

# Small Area Estimation – New Developments and Directions

**Danny Pfeffermann**

*appears in International Statistical Review, 2002*

*Department of Statistics, Hebrew University, Jerusalem, Israel*

*Department of Social Statistics, University of Southampton, United Kingdom*

## Summary

The purpose of this paper is to provide a critical review of the main advances in small area estimation (SAE) methods in recent years. We also discuss some of the earlier developments, which serve as a necessary background for the new studies. The review focuses on model dependent methods with special emphasis on point prediction of the target area quantities, and mean square error assessments. The new models considered are models used for discrete measurements, time series models and models that arise under informative sampling. The possible gains from modeling the correlations among small area random effects used to represent the unexplained variation of the small area target quantities are examined. For review and appraisal of the earlier methods used for SAE, see Ghosh and Rao (1994).

*Key words:* Best linear unbiased prediction; Cross-sectional correlations; Empirical Bayes; Hierarchical Bayes; Informative sampling; Mixed models; Time series models

## 1 Introduction

Small area estimation (SAE) is a topic of great importance due to the growing demand for reliable small area statistics even when only very small samples are available for these areas. Over the years, many statistical agencies have introduced vigorous programs to meet this demand. Extensive research on the theoretical and practical aspects of SAE is carried out and many international conferences and workshops are held in order to share the results of this research effort. Interest in small area estimation methods has further enhanced in recent years due to the tendency of many European countries to base future censuses on administrative record systems. Recognizing the inaccuracies of the administrative data and the fact that even the richest records cannot cover all the detailed information required for small census tracts, the idea is to test, correct and supplement the administrative information by sample data.

The problem of SAE is twofold. First is the fundamental question of how to produce reliable estimates of characteristics of interest, (means, counts, quantiles, etc.) for small areas or domains, based on very small samples taken from these areas. The second related question is how to assess the estimation error. Note in this respect that except in rare cases, sampling designs and in particular sample sizes are chosen in practice so as to provide reliable estimates for aggregates of the small areas such as large geographical regions or broad demographic groups. Budget and other constraints usually prevent the allocation of sufficiently large samples to each of the small areas. Also, it is often the case that domains of interest are only specified after the survey has already been designed and carried out. Having only a small sample (and possibly an empty sample) in a given area, the only possible solution to the estimation problem is to *borrow information* from other related data sets. Potential data sources can be divided into two broad categories:

- ☐ Data measured for the characteristics of interest in other ‘similar’ areas,
- ☐ Data measured for the characteristics of interest on previous occasions.

The methods used for SAE can be divided accordingly by the related data sources that they employ, whether cross-sectional (from other areas), past data or both. A further division classifies the methods by the type of inference: ‘design based’, ‘model dependent’ (with sub-division into the frequentist and Bayesian approaches), or the combination of the two. In what follows I describe briefly two real applications of SAE that illustrate what the areas or domains might be, the quantities of interest and the kind of concomitant variables that are used for estimation. Other examples are mentioned in subsequent sections.

*A- Estimation of illicit drug use-* Over the last six years, the Research Triangle Institute (RTI) in the U.S.A. is engaged in producing small area estimates of drug use for all the 50 states and the District of Columbia. Estimates of rates of use (prevalence), numbers of users and other measures related to the use of illicit drugs, alcohol and tobacco are derived for 32 demographic cells defined by age, race and sex within 300 regions (sub-states), based on the National Household Survey on Drug Abuse (NHSDA). The total sample size of the 1999 survey is about 67500 persons, with an average of 75 persons per age group. The estimates are derived by use of logistic models that contain 12-25 regressor variables with fixed

coefficients and state and sub-state random effects with age group specific elements as the independent variables (see Section 4.) The regressor variables comprise person and block group level demographic variables, census tract level demographic and socio-economic status variables and county level rates of drug-related arrests and deaths. The auxiliary information is available from various administrative records. For a detailed description of the NHSDA and the other data sources, the modeling process and estimates released, see the web site:

<http://www.samhsa.gov/oas/NHSDA/1999/Table%20of%20Contents.htm>. The article by Folsom *et al.* (1999) expands on methodological issues.

*B- Estimation of Small Area Employment-* The Bureau of Labor Statistics (BLS) in the U.S.A. is running a monthly survey of businesses within states, called the Current Employment Statistics (CES) survey (also known as the payroll survey). The survey is designed to produce monthly estimates of employment for major industry divisions within states and large metropolitan statistical areas. Monthly estimates for major industry divisions are, however, desired also for about 320 local labor market areas, defining a total of over than 2500 nonempty small domains, with many of the domains having only 10 or fewer responding units. The direct estimates obtained from the CES are therefore very erratic. Administrative data of total employment within the same domains are obtained from state unemployment insurance reports, collected for virtually all the businesses, but these reports, known as the ES202 data become available only with a time lag of 6-12 months. Recent studies (Harter *et al.* 1999) suggest that the direct small domain CES estimates can be improved by estimates of the form,

*employment in CES sample + non-sample predicted employment + “add-ons”.*

The first component of this estimate is the total measured employment for businesses in the CES sample in the month of investigation. The second component is a Ratio predictor of total employment in non-sample businesses of the form,  $\hat{E}_s = X_s \hat{R}_1$  where  $X_s$  denotes the small domain ES202 benchmark employment for non-sample businesses and  $\hat{R}_1$  is the ratio between the state-wide employment estimated from the CES sample and the state-wide employment obtained from the ES202 data in the corresponding industry. This predictor is an example of a ‘synthetic estimate’ (see Section 2) in that the ratio between the CES and ES202 estimates at the state level is applied to the ES202 data locally. The supplemental “add-ons”

adjust for employment that cannot be assigned to the industry division through specific firms. Enhancements to the above small domain estimates are currently investigated.

The purpose of this paper is to discuss some of the recent developments in SAE. The main emphasis is on general methodological issues rather than on detailed technical solutions. For these, as well as some recent applications of SAE, the reader is referred to the new review article by Rao (1999), which is an update of the more extensive review of Ghosh and Rao (1994). Another recent review of SAE methods is the article by Marker (1999).

Section 2 illustrates the important role of administrative information for SAE and introduces the family of synthetic regression estimators. Section 3 reviews several cross-sectional models in common use for continuous measurements. These models are extended to handle discrete measurements in Section 4 and to account for time series relationships in Section 5. The analysis in Sections 3-5 assumes implicitly that the sampling process can be ignored for inference. Section 6 considers the case of informative sampling under which the sample data no longer represent the population model. Section 7 examines the importance of modeling the correlations among the random area effects. I conclude with brief remarks in Section 8.

## 2 The importance of concomitant administrative data, synthetic estimation

In this and the next three sections we assume for convenience that the sample is selected by simple random sampling without replacement. The possible implications of the use of complex sampling schemes with unequal selection probabilities are discussed in Section 6. Let  $y$  define the characteristic of interest and denote by  $y_{ij}$  the outcome value for unit  $j$  belonging to area  $i$ ,  $i = 1 \dots m$ ;  $j = 1 \dots N_i$ , where  $N_i$  is the area size. Let  $s = s_1 \cup \dots \cup s_m$  signify the sample where  $s_i$  of size  $n_i$  defines the sample observed for area  $i$ . Notice that the  $n_i$ 's are random unless a separate sample with fixed sample size is taken in every area. Suppose that the objective is to estimate the true area mean  $\bar{Y}_i = \sum_{j=1}^{N_i} y_{ij} / N_i$ . If no auxiliary information is available, the ordinary *direct* design unbiased estimator and its design variance over the *randomization distribution* (the distribution induced by the random selection of the sample with the population values held fixed), are given by

$$\bar{y}_i = \sum_{j=1}^{n_i} y_{ij} / n_i \quad ; \quad V_D[\bar{y}_i | n_i] = (\tilde{S}_i^2 / n_i)[1 - (n_i / N_i)] = S_i^{*2} \quad (2.1)$$

where  $\tilde{S}_i^2 = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 / (n_i - 1)$ . Clearly, for small  $n_i$  the variance will be large unless the variability of the  $y$ -values is sufficiently small. Suppose, however, that in addition to measuring  $y$ , values  $x_{ij}$  of  $p$  concomitant variables  $x(1) \dots x(p)$  are also known for each of the sample units and that the row area means  $\bar{X}_i = \sum_{j=1}^{n_i} x_{ij} / n_i$  are likewise known. Such information may be obtained from a recent census or some other administrative records, see the examples in the introduction. In this case, a more efficient design unbiased estimator is the regression estimator,

$$\bar{y}_{reg,i} = \bar{y}_i + (\bar{X}_i - \bar{x}_i)' \beta_i \quad ; \quad V_D(\bar{y}_{reg,i} | n_i) = S_i^{*2} (1 - R_i^2) \quad (2.2)$$

where  $\bar{x}_i = \sum_{j=1}^{n_i} x_{ij} / n_i$ , and  $\beta_i$  and  $R_i$  are correspondingly the vector of regression coefficients and the multiple correlation coefficient between  $y$  and  $x(1) \dots x(p)$  computed from all the  $N_i$  measurements in area  $i$ . Thus, by use of the concomitant variables, the variance is reduced by the factor  $(1 - R_i^2)$ , illustrating very clearly the importance of using auxiliary information with good prediction power in SAE. Other well-known uses of auxiliary information for direct estimation are the Ratio estimator and Poststratification.

