UNIVERSITY OF Southampton
Southampton Statistical Sciences Research Institute

Working Paper M10/11

Methodology

New Important Developments In Small Area Estimation

Danny Pfeffermann

Abstract

The purpose of this paper is to review and discuss some of the new important developments in small area estimation (SAE) methods. Rao (2003) wrote a very comprehensive book, which