington, DC.

STATISTICS CANADA (1987). *Population Estimation Methods, Canada*, Catalogue 91-528E. Statistics Canada, Ottawa.

STEFFEY, D. and KASS, R. E. (1991). Comment on "That BLUP is a good thing: The estimation of random effects," by G. K. Robinson. *Statist. Sci.* **6** 45–47.

STUKEL, D. (1991). *Small Area Estimation Under One and Two-Fold Nested Error Regression Model*. Ph.D. Thesis, Carleton Univ.

U.S. BUREAU OF THE CENSUS (1966). Methods of population estimation: Part I, Illustrative procedure of the Bureau's component method II. *Current Population Reports, Series P-25, No. 339*. U.S. Government Printing Office, Washington, DC.

ZIDEK, J. V. (1982). A review of methods for estimating the populations of local areas. Technical Report 82–4, Univ. British Columbia, Vancouver.

Comment

Noel Cressie and Mark S. Kaiser

Malay Ghosh and Jon Rao have presented us with a well written exposition of the topic of small area estimation. The past literature has been decidedly influenced by linear modeling, and we see that clearly in their paper. There has also been a tendency to judge the performance of the estimation methods by concentrating on a single, arbitrary small area. In our comment, we shall discuss what opportunities there might be to expand the class of statistical models for small area data and to consider multivariate aspects of small area estimation.

Noel Cressie is Professor of Statistics and Distinguished Professor in Liberal Arts and Sciences and Mark S. Kaiser is Assistant Professor, Department of Statistics, Iowa State University, Snedecor Hall, Ames, Iowa 50011-1201.