The problem with the use of  $\bar{y}_{reg,i}$  is that in practice the coefficients  $\beta_i$  are seldom known. Replacing  $\beta_i$  by its ordinary least square estimator from the sample  $s_i$  is not effective because of the small sample size. If, however, the  $\beta_i$ 's are known to be 'similar' across the small areas and likewise for the 'intercepts'  $(\bar{y}_i - \bar{X}_i' \beta_i)$ , a more stable estimator is the *synthetic regression estimator*  $\bar{y}_{reg,i}^{syn} = (\bar{y} - \bar{x}'b) + \bar{X}_i' b$ , where  $(\bar{y}, \bar{x}') = \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij}, x_{ij}') / \sum_{i=1}^m n_i$  are the global sample means and  $b = [\sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)']^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i)$  is a pooled estimator computed likewise from all the samples  $s_i$ . In the special case of a single concomitant variable and 'zero intercepts',  $\bar{y}_{reg,i}^{syn}$  is replaced by the *synthetic Ratio estimator*  $\bar{y}_{ratio,i}^{syn} = \bar{X}_i \bar{y} / \bar{x}$ . This is basically the ratio predictor  $\hat{E}_{\bar{y}}$  used in Example B of the introduction.

The term “synthetic” refers to the fact that an estimator computed from a large domain is used for each of the separate areas comprising that domain assuming that the areas are ‘homogeneous’ with respect to the quantity that is estimated. Thus, synthetic estimators already borrow information from other ‘similar areas’. Another, even simpler example is the use of the global mean  $\bar{y}$  for estimating each of the small area means when no auxiliary information is available. The article by Marker (1999) contains a thorough discussion of synthetic estimators with many examples. Ghosh and Rao (1994) provide model-based justifications for some of the synthetic estimators in common use.

The prominent advantage of synthetic estimation is the potential for substantial variance reduction but it can lead to severe biases if the assumption of homogeneity within the larger domain is violated. For example, the (unconditional) design bias of the synthetic regression estimator is approximately  $Bias_D(\bar{y}_{reg,i}^{syn}) \cong \bar{Y} - \bar{X}B - (\bar{Y}_i - \bar{X}_i'B)$ , where  $\bar{Y}$  and  $\bar{X}$  are the true large domain means of  $y$  and  $x$  and  $B$  is the corresponding regression coefficient. The bias can be large unless the intercept and slope coefficients are similar across the areas. Note again that the design bias is computed with respect to the randomization distribution (repeated sampling). Bias reduction under this distribution (but at the expense of increased variance) can be achieved by the use of *composite estimators*. A composite estimator is a weighted sum of the area direct estimator (has small or no bias but large variance) and the synthetic estimator (has small variance but possibly large bias). Thus, denoting more generally by  $\theta_i$  the small area characteristic of interest, a composite estimator has the general form,

$$\hat{\theta}_{com,i} = w_i \tilde{\theta}_i + (1 - w_i) \hat{\theta}_i^{syn} \quad (2.3)$$

where  $\tilde{\theta}_i$  is the direct estimator and  $\hat{\theta}_i^{syn}$  is the synthetic estimator. The question underlying the use of composite estimators is the choice of the weights  $w_i$ . Ideally, the weights should be selected so as to minimize the mean square error (MSE) but this is problematic since the MSE of the synthetic estimator is usually unknown because of its bias. One simple alternative therefore is to set  $w_i = f_i = (n_i / N_i)$  so that more weight is given to the direct unbiased estimator as the sampling fraction  $f_i$  increases. However, the sampling fractions are ordinarily very small and hence the use of this weight practically implies the use of the synthetic estimator, irrespective of the variability of the outcome variable. Other plausible

specifications of the weights are discussed in Ghosh and Rao (1994), Thomsen and Holmoy (1998) and Marker (1999). Optimal choices of the weights under mixed linear models are considered in Section 3.

A common feature of the estimators considered in this section is that they are ‘model free’ in the sense that no explicit model assumptions are used for their derivation, and the variance and bias are computed with respect to the randomization distribution. The article by Marker (1999) contains a historical survey of design-based estimators with many references. In the rest of the paper I consider model dependent estimators.

### 3 Cross-sectional models for continuous measurements

SAE is widely recognized as one of the few problems in survey sampling where the use of models is often inevitable. The specification of an appropriate working model permits the construction of correspondingly efficient estimators and the computation of variances and confidence intervals, which may not be feasible under the randomization distribution. (With the very small sample sizes often encountered in practice, large sample normal theory does not apply.) The models reviewed in this and the next sections are ‘mixed effects models’, containing fixed and random effects. Some special features of the application of these models for SAE problems are summarized at the end of the section.

One of the simplest models in common use is the ‘nested error unit level regression model’, employed originally by Battese *et al.* (1988) for predicting areas under corn and soybeans in 12 counties of the state of Iowa in the U.S. Suppose that the values of concomitant variables  $x(1)...x(p)$  are known for every unit in the sample and that the true area means of these variables are also known. Denoting by  $x_{ij}$  the concomitant values for unit  $j$  in area  $i$ , the model has the form,

$$y_{ij} = x_{ij}'\beta + u_i + \varepsilon_{ij} \quad (3.1)$$

where  $u_i$  and  $\varepsilon_{ij}$  are mutually independent error terms with zero means and variances  $\sigma_u^2$  and  $\sigma_\varepsilon^2$  respectively. The random term  $u_i$  represents the joint effect of area characteristics that are not accounted for by the concomitant variables. (In Battese *et al.* 1988  $y_{ij}$  is the reported number of hectares of corn (or soybeans) in sample segment  $j$  of county  $i$  and

$x_{ij}' = (x_{1ij}, x_{2ij})$  denotes the numbers of pixels classified as corn and soybeans from satellite pictures. The satellite information is known for both the sample and nonsample segments.) Under the model the true small area means are  $\bar{Y}_i = \bar{X}_i' \beta + u_i + \bar{\varepsilon}_i$  but since  $\bar{\varepsilon}_i = \sum_{j=1}^{N_i} \varepsilon_{ij} / N_i \cong 0$  for large  $N_i$ , the target parameters are ordinarily defined to be  $\theta_i = \bar{X}_i' \beta + u_i$ . For known variances  $(\sigma_u^2, \sigma_\varepsilon^2)$ , the Best Linear Unbiased Predictor (BLUP) of  $\theta_i$  under the model is,

$$\hat{\theta}_i = \gamma_i [\bar{y}_i + (\bar{X}_i - \bar{x}_i)' \hat{\beta}_{GLS}] + (1 - \gamma_i) \bar{X}_i' \hat{\beta}_{GLS} \quad (3.2)$$

where  $\hat{\beta}_{GLS}$  is the (optimal) Generalized Least Square (GLS) estimator of  $\beta$  computed from all the observed data and  $\gamma_i = \sigma_u^2 / (\sigma_u^2 + \sigma_\varepsilon^2 / n_i)$ . For areas  $k$  with no samples,  $\hat{\theta}_k = \bar{X}_k' \hat{\beta}_{GLS}$ . The coefficient  $\gamma_i$  is a “shrinkage factor” providing a trade-off between the (usually large) variance of the regression predictor  $\bar{y}_i + (\bar{X}_i - \bar{x}_i)' \hat{\beta}_{GLS}$ , and the bias of the synthetic estimator  $\bar{X}_i' \hat{\beta}_{GLS}$  for given value  $\theta_i$ . (The synthetic estimator and hence the BLUP are biased when conditioning on  $\theta_i$  or equivalently on  $u_i$ . The two predictors are unbiased unconditionally under the model since  $E(\bar{Y}_i) = \bar{X}_i' \beta$ .) The predictor  $\hat{\theta}_i$  has the structure of the composite estimator  $\hat{\theta}_{com,i}$  defined by (2.3). However, the weight  $\gamma_i$  is chosen in an optimal way under the model so that it accounts for the magnitude of the differences between the area effects  $u_i$ . Thomsen (in Ghosh and Rao, 1994), and Thomsen and Holmoy (1998) comment that predictors of the form (3.2) tend to over-estimate area means with small random effects and under-estimate area means with large effects such that the variation between the predictors is smaller than the variation between the true means. This is clear considering the structure of these predictors but it raises the question of the appropriate loss function in a particular application.

The BLUP  $\hat{\theta}_i$  is also the Bayesian predictor (posterior mean) under normality of the error terms and diffuse prior for  $\beta$ . In practice, however, the variances  $\sigma_u^2$  and  $\sigma_\varepsilon^2$  are seldom known. A common procedure is to replace them in the BLUP formula by standard variance components estimates like Maximum Likelihood Estimators (MLE), Restricted MLE (REML) or Analysis of Variance (ANOVA) estimators. The resulting predictors are known as



the Empirical BLUP (EBLUP) or Empirical Bayes (EB) predictors, see, for example, Prasad and Rao (1990) for details. Alternatively, Hierarchical Bayes (HB) predictors can be developed by specifying prior distributions for  $\beta$  and the two variances and computing the posterior distribution  $f(\theta_i | y, X)$  given all the observations in all the areas, see, e.g., Datta and Ghosh (1991). The actual application of this approach can be quite complicated since the posterior mean,  $E(\theta_i | y, X)$  has generally no close form. Recent studies use the Gibbs sampler (Gelfand and Smith, 1990) or other Markov Chain Monte Carlo (MCMC) techniques for stochastic simulation. The HB approach is very general and also very appealing since it produces the posterior variances associated with the point predictors (see below), but it requires in addition to the specification of the prior distributions good computing skills and intensive computations.

A somewhat different model from (3.1) discussed extensively in the literature is the ‘area level random effects model’, which is used when the concomitant information is only at the area level. Let  $x_i$  represent this information. The model, used originally by Fay and Herriot (1979) for the prediction of mean per capita income (PCI) in small geographical areas within counties (less than 500 persons) is defined as,

$$\tilde{\theta}_i = \theta_i + e_i ; \quad \theta_i = x_i' \beta + u_i \quad (3.3)$$

where  $\tilde{\theta}_i$  denotes the direct sample estimator (for example, the sample mean  $\bar{y}_i$ ), so that  $e_i$  represents in this case the sampling error, assumed to have zero mean (see comment below) and known design variance  $Var_D(e_i) = \sigma_{Di}^2$ , ( $S_i^{*2}$  if  $\tilde{\theta}_i = \bar{y}_i$ , equation 2.1). The model (3.3) integrates therefore a model dependent random effect  $u_i$  and a sampling error  $e_i$  with the two errors being independent. (In Fay and Herriot  $\tilde{\theta}_i$  is the direct sample estimate of mean PCI in ‘local government unit’  $i$  and  $x_i$  contains data on average county PCI, tax returns and values of housing. All the variables are measured in the log scale.) The BLUP under this model is,

$$\hat{\theta}_i = \gamma_i \tilde{\theta}_i + (1 - \gamma_i) x_i' \hat{\beta}_{GLS} = x_i' \hat{\beta}_{GLS} + \gamma_i (\tilde{\theta}_i - x_i' \hat{\beta}_{GLS}) \quad (3.4)$$

which again is a composite estimator with weight  $\gamma_i = \sigma_u^2 / (\sigma_{Di}^2 + \sigma_u^2)$ . In practice, the variances  $\sigma_u^2$  and  $\sigma_{Di}^2$  are usually unknown and they are replaced by sample estimates,

yielding in turn the corresponding EBLUP predictors. Arora and Lahiri (1997) show that if the variances  $\sigma_{Di}^2$  are considered random with a non-degenerate prior, then for known  $\beta$  and  $\sigma_u^2$  the Bayesian predictor has a smaller MSE than the corresponding BLUP.

*Comment:* The target area quantity  $\theta_i$  is often a nonlinear function of the area mean  $\bar{Y}_i$  so that the direct estimator  $\tilde{\theta}_i$  is a nonlinear function of the sample mean  $\bar{y}_i$ . For example, Fay and Herriot (1979) use the log transformation for estimating area per capita incomes. The problem arising in the case of nonlinear transformations is that the assumption  $E_D(e_i) = 0$  (design unbiasedness of the direct estimator) may not hold even approximately if the sample sizes are too small, requiring instead the use of Generalized Linear Models (see Section 4).

As defined in Section 1, an important aspect of SAE is the assessment of the prediction errors. This problem does not exist in principle under the full Bayesian paradigm, which produces the posterior variances of the target quantities around the HB predictors (the posterior means). However, as already stated, the implementation of this approach requires the specification of prior distributions and the computations can become very intensive. Assessment of the prediction errors under the EBLUP and EB approaches is also complicated because of the added variability induced by the estimation of the model parameters. To illustrate the problem, consider the model defined by (3.3) and suppose that the design variances  $\sigma_{Di}^2$  are known. (Stable variance estimators are often calculated as  $\hat{\sigma}_{Di}^2 = \hat{\sigma}_D^2 / n_i$  with  $\hat{\sigma}_D^2$  computed from all the data or obtained from other sources. The estimators  $\hat{\sigma}_{Di}^2$  are treated as the true variances.) If  $\beta$  and  $\sigma_u^2$  were also known, the variance of the BLUP (or HB under normality assumptions) is,  $Var[\hat{\theta}_i(\sigma_u^2, \beta)] = \gamma_i \sigma_{Di}^2 = g_{li}$ . Under the EBLUP and EB approaches,  $\beta$  and  $\sigma_u^2$  are replaced by sample estimates and a naïve variance estimator is obtained by replacing  $\sigma_u^2$  by  $\hat{\sigma}_u^2$  in  $g_{li}$ . This estimator ignores the variability of  $\hat{\sigma}_u^2$  and hence underestimates the true variance. Prasad and Rao (1990), extending the work of Kackar and Harville (1984) approximate the true prediction MSE of the EBLUP under normality of the two error terms and for the case where  $\sigma_u^2$  is estimated by the ANOVA (fitting of constants) method as,

$$MSE[\hat{\theta}_i(\hat{\sigma}_u^2, \hat{\beta})] = E[\hat{\theta}_i(\hat{\sigma}_u^2, \hat{\beta}) - \theta_i]^2 = g_{1i} + g_{2i} + g_{3i} \times Var(\hat{\sigma}_u^2) \quad (3.5)$$

where  $g_{2i} = (1 - \gamma_i)^2 x_i Var(\hat{\beta}_{GLS}) x_i$  and  $g_{3i} = [\sigma_{Di}^4 / (\sigma_{Di}^2 + \sigma_u^2)^3]$ . The term  $g_{2i}$  is the excess in MSE due to estimation of  $\beta$  and  $g_{3i} \times Var(\hat{\sigma}_u^2)$  is the excess in MSE due to estimation of  $\sigma_u^2$ . The neglected terms in the approximation are of order  $o(1/m)$  with  $m$  denoting the number of sampled areas. Note that the leading term in (3.5) is  $g_{1i} = \gamma_i \sigma_{Di}^2$  such that for large  $m$  and small values  $\gamma_i$ ,  $MSE[\hat{\theta}_i(\hat{\sigma}_u^2, \hat{\beta})] \ll \sigma_{Di}^2 = Var_D(\tilde{\theta}_i)$  illustrating the possible gains from using the model dependent predictor. Building on the approximation (3.5), Prasad and Rao (1990) develop a MSE estimator with bias of order  $o(1/m)$  as,

$$M\hat{S}E[\hat{\theta}_i(\hat{\sigma}_u^2, \hat{\beta})] = g_{1i}(\hat{\sigma}_u^2) + g_{2i}(\hat{\sigma}_u^2) + 2g_{3i}(\hat{\sigma}_u^2) \times V\hat{a}r(\hat{\sigma}_u^2) \quad (3.6)$$

where  $g_{ki}(\hat{\sigma}_u^2)$  is obtained from  $g_{ki}$  by substituting  $\hat{\sigma}_u^2$  for  $\sigma_u^2$ ,  $k=1,2,3$ . A similar estimator is developed for the nested error regression model (3.1). Lahiri and Rao (1995) show that the estimator (3.6) is robust to departures from normality of the random area effects  $u_i$  (but not the sampling errors  $e_i$ ).

Datta and Lahiri (2000) extend the results of Prasad and Rao to other variance components estimators and general linear mixed models of the form,

$$Y_i = X_i \beta + Z_i u_i + e_i = \xi_i + e_i, \quad i = 1 \dots m. \quad (3.7)$$

In (3.7),  $Y_i$  is a vector observation of order  $n_i$ ,  $\beta$  is a vector of fixed coefficients,  $Z_i$  is a fixed matrix of order  $n_i \times d$  and  $u_i$  and  $e_i$  are independent vector random effects and residual terms of orders  $d$  and  $n_i$  respectively. It is further assumed that  $u_i \sim N_d(0, G_i)$ ,  $e_i \sim N_{n_i}(0, \Sigma_i)$ , with  $G_i$  and  $\Sigma_i$  being functions of some vector parameter  $\lambda$ . The model contains as special cases familiar models used for SAE like the models defined by (3.1) and (3.3), the model considered by Moura and Holt (1999) that permits the regression coefficients in (3.1) to depend stochastically on area level auxiliary variables, and the two-folded nested error regression model studied by Stukel and Rao (1999). Explicit expressions of the BLUP (or posterior means) of linear combinations of the fixed and random effects and their variances (posterior variances) are derived for the case of known  $\lambda$  using results from

Henderson's (1975). The authors develop MSE estimators with bias of order  $o(1/m)$  for the EBLUP obtained when estimating  $\lambda$  by MLE or REML.

The MSE approximations discussed so far are under the frequentist approach. The use of the EB approach again requires appropriate measures of errors. Kass and Steffey (1989) develop first and second order EB MSE estimators for a general two-stage hierarchical model by approximating the posterior distribution of the hyper-parameters indexing the second stage model ( $\beta$  and  $\lambda$  in the notation of 3.7) by the normal distribution. Singh *et al.* (1998) study modifications to the Kass-Steffey approximations and propose Monte Carlo alternatives to some of the frequentist and EB measures of error.

*Comment:* The MSE approximations under both the frequentist and the EB approaches involve bias corrections of desired order and as a result, the use of these approximations improves the coverage properties of standard confidence intervals for the small area quantities. However, as found empirically by Singh *et al.* (1998), the addition of bias corrections can inflate the variance of the MSE approximations to a degree where it offsets the reduction in the bias.

We conclude this section by pointing out the special features of the use of mixed effects models for SAE. The most prominent feature is that in SAE the target parameters are the realized (actual) small area quantities, which renders the problem into a prediction problem. In other familiar applications of these models like in animal breeding or in education, the prime interest is in inference on the fixed regression coefficients and the variance components, which is only an intermediate step in SAE. The fact that the target quantities in SAE are the random area realizations complicates the evaluation of the prediction MSE very substantially since the unknown variance components are replaced by sample estimates. Indeed, the major developments in evaluation of prediction MSE for mixed effect models over the last decade have originated and evolved in the SAE context.

A second important feature of the application of mixed effects models for SAE is the kind of data available for analysis. On the one hand there often exists a large volume of administrative data at the unit and/or the area level. On the other hand, agencies collecting the data are often reluctant to release the micro data for secondary analysis so that the data available on the response variable consists of only the direct estimators that are subject to sampling errors that need to be modeled. The sampling designs used to select the sample may

dictate the use of weighted estimators, which further complicates the modeling process and the evaluation of the prediction errors. This issue is discussed in section 6.

#### 4 Models for discrete measurements

Recent research in SAE with some real breakthroughs focuses on situations where the measurements  $y_{ij}$  are categorical or discrete and the small area quantities of interest are proportions or counts. In such cases, the mixed linear models considered before are no longer applicable. MacGibbon and Tomberlin (1989) consider the following model for the case of binary measurements,

$$\begin{aligned} \Pr(y_{ij} = 1 | p_{ij}) &= p_{ij} ; \Pr(y_{ij} = 0 | p_{ij}) = (1 - p_{ij}); \\ \log it(p_{ij}) &= \log[p_{ij} / (1 - p_{ij})] = x_{ij}'\beta + u_i ; \quad u_i \sim N(0, \sigma_u^2). \end{aligned} \quad (4.1)$$

The outcomes  $y_{ij}$  are assumed to be conditionally independent and likewise for the random effects  $u_i$ . The purpose is to predict the true area proportions  $p_i = \sum_{j=1}^{N_i} y_{ij} / N_i$ . In an application considered by the authors  $y_{ij}$  defines labour force participation for sample individual  $j$  in county  $i$  and  $x_{ij}$  defines ‘sex’ and ‘age’. Assuming a diffuse prior for  $\beta$  and known variance  $\sigma_u^2 = \text{Var}(u_i)$ , the authors approximate the joint posterior distribution of  $\beta$  and  $\{u_i, i = 1 \dots m\}$  given the data by the multivariate normal distribution with mean equal to the mode of the true posterior and covariance matrix equal to the inverse information matrix evaluated at the mode. Denoting by  $\tilde{\beta}$  and  $\tilde{u}_i$  the respective modes,  $p_{ij}$  is estimated as  $\tilde{p}_{ij} = [1 + \exp\{-(x_{ij}'\tilde{\beta} + \tilde{u}_i)\}]^{-1}$  and for small sampling fractions  $n_i / N_i$ ,  $\hat{p}_i = \sum_{j=1}^{N_i} \tilde{p}_{ij} / N_i$ . The case of unknown  $\sigma_u^2$  is treated by repeating the same analysis with  $\sigma_u^2$  set to its MLE, thus yielding the corresponding EB predictor. A *naive* estimator for the variance of the empirical predictor is obtained by use of Taylor expansion but this estimator ignores the variance component resulting from the estimation of  $\sigma_u^2$ . Farrell, MacGibbon and Tomberlin (1997) show that the naive variance estimator can be improved by use of a parametric Bootstrap procedure proposed by Laird and Louis (1987).

The model defined by (4.1) has been studied more recently by Jiang and Lahiri (1998). The authors derive the Best Predictor (BP) of  $\log it(p_{ij})$  for given values  $(\sigma_u^2, \beta)$  as

$\log \hat{it}(p_{ij}) = x_{ij}'\beta + E(u_i | \sum_{j=1}^{n_i} y_{ij})$ , with the conditional expectation obtained as a ratio of two one-dimensional integrals. The estimator has generally no closed form because of the complex integration involved, but for the special case where  $x_{ij} = x_i$  for all  $j$  and  $\sigma_u^2 \rightarrow \infty$ ,  $E(u_i | y_i) \cong [\log \hat{it}(\bar{y}_i) - x_i'\beta]$ . Notice that in this case  $\log \hat{it}(p_{ij}) = \log \hat{it}(\bar{y}_i)$ . An Empirical BP (EBP) is obtained by replacing  $\beta$  and  $\sigma_u^2$  by sample estimates obtained by the *method of moments*. An approximation to the MSE of the EBP with bias of order  $o(1/m)$  is developed.

Malec *et al.* (1997) also consider Bernoulli outcomes imposing the same probabilities of ‘ones’ for all individuals  $j$  in area  $i$  belonging to the same socio-economic/demographic class  $k$  and the same ‘cluster’  $c$  (homogeneous geographic unit). Specifically, the model assumes,

$$\Pr(y_{ijk} = 1) = p_{kc}; \quad \log \hat{it}(p_{kc}) = x_k'\beta_c, \quad \beta_c \sim N(z_c\eta, \Gamma) \quad (4.2)$$

where  $x_k$  is a vector of regressor values defining the  $k$ -th class and  $z_c$  represents cluster level covariates. The random coefficients  $\beta_c$  are independent between the clusters. The authors compare several predictors of the totals  $\theta_i = \sum_c \sum_k \sum_{j=1}^{N_{ikc}} y_{ijk}$  where  $N_{ikc}$  is the number of individuals in class  $k$  belonging to cluster  $c$  of area  $i$ . Model dependent predictors are derived by evaluation of the conditional expectations  $E(p_{kc} | y, x)$ . For application of the Hierarchical Bayes (HB) approach, the hyper-parameters  $(\eta, \Gamma)$  are assigned the improper prior  $p(\eta, \Gamma) = \text{constant}$ . The computation of the HB predictors is carried out by use of the Gibbs sampler. The authors consider also several ‘time saving’ approximations to the HB predictors, obtained by replacing some of the conditional densities underlying the use of the Gibbs sampler by normal densities. Empirical Bayes (EB) predictors are obtained by computing (numerically) the expectations  $E(p_{kc} | y_s, \eta, \Gamma)$  with  $(\eta, \Gamma)$  assumed known (set at their MLE). Other estimators considered are “*synthetic*” estimators obtained by setting  $\beta_c \equiv z_c\eta$  ( $\Gamma = 0$ ) and  $p(\eta) = \text{constant}$  in the equation defining the HB point estimator, and *direct* design based estimators. Empirical results based on simulated and real data indicate that unless the sample sizes within the areas are sufficiently large, the exact HB predictors dominate the other estimators both with respect to point estimation and variance measures. It should be noted, however, that no bias corrections that account for parameter estimation have

been applied when estimating the variances of the EB estimators. As noted before, the HB approach ‘automatically’ accounts for all sources of error.

Ghosh *et al.* (1998) develop a general methodology for HB inference under the Generalized Linear Model (GLM) with random effects that includes the model (4.1) as a special case. The GLM (McCullagh and Nelder, 1989) is defined as,

$$f(y_{ij} | \theta_{ij}, \phi_{ij}) = \exp\{[(y_{ij}\theta_{ij} - b(\theta_{ij}))/\phi_{ij}] + a(\phi_{ij}) + c(y_{ij}, \phi_{ij})\} \quad (4.3)$$

where  $a(\cdot) > 0$ ,  $b(\cdot)$  and  $c(\cdot)$  are known functions. As easily verified,  $E(y_{ij}) = \partial b(\theta_{ij}) / \partial \theta_{ij}$ ;  $Var(y_{ij}) = \{\partial^2 [b(\theta_{ij})] / \partial \theta_{ij}^2\} \phi_{ij}$ . The scale parameters  $\phi_{ij}$  are assumed known. In order to borrow information across the areas, the ‘natural parameters’  $\theta_{ij}$  are modeled as  $h(\theta_{ij}) = x_{ij}'\beta + u_i + \varepsilon_{ij}$  where  $h(\cdot)$  is a strictly increasing *link function* with  $u_i$  and  $\varepsilon_{ij}$  playing a similar role as in the mixed linear model. The class of HB models studied by Ghosh *et al.* (1998) consists of the following equations (see below for an application):

$$\begin{aligned} y_{ij} | \theta_{ij}, \beta, u_i, \sigma_u^2, \sigma_\varepsilon^2 &\stackrel{ind}{\sim} GLM \\ h(\theta_{ij}) | \beta, u_i, \sigma_u^2, \sigma_\varepsilon^2 &\stackrel{ind}{\sim} N(x_{ij}'\beta + u_i, \sigma_\varepsilon^2) \quad ; \quad u_i | \beta, \sigma_u^2, \sigma_\varepsilon^2 \stackrel{ind}{\sim} N(0, \sigma_u^2) \end{aligned} \quad (4.4)$$

$$\beta \sim Uniform(R^k) ; (\sigma_u^2)^{-1} \sim Gamma(a, b) ; (\sigma_\varepsilon^2)^{-1} \sim Gamma(c, d) .$$

with  $\beta, \sigma_u^2, \sigma_\varepsilon^2$  mutually independent and  $(a, b, c, d)$  denoting fixed parameter values. The joint posterior distribution of the parameters  $\theta_{ij}$  or functions of them (like the means  $E(y_{ij} | \theta_{ij})$ ) is evaluated by the Gibbs sampler. It is essential, however, to ensure that the resulting posterior is *proper* and the authors establish sufficient conditions to this effect. In the special case of the model (4.1), the conditions reduce to the requirement that the measurements  $y_{ij}$  for a given area  $i$  are not all zeroes or ones. When the  $y_{ij}$  are Poisson (see below), the condition is  $y_i = \sum_{j=1}^{n_i} y_{ij} > 0$ .

As an example, consider the model applied by Ghosh *et al.* (1998) for estimating lung cancer death counts in 115 counties of the State of Missouri in the U.S. for the



period 1972-1981. Let  $y_{ias}$  define the lung cancer death count in cell  $(a, s)$  of county  $i$ , where  $a$  and  $s$  define age/sex groups. The model assumes,

$$\begin{aligned}
y_{ias} \mid \lambda_{ias}^{ind} &\sim \text{Poisson}(\lambda_{ias}) \\
\log(\lambda_{ias}) &= \log(rN_{ias}) + x_s\beta_1 + x_a\beta_2 + x_sx_a\beta_3 + u_i + \varepsilon_{ias} \\
u_i \mid \{u_k, k \neq i\} &\sim N(\bar{u}_{-i}, \sigma_u^2 / m_i) ; \varepsilon_{ias} \sim N(0, \sigma_\varepsilon^2) \\
\beta &\sim \text{Uniform}(R^3) ; (\sigma_u^2)^{-1} \sim \text{gamma}(.01, .01) ; (\sigma_\varepsilon^2)^{-1} \sim \text{gamma}(1, 1)
\end{aligned} \tag{4.5}$$

where  $N_{ias}$  is the mid-period population size in cell  $ias$ ,  $r$  is the statewide lung cancer rate,  $x_s$  is a sex indicator variable and  $x_a$  defines three age groups taking the values  $(-1, 0, 1)$ . The model for the area effects  $u_i$  accounts for spatial clustering by defining the means  $\bar{u}_{-i}$  to be the average of the random effects  $\{u_k, k \neq i\}$  that are “neighbors” of  $u_i$ , with  $m_i$  denoting the number of these neighbors. (Two counties are defined as neighbors if and only if they are physically adjacent to each other.) The distribution postulated for the random effects has a ‘Conditional Autoregressive structure’; see Clayton and Kaldor (1987) for discussion and uses of this kind of model. As implied by the second and third equations, the model dependent estimate of  $\lambda_{ias}$  borrows strength from other area estimates and from estimates of other cells within the same area. Rao (1999) reviews several new studies with interesting variations of the model (4.4).

## 5 Time series models

The models and estimators considered so far borrow strength from administrative data and neighboring areas. Another valuable source of information when available is data measured for the target characteristics on previous occasions. The (direct) estimators obtained from the different surveys are usually correlated even when independent samples are selected on different occasions because of the correlations among the true area characteristics over time, giving rise to time series modeling. A typical time series model fitted to survey data consists of two parts:

- A model fitted to the population (area) quantity of interest,
- A model fitted to the sampling errors.

Consider the following general class of *state-space* models for a single area  $i$  with the index  $t$  designating time. Below we consider time series models that borrow also strength across areas.

$$\begin{aligned} y_{it} &= \theta_{it} + e_{it} = x_{it}' \beta_{it} + e_{it} \\ \beta_{it} &= T_t \beta_{t-1,i} + \eta_{it} ; e_{it} \sim ARMA(a,b) \end{aligned} \quad (5.1)$$

where  $\beta_{it}(p \times 1)$  is a random *state vector*,  $T_t(p \times p)$  is a fixed *transition matrix* and  $e_{it}$  and  $\eta_{it} = (\eta_{it1} \dots \eta_{itp})'$  are independent random errors with  $Var(e_{it}) = \sigma^2$  and  $V(\eta_{it}) = Q$  respectively. It is assumed also that  $E(\eta_{it} \eta_{t-j,i}') = 0$  for  $j > 0$ . In (5.1),  $y_{it}$  is the (direct) sample estimate for area  $i$  at time  $t$  and  $\theta_{it} = x_{it}' \beta_{it}$  is the target quantity, modeled as a linear combination of known concomitant variables with random coefficients so that  $e_{it} = (y_{it} - \theta_{it})$  is the sampling error. The notation  $ARMA(a,b)$  defines the Auto-Regressive Moving Average model of order  $(a,b)$ , (Box and Jenkins 1976). As can be seen, the model (5.1) accounts for the time series relationships between the true area quantities via the model postulated for the state vectors and for the autocorrelations between the sampling errors. Note in this regard that the sampling errors may be correlated even when there is no sample overlap. This is so because in repeated surveys it is often the case that units joining the sample are from the same small geographical areas (like census tracts) as units leaving the sample.

The following models can be represented in the form (5.1), see Pfeiffermann and Burck (1990) for details.

1- the *area level random effects* model (3.3); the model defined by (5.1) can easily be extended to the case where individual measurements  $y_{itj}$  with corresponding covariate values  $x_{itj}$  are available, in which case the model contains also the *unit level regression* model (3.1) as a special case.

2- the *random coefficient regression* model,  $\beta_{itk} = \beta_k + \eta_{itk}$ . This is one of the models for which Prasad and Rao (1990) developed the EBLUP MSE bias corrections. By imposing  $Var(\eta_{itj}) = 0$  for some  $j$ , the corresponding regression coefficient is set fixed over time.

3- the *first order autoregression* model,  $(\beta_{tik} - \beta_k) = \phi_k (\beta_{t-1,ik} - \beta_k) + \eta_{tik}$ . By setting  $\phi_k = 1$ , the model simplifies to the *random walk* model.

4- As last example consider the model used in the U.S. for the production of all the major state employment and unemployment statistics. The model is fitted separately for each of the 50 states and the District of Columbia, see Tiller (1992) for details. In what follows  $y_{ti}$  is the direct estimate (say, the observed unemployment rate) in state  $i$  at month  $t$ ,  $L_t$  is a trend component with (random) slope  $R_t$  and  $S_t$  is a seasonal effect.

Observation equation

$$y_{ti} = x_{ti}\alpha_{ti} + L_{ti} + S_{ti} + e_{ti} ; e_{ti} \sim AR(15) \quad (5.2a)$$

Model for regression coefficients

$$\alpha_{ti} = \alpha_{t-1,i} + \eta_{\alpha ti} \quad (5.2b)$$

Model for Trend component

$$L_{ti} = L_{t-1,i} + R_{t-1,i} + \eta_{Lti} ; R_{ti} = R_{t-1,i} + \eta_{Rti} \quad (5.2c)$$

Model for Seasonal effects

$$\begin{aligned} S_{jti} &= \cos \omega_j S_{j,t-1,i} + \sin \omega_j S_{j,t-1,i}^* + \eta_{Sjti} \\ S_{jti}^* &= -\sin \omega_j S_{j,t-1,i} + \cos \omega_j S_{j,t-1,i}^* + \eta_{Sjti}^* \quad S_{ti} = \sum_{j=1}^6 S_{jti} \\ \omega_j &= 2\pi j/12, j = 1, \dots, 6 \end{aligned} \quad (5.2d)$$

The error terms  $\eta_{\alpha ti}, \eta_{Lti}, \eta_{Rti}, \eta_{Sjti}, \eta_{Sjti}^*$  are independent white noise. Note that the observed series consists of only the direct estimates  $y_{ti}$ . The model postulated for the trend is a local approximation to a linear trend. The AR(15) model accounts for the short and long term autocorrelations between the sampling errors as implied by the particular rotating panel sampling scheme underlying the U.S. Labor Force Survey. The covariate  $x_{ti}$  represents the ‘number of unemployed workers claiming unemployment insurance’. Pfeiffermann *et al.* (1998a) fit a similar model for the production of labor force statistics and trend estimates in small regions of Australia. Binder and Dick (1989) and Bell and Hillmer (1990) fit Multiplicative Seasonal ARIMA models (Box and Jenkins 1976) to the “regression residuals”  $(\theta_{ti} - x_{ti}'\gamma)$  where  $\theta_{ti}$  defines as before the true area quantity of interest and  $x_{ti}$  represents area level covariate variables. The fitting of these models requires differencing of the observed series to guarantee stationarity. The area estimates are obtained by signal extraction.

The model (5.1) contains in general as unknown hyper-parameters the variances and covariances in  $Q$ , the parameters of the ARMA model of the sampling error and possibly also some of the elements of the transition matrices  $T_t$ . Because of possible identification problems and in order to simplify the maximization of the likelihood, it is customary to estimate the ARMA parameters based on external estimates of the variance and autocorrelations of the sampling error. Pfeiffermann *et al.* (1998) develop a simple method of estimating the sampling error autocorrelations for rotating panel sampling designs. The other parameters are estimated by the values maximizing the likelihood when fixing the ARMA parameters at their estimated values. The likelihood function is conveniently obtained by use of the *Kalman filter*, which for known hyper-parameters yields the BLUP of the state vector and the corresponding prediction error variance-covariance matrix for every time  $t$ . See Harvey (1989) for details.

Like with cross-sectional models, an important issue when fitting time series models with estimated parameters is how to obtain reliable estimates for the MSE of the state vector predictors  $\hat{\beta}_{ii}$ , and hence for the MSE of the small area predictors  $\hat{\theta}_{ii} = x_{ii}' \hat{\beta}_{ii}$ . A “naive” MSE estimator is obtained by replacing the unknown parameters in the MSE expressions produced by the Kalman filter by the parameter estimates, but this again results in underestimation of the true MSE. Ansley and Kohn (1986) develop a bias correction to the naive MSE estimator by expanding the MLE around the true parameters. The bias correction is derived under the frequentist approach but it is shown to have also a Bayesian interpretation. Hamilton (1986) adopts the Bayesian perspective and proposes a Monte Carlo method that consists of generating samples from a normal approximation to the posterior distribution of the unknown parameters. The normal approximation is obtained by fixing the mean at the MLE (or REML) and the covariance matrix at the inverse information matrix evaluated at the MLE. Quenneville and Singh (2000) show that both approaches have a bias of order  $O(1/n)$  where  $n$  is the length of the series. The authors propose a second order approximation that reduces the bias to the order of  $o(1/n)$ . This approximation again assumes normality of the posterior distribution of the parameters and requires an evaluation of the information matrix. Pfeiffermann and Tiller (2000) propose a simple parametric bootstrap method that yields frequentist MSE estimates with bias of order  $O(1/n^2)$ . The use of

this approach does not assume normality of the unknown parameters or their estimators and it does not require evaluation of the inverse information matrix. This matrix is often unstable, particularly when the number of unknown parameters is large.

Evaluation of the prediction MSE of the small area predictors  $\hat{\theta}_{ii} = x_{ii}'\hat{\beta}_{ii}$  creates no extra problem under the full HB modeling approach because the prediction MSE is defined by the posterior variance. The use of this approach requires however the specification of prior distributions for all the unknown hyper-parameters and as mentioned before, the computations can be very intensive; see, for example, Ghosh *et al.* (1996) and Datta *et al.* (1999). (The latter study also considers estimation of unemployment rates in states of the U.S. but uses a different model from the model defined by (5.2). In particular, this model permits cross-sectional correlations between the state unemployment rates  $\theta_{ii}$ , see also below.)

The models considered so far pertain to time series observed in a single area. In practice, data are often available for many small areas simultaneously, although possibly for only few time points. In such cases it is desired to borrow information both cross-sectionally and over time. The combined modeling of cross-sectional and time series data is a classical problem in econometrics; see, e.g., Johnson (1977, 1980) for annotated bibliographies, but the econometric literature does not address SAE problems. One way of borrowing information cross-sectionally and over time within the framework of the model (5.1) is by including in the model both contemporary random effects and time varying effects. Rao and Yu (1994) consider the model

$$y_{ii} = \theta_{ii} + e_{ii} = [x_{ii}'\beta + u_i + v_{ii}] + e_{ii} \quad (5.3)$$

where, using previous notation,  $y_{ii}$  is the direct estimator for area  $i$  at time  $t$ ,  $e_{ii}$  is the sampling error (assumed to be correlated over time with known variances and covariances),  $u_i$  is the contemporary random effect and  $v_{ii}$  is a time effect, assumed to evolve as in an AR(1) process. The random errors  $u_i, v_{ii}, e_{ii}$  are independent within and between the areas for any given time point  $t$ . The authors derive the EBLUP and valid (correct order) MSE estimators. Datta *et al.* (1999) extend the model (5.3) by permitting random regression coefficients  $\beta_i$  instead of the fixed vector  $\beta$  and adding extra random terms to account for

seasonal variations. The authors develop an appropriate HB methodology for point and MSE estimation using the Gibbs sampler.

An alternative way of accounting for cross-sectional relationships under the model (5.1) is by permitting corresponding components of the error terms  $\eta_{ii}$  pertaining to different areas  $i$  to be correlated. Pfeiffermann and Burck (1990) derive the explicit expression of the small area predictor obtained this way for the case where the state vector is a random walk and the sampling errors are uncorrelated, illustrating how the time series and cross-sectional data combine to strengthen the direct estimator. The authors apply the model for predicting mean house sale prices in cells defined by size (number of rooms) and cities. Ghosh *et al.* (1996) consider a more extreme case by which the state vectors are the same across the areas ( $\beta_{ii} \equiv \beta_i \sim \text{random walk}$ ), and apply the HB methodology with appropriate prior distributions using the Gibbs sampler. The specific problem considered is the prediction of median income of four person families in states of the U.S. An interesting feature of this study is that the direct estimators  $y_{ii}$  are multivariate, containing also the median income estimates of three and five person families (with the other model components modified accordingly). Since the three contemporary direct estimates are correlated, including all of them in the model potentially improves the prediction compared to modeling of only the univariate estimators. In this particular application the best predictors were actually obtained when using a bivariate model with the direct median estimates of four and five person families as the input data.

## 6 Accounting for sampling effects

The models and inference procedures considered so far assume implicitly simple random sampling designs. In practice, the sample selection process often involves unequal selection probabilities at least at some stages of the selection process. When the selection probabilities are not related to the values of the response variable, the models holding for the population hold also for the sample data and the sample selection process can be ignored. If, however, the selection probabilities are related to the response variable values even after conditioning on the values of the explanatory variables included in the model, the sampling design becomes *informative* and the model holding for the sample data differs from the model holding in the population. Ignoring the sample selection in such cases may yield biased predictors for the target characteristics of interest.

As a simple (but extreme) example of an informative sampling scheme suppose that the outcomes  $y_{ij}$  in a given area  $i$  are binary with  $\Pr(y_{ij} = 1) = p$ . Let the sample be selected with probabilities,  $\Pr(j \in s_i | y_{ij} = k) = \pi_k$ ,  $k = 0, 1$ . Thence, by application of Bayes rule,  $\Pr(y_{ij} = 1 | j \in s_i) = \pi_1 p / [\pi_1 p + \pi_2 (1 - p)] = p_s$ . Clearly,  $p \neq p_s$  unless  $\pi_1 = \pi_2$  and inference on  $p$  that ignores the selection scheme actually means inference on  $p_s$ . For recent discussion of informative probability sampling with key references, see Pfeffermann *et al.* (1998b).

A common approach to account for possible sampling effects is to weight the sample measurements by the sampling weights, defined here as the inverse of the sample inclusion probabilities. (In practice the weights are often modified to account for missing data and post-stratification adjustments.) In the context of small area estimation, Kott (1989) and Prasad and Rao (1999) consider the simple ‘unit level random effect model’,  $y_{ij} = \mu + u_i + \varepsilon_{ij} = \theta_i + \varepsilon_{ij}$ , with the same assumptions on  $u_i$  and  $\varepsilon_{ij}$  as in (3.1), and propose to replace the simple means  $\bar{y}_i$  by the weighted means  $\bar{y}_{iw} = \sum_{j=1}^{n_i} w_{ij} y_{ij} / \sum_{j=1}^{n_i} w_{ij}$  as the input *direct* estimators where  $w_{ij} = 1/\pi_{ij}$ ;  $\pi_{ij} = P(j \in s_i)$ . The estimator  $\bar{y}_{iw}$  is approximately design unbiased (over repeated sampling) and consistent for the corresponding area mean  $\theta_i$ . The predictor  $\hat{\theta}_i$  in both articles has the form of a composite estimator,  $\hat{\theta}_i = \hat{\alpha}_i \bar{y}_{iw} + (1 - \hat{\alpha}_i) \hat{\mu}_w$ , where  $\hat{\mu}_w$  is a weighted average of the unweighted sample means  $\bar{y}_i$  in Kott (1989), and a weighted average of the weighted means  $\bar{y}_{iw}$  in Prasad and Rao (1999). The coefficients  $\hat{\alpha}_i$  and the vector coefficients defining  $\hat{\mu}_w$  are functions of the sampling weights and the unweighted ANOVA estimators of the model variances. Prasad and Rao (1999) derive MSE estimators of correct order under the assumption that the model holds for the sample data (implying *de facto* non-informative sampling). The authors extend the results to the unit level regression model defined by (3.1). Arora and Lahiri (1997) model the weighted area means in a HB analysis that uses the Gibbs sampler to obtain the posterior means and variances. The authors allow for different (unknown) variances  $k_i \sigma_i^2 = \text{Var}(\bar{y}_{iw} | \theta_i)$  in different areas, postulating the general prior  $\sigma_i^{-2} \sim \text{gamma}(a, b)$  (the  $k_i$ 's are known), but like in Prasad and Rao (1999) the variances are taken under the population model. An empirical investigation

with data collected in the U.S. family expenditure survey illustrates the better performance of the HB estimators compared to the EBLUP and the direct weighted means.

The use of the estimators  $\bar{y}_{iw}$  protects against informative sampling *within* the areas. Furthermore, as  $n_i$  increases, the weight  $\hat{\alpha}_i$  attached to the direct estimator increases so that the estimator  $\hat{\theta}_i$  is design consistent for  $\theta_i$ . Note also that by estimating  $\mu$  as a weighted average of the weighted means  $\bar{y}_{iw}$  like in Prasad and Rao (1999), the estimator  $\hat{\theta}_i$  is approximately unbiased for  $\theta_i$  if the expectation is taken over both the randomization distribution and the model. The estimator is neither approximately design unbiased nor approximately model unbiased under informative sampling. Folsom *et al.* (1999) incorporate the sampling weights for defining the conditional posterior densities of the fixed and random effects in a logistic mixed model (extension of the model 4.1). The resulting densities are employed for a survey weighted full HB analysis using the Gibbs sampler. The model is used for estimating illicit drug use in the U.S. (first example in Section 1).

Weighting the sample measurements by the sampling weights does not protect against informative selection of the areas when only some of the areas are selected to the sample. Note in this regard that under the models (3.1) and (3.3) considered in section 3, the EBLUP or EB predictors for areas not included in the sample are the synthetic estimators  $\bar{X}_i' \hat{\beta}_{GLS}$  and  $x_i' \hat{\beta}_{GLS}$  respectively. These estimators may be severely biased if the selection of the areas is informative, since it is no longer necessarily true that  $E(u_i | i \notin s) = 0$ , a condition validating the use of the synthetic estimators under noninformative sampling. For the model (3.1) (with possibly more than one random effect), Pfeffermann *et al.* (1998c) propose a weighting system that yields consistent estimators for all the model parameters, but that article does not address the prediction of small area means.

A different way of protecting against informative sampling is to base the inference on the *sample* probability density function (pdf), defined for *single stage* sampling as,

$$f_s(y_k | x_k) = f(y_k | x_k, k \in s) = \Pr(k \in s | y_k, x_k) f_p(y_k | x_k) / \Pr(k \in s | x_k) \quad (6.1)$$

where  $f_p(y_k | x_k)$  defines the population pdf (Pfeffermann *et al.* 1998b). Denote the sample inclusion probabilities by  $\pi_k = \Pr(k \in s)$ . These probabilities depend in general on the all the



population values of  $y$  and  $x$ , and possibly also on the population values of design variables used for the sample selection but not included in the model. By viewing the population measurements as random realizations under the model, the probabilities  $\pi_k$  are likewise random and the pdf (6.1) can be written alternatively as,

$$f_s(y_k | x_k) = E_p(\pi_k | y_k, x_k) f_p(y_k | x_k) / E_p(\pi_k | x_k). \quad (6.2)$$

Note that  $E_p(\pi_k | y_k, x_k)$  is generally not the same as  $\pi_k = \Pr(k \in s)$ , which may depend on all the population measurements. It follows from (6.2) that for a given population pdf, the sample pdf is fully specified by the expectation  $E_p(\pi_k | y_k, x_k) = 1 / E_s(w_k | y_k, x_k)$ , where  $E_s$  defines the expectation with respect to the sample distribution. The sample expectation can be identified and estimated from the sample data and knowledge of the sampling design and, see Pfeffermann and Sverchkov (1999) for discussion and examples. Furthermore, Pfeffermann *et al.* (1998b) establish that for independent population measurements, the sample measurements are asymptotically independent with respect to the sample distribution under commonly used sampling schemes for selection with unequal probabilities. The asymptotic assumes  $N \rightarrow \infty$  with  $n$  held fixed. ( $N$  and  $n$  are the population and sample sizes respectively.)

How can the sample distribution be used for SAE? Consider for convenience the unit level random effect model  $y_{ij} = \mu + u_i + \varepsilon_{ij} = \theta_i + \varepsilon_{ij}$  mentioned earlier in this section. The corresponding sample model is,

$$f_s(y_{ij} | \theta_i) = E_p(\pi_{j|i} | y_{ij}, \theta_i) f_p(y_{ij} | \theta_i) / E_p(\pi_{j|i} | \theta_i) \quad (6.3a)$$

$$f_s(\theta_i) = E_p(\pi_i | \theta_i) f_p(\theta_i) / E_p(\pi_i) \quad (6.3b)$$

where  $\pi_{j|i}$  is the conditional sample selection probability given that area  $i$  is in the sample,  $\pi_i = \Pr(i \in s)$  and  $f_p(y_{ij} | \theta_i)$  and  $f_p(\theta_i)$  are the corresponding population densities. By modelling the expectations defining the sample distributions in (6.3a) and (6.3b), one can obtain the ‘model dependent’ predictors of the small area means, employing the methods discussed in previous sections. As a simple illustration suppose that the selection of the areas is carried out with probabilities  $\pi_i$  such that  $E(\pi_i | \theta_i) = \Pr(i \in s) = k \times \exp(a_0 + a_1 \theta_i)$  where

$k$  is some constant and that the sampling process within the areas is noninformative. Simple calculations show that for known variances  $(\sigma_u^2, \sigma_\varepsilon^2)$  the optimal (BLUP or HB) predictors of the area means for areas in the sample and areas not in the sample are correspondingly,

$$\begin{aligned}\hat{\theta}_{s,i} &= E(\theta_i | i \in s, y) = \hat{\theta}_i + a_1 \sigma_\varepsilon^2 \gamma_i \\ \hat{\theta}_{c,i} &= E(\theta_i | i \notin s, y) = \hat{\theta}_i - \frac{a_1 \sigma_\varepsilon^2 \gamma_i k \exp(a_0 + d_i)}{1 - k \exp(a_0 + d_i)}\end{aligned}\quad (6.4)$$

where  $y$  denotes the sample data,  $d_i = a_1 \hat{\theta}_i + (a_1^2 / 2) \sigma_\varepsilon^2 \gamma_i$  and  $\hat{\theta}_i = \gamma_i \bar{y}_i + (1 - \gamma_i) \bar{y}$  is the optimal predictor under noninformative selection of the areas (see Section 3). As clearly indicated by (6.4), the optimal predictors for areas in the sample are higher than the optimal predictors under noninformative sampling since under the assumed selection process the selected areas tend to have larger means. Conversely, the optimal predictors for areas not in the sample are lower than the optimal predictors under noninformative sampling. Thus, the use of the predictors  $\hat{\theta}_i$  that ignore the sampling process yields biased prediction. We are currently studying the application of this approach in conjunction with the model (3.1) using data from the Brazilian “Basic Education Evaluation Study”.

*Comment:* Application of the EB or HB approaches under informative sampling does not require changing the prior distributions since they refer to the population model. Ghosh (1999) explores the properties of the EB predictor obtained by use of the sample distribution in the case of the Fay-Herriot model (3.3).

## 7 The importance of modeling the correlations among random area effects

Many of the models in common use for SAE assume that the random small area effects are independent between the areas. In practice, it would often be reasonable to assume that the random effects associated with ‘neighboring areas’ by some distance measure (not necessarily geographical) are correlated, with the correlations decaying to zero as the distance increases. Such models are very common in *spatial analysis* (see, e.g., Cressie 1993), but are not in wide use for SAE problems. As few exceptions we mention the articles by Datta and Ghosh (1991), Datta *et al.* (1999), Ghosh *et al.* (1996), Ghosh *et al.* (1998), Rao and Yu (1994) and Pfeiffermann and Burck (1990) that have been reviewed in previous sections.

How much can be gained by accounting for existing correlations among the area effects? In order to study this question, we consider two simple examples. In the first example we let the ‘*working model*’ be defined by (3.3) but with no concomitant variables and with the added simplification of equal sample sizes,  $n_i = n$ . Thus, the working model is,

$$\bar{y}_i = \mu + u_i + e_i = \theta_i + e_i ; \text{Var}_D(e_i) = \sigma_e^2 / n = \sigma^2 ; \text{Var}(u_i) = \sigma_u^2 \quad (7.1)$$

with  $(u_i, e_i)$  being independent within and between the areas. Let  $\sigma^2$  and  $\sigma_u^2$  be known but  $\mu$  unknown. Under this model, the optimal predictor of the small area mean is,

$$\hat{\theta}_i = \gamma \bar{y}_i + (1 - \gamma) \bar{y} ; \quad \bar{y} = \sum_{i=1}^m \bar{y}_i / m ; \quad \gamma = \sigma_u^2 / (\sigma_u^2 + \sigma^2). \quad (7.2)$$

Suppose now that under the ‘*correct model*’,  $\text{Corr}(u_i, u_k) = \rho > 0$  for  $i \neq k$ . Simple calculations show that for this model the optimal predictor and its MSE are,

$$\theta_i^* = \tilde{\gamma} \bar{y}_i + (1 - \tilde{\gamma}) \bar{y} ; \quad \tilde{\gamma} = [\sigma_u^2 (1 - \rho)] / [\sigma_u^2 (1 - \rho) + \sigma^2] \quad (7.3)$$

$$\begin{aligned} \text{MSE}(\theta_i^*) &= \tilde{\gamma}^2 \delta^2 + (1 - \tilde{\gamma})^2 [\delta^2 + (m - 1) \rho \sigma_u^2] / m \\ &\quad + (1 - 2\tilde{\gamma}) \sigma_u^2 - 2(1 - \tilde{\gamma}) [\sigma_u^2 + (m - 1) \rho \sigma_u^2] / m \end{aligned} \quad (7.4)$$

where  $\delta^2 = (\sigma_u^2 + \sigma^2) = \text{Var}(\bar{y}_i)$ . The ‘working predictor’  $\hat{\theta}_i$  is unbiased under the correct model and its MSE is obtained from (7.4) by replacing  $\tilde{\gamma}$  by  $\gamma$ .

The question arising is: what is the loss from using  $\hat{\theta}_i$  instead of  $\theta_i^*$ ? We mention first that  $\lim_{m \rightarrow \infty} \text{MSE}(\theta_i^*) / \text{MSE}(\hat{\theta}_i) = [\tilde{\gamma} / \gamma] / [1 - \rho(1 - \gamma)]$  with the limit decreasing to 0 as  $\rho$  increases to 1, showing that with many areas and large cross-sectional correlations, the loss in efficiency can be substantial. Table 1 shows the relative efficiencies  $R = \text{MSE}(\theta_i^*) / (\text{MSE} \hat{\theta}_i)$  for selected numbers of areas and different values of  $\rho$  and  $L = (\sigma^2 / \sigma_u^2)$ .

**Table 1: Relative efficiencies of working predictor compared with the optimal predictor under equal correlations model**

		L=0.1	L=1.0	L=10	L=100
m=5	$\rho = .25$	1.00	0.99	1.00	1.00
	$\rho = .50$	0.97	0.93	0.99	1.00
	$\rho = .75$	0.88	0.80	0.98	1.00
m=40	$\rho = .25$	0.99	0.98	1.00	1.00
	$\rho = .50$	0.96	0.90	0.97	1.00
	$\rho = .75$	0.85	0.67	0.92	1.00
m=100	$\rho = .25$	0.99	0.98	0.99	1.00
	$\rho = .50$	0.96	0.89	0.97	1.00
	$\rho = .75$	0.85	0.65	0.88	1.00

For the second example we use the same working model but let  $Corr(u_i, u_k) = \rho^{|i-k|}$  to represent the true correlations among the random area effects. This correlation pattern corresponds to the case where the means  $\bar{y}_i$  refer to different time points and the area effects follow the AR(1) model,

$$u_i = \rho u_{i-1} + \varepsilon_i, \quad Var(\varepsilon_i) = \sigma_\varepsilon^2 \Rightarrow \sigma_u^2 = \sigma_\varepsilon^2 / (1 - \rho^2). \quad (7.5)$$

If  $i = m$  signifies the last time point with observations, the optimal predictor of  $\theta_m$  under this model is easily obtained by application of the recursive Kalman filter equations (Harvey 1989). For ‘large  $m$ ’ the MSE of the optimal predictor can be computed explicitly by solving the corresponding Riccati equation. We find that

$$MSE(\theta_m^*) = \{[(1 - \rho^2)^2 \delta^4 + 4\sigma_u^2 \sigma^2 \rho^2 (1 - \rho^2)]^{1/2} - (1 - \rho^2) \delta^2\} / 2\rho^2. \quad (7.6)$$

As in the previous example, the working predictor  $\hat{\theta}_i$  defined by (7.2) is unbiased under the model (7.5) and as  $m \rightarrow \infty$ ,  $MSE(\hat{\theta}_i) \rightarrow \gamma \sigma^2$  under both the working model and the true model. Table 2 shows the relative efficiencies  $R = MSE(\theta_i^*) / (MSE \hat{\theta}_i)$  for large  $m$  and different values of  $\rho$  and  $L = (\sigma^2 / \sigma_u^2)$ .

**Table 2: Relative efficiencies of working predictor compared with optimal predictor under Auto-regression model, large m**

	L=0.1	L=1	L=10	L=100
--	-------	-----	------	-------

$\rho = 0.25$	0.99	0.98	0.99	1.00
$\rho = 0.50$	0.97	0.93	0.97	1.00
$\rho = 0.75$	0.91	0.80	0.91	0.99
$\rho = 0.95$	0.66	0.48	0.66	0.93

The first notable (but expected) result emerging from the two tables is that for fixed ratios  $L$ ,  $R$  decreases as  $\rho$  increases, implying that the loss from using the working predictor increases. Yet, unless  $\rho > 0.5$ ,  $R \geq 0.90$ . Another less expected result is that increasing  $L$  does not necessarily imply a corresponding decrease in  $R$ . In fact, in both tables  $R$  decreases as  $L$  increases from 0.1 to 1 but then it increases when further increasing  $L$ . (It is easily shown that for the AR(1) model and for the limiting case  $m \rightarrow \infty$  under the equal correlations model,  $R(L) = R(1/L)$  for all  $L$ , with the minimum value of  $R$  attained at  $L=1$  for which  $\gamma=0.5$ .) This result is explained by the fact that as  $\sigma^2$  increases in relation to  $\sigma_u^2$ , more weight is assigned to the synthetic estimator  $\bar{y}$  which is common to both the working and the true model so that the differences between the MSE of the two predictors diminish. The results presented in Table 2 for the AR(1) model are ‘one sided’ in the sense that the ‘borrowing of strength’ is only from previous occasions. We studied also the ‘two sided’ case where data collected before and after the time point of interest are used for deriving the optimal predictor and obtained very similar results.

The overall conclusion from this study is that unless the correlations between the area random effects are large, the loss in efficiency from using the working model is small. Notice also that for small values of  $\rho$  the MSE of the working predictor  $\hat{\theta}_i$  under the correct model is similar to the MSE evaluated under the working model. Clearly, the results of this rather limited study need to be tested under different models.

## 8 Concluding remarks

This article attempts to overview the main topics in SAE research in recent years, emphasizing the new models and inference methods with particular attention to point estimation and MSE evaluations. Two important issues not considered are model selection and model diagnostics. As mentioned before, SAE is one of the few fields in survey sampling

where it is widely recognized that the use of model dependent inference is often inevitable. Given the growing use of small area statistics and their immense importance, it is imperative to develop efficient tools for the selection of models and the ascertainment of their goodness of fit. This is a difficult problem because small area models contain assumptions on unobservable random effects, which are therefore difficult to verify. We mention also that under informative sampling discussed in Section 6, the observed data no longer represent the population model, making the model selection and diagnostics even harder. Ghosh and Rao (1994) discuss a few formal and informal diagnostic tests that have been developed until that year. Recently, Jiang, Lahiri and Wu (1999) developed a ‘chi-square statistic’ for testing the distribution of the combined error components. A third related issue is how to secure the robustness of the estimators after a model and inference method have been selected. Several sensitivity analyses are reported in the literature but more theoretical research is needed for studying the robustness of the SAE methods in common use. With the extensive research on SAE taking place in so many countries, I expect many new innovative studies on these and other issues in the near future.

## References

- Arora, V., and Lahiri, P. (1997). On the superiority of the Bayesian method over the BLUP in small area estimation problems. *Statistica Sinica*, **7**, 1053-1063.
- Ansley, C.F., and Kohn, R. (1986). Prediction mean squared error for state space models with estimated parameters. *Biometrika*, **73**, 467-473.
- 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. Am. Statist. Ass.*, **83**, 28-36.
- Bell, W.R., and Hillmer, S.C. (1990). The time series approach to estimation for repeated surveys. *Survey Methodology*, **16**, 195-215.
- Binder, D.A., and Dick, J.P. (1989). Modeling and estimation for repeated surveys. *Survey Methodology*, **15**, 29-45.
- Box, G.E.P., and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, 2<sup>nd</sup> ed. San Francisco: Holden-Day.
- Clayton, D.G., and Kaldor, J. (1987). Empirical Bayes estimates of age standardized relative risks for use in disease mapping. *Biometrics*, **43**, 671-681.

- Cressie, N. (1993). *Statistics for Spatial Data*. New York: Wiley.
- Datta, G.S., and Ghosh, M. (1991). Bayesian prediction in linear models: Application to small area estimation, *Ann. Statist.*, **19**, 1748-1770.
- Datta, G.S., Lahiri, P., Maiti, T., and Lu, K.L. (1999). Hierarchical Bayes estimation of unemployment rates for the states of the U.S. *J. Am. Statist. Ass.*, **94**, 1074-1082.
- Datta, G.S., and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small-area estimation problems. *Statistica Sinica*, **10**, 613-627.
- Farrell, P.J., MacGibbon, B., and Tomberlin, T.J. (1997). Empirical Bayes small-area estimation using logistic regression models and summary statistics. *J. Bus. Econ. Statist.* **15**, 101-108.
- Fay, R. E., and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data, *J. Am. Statist. Ass.*, **74**, 269-277.
- Folsom, R., Shah, B., and Vaish, A. (1999). Substance abuse in states: A methodological report on model based estimates from the 1994-1996 national household surveys on drug abuse. In *Proceedings of the Section on Survey Research methods*, American Statistical Association, pp. 371-375.
- Gelfand, A.E., and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Statist. Ass.*, **85**, 398-409.
- Ghosh, M. (1999). Some recent results on empirical Bayes methods for small area estimation. In *Proceedings of conference on small area estimation*, pp. 51-64, Riga, Latvia.
- Ghosh, M., and Rao, J.N.K. (1994). Small area estimation: an appraisal (with discussion). *Statistical Science*, **9**, 65-93.
- Ghosh, M., Nangia, N., and Kim, D. H. (1996). Estimation of median income of four-person families: a Bayesian time series approach. *J. Am. Statist. Ass.*, **91**, 1423-1431.
- Ghosh, M., Natarajan, K., Stroud, T.W.F., and Carlin, B. (1998). Generalized linear models for small-area estimation. *J. Am. Statist. Ass.*, **93**, 273-282.
- Hamilton, J. D. (1986). A standard error for the estimated state vector of a state space model. *Journal of Econometrics*, **33**, 387-397.
- Harter, R., Wolter, K., and Macalsu, M. (1999). Small domain estimation of employment using CES and ES202 data. Statistical Policy Working Paper 30, Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget, Washington DC.

- Harvey, A. C. (1989). *Forecasting Structural Time Series with the Kalman Filter*. Cambridge: Cambridge University Press.
- Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a linear model. *Biometrics*, **31**, 423-448.
- Jiang, J., and Lahiri, P. (1998). Empirical best prediction for small area inference with binary data. Technical report, Department of Mathematics and Statistics, University of Nebraska.
- Jiang, J., Lahiri, P., and Wu, C.H. (1999). On Pearson- $\chi^2$  testing with unobservable cell frequencies and mixed model diagnostics. Technical report, Department of Statistics, Case Western Reserve University.
- Johnson, L.W. (1977). Stochastic parameter regressions: An annotated bibliography. *Inter. Statist. Rev.*, **45**, 257-272.
- Johnson, L.W. (1980). Stochastic parameter regressions: An additional annotated bibliography, *Inter. Statist. Rev.*, **48**, 95-102.
- Kackar, R.N., and Harville, D.A. (1984). Approximations for standard errors of estimators for fixed and random effects in mixed models. *J. Am. Statist. Ass.*, **79**, 853-862.
- Kass, R. E., and Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *J. Am. Statist. Ass.*, **84**, 717-726.
- Kott, P. (1989). Robust small domain estimation using random effects modelling. *Survey Methodology*, **15**, 1-12.
- Lahiri, P., and Rao, J.N.K. (1995). Robust estimation of mean square error of small area predictors. *J. Am. Statist. Ass.*, **90**, 758-766.
- Laird, N.M., and Louis, T. A. (1987). Empirical Bayes confidence intervals based on bootstrap samples. *J. Am. Statist. Ass.*, **82**, 739-750.
- MacGibbon, B., and Tomberlin, T.J. (1989). Small area estimates of proportions via empirical Bayes techniques. *Survey Methodology*, **15**, 237-252.
- Malec, D., Sedransk, J., Moriarity, C. L., and LeClere, F.B. (1997). Small area inference for binary variables in the national health interview survey. *J. Am. Statist. Ass.*, **92**, 815-826.
- Marker, D.A. (1999). Organization of small area estimators using a generalized linear regression framework. *Journal of Official Statistics*, **15**, 1-24.
- McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models* 2nd ed. London: Chapman and Hall.



- Moura, F.A.S., and Holt, D. (1999). Small area estimation using multilevel models. *Survey Methodology*, **25**, 73-80.
- Quenneville, B., and Singh, A.C. (2000). Bayesian prediction mean square error for state space models with estimated parameters. *J. Time Series Anal.* **21**, 219-236.
- Pfeffermann, D., and Burck, L. (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methodology*, **16**, 217-237.
- Pfeffermann, D., Feder, M., and Signorelli, D. (1998a). Estimation of autocorrelations of survey errors with application to trend estimation in small areas. *J. Bus. Econ. Statist.*, **16**, 339-348.
- Pfeffermann, D., Krieger, M.A, and Rinott, Y. (1998b). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica*, **8**, 1087-1114.
- Pfeffermann, D., Skinner, C.J., Holmes, D., Goldstein, H., and Rasbash, J. (1998c). Weighting for unequal selection probabilities in multilevel models, *J. Roy. Statist. Soc. B*, **60**, 23-40.
- Pfeffermann, D., and Sverchkov, M. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhya B*, **61**, 166-186.
- Pfeffermann, D. and Tiller, R.B. (2000). Variance estimation of seasonally adjusted and trend estimators accounting for survey error autocorrelations. Technical report, Department of Statistics, Hebrew University.
- Prasad, N.G.N., and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *J. Am. Statist. Ass.*, **85**, 163-171.
- Prasad, N.G.N., and Rao, J.N.K. (1999). On robust small area estimation using a simple random effects model. *Survey Methodology*, **25**, 67-72.
- Rao, J.N.K., and Yu, M. (1994). Small area estimation by combining time series and cross-sectional data. *Can. J. Statist.*, **22**, 511-528.
- Rao, J.N.K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, **25**, 175-186.
- Singh, A.C., Stukel, D. M., and Pfeffermann, D. (1998). Bayesian versus frequentist measures of error in small area estimation. *J. Roy. Statist. Soc. B*, **60**, 377-396.
- Stukel, D. M., and Rao, J.N.K. (1999). On small area estimation under two-fold nested error regression models. *J. Statist. Plan. Infer.*, **78**, 131-147.

Tiller, R.B. (1992). Time series modeling of sample survey data from the U.S. Current Population Survey. *Journal of Official Statistics*, **8**, 149-166.

Thomsen, I., and Holmoy, A.M.K. (1998). Combining data from surveys and administrative record systems. The Norwegian experience, *Inter. Statist. Rev.*, **66**, 201-221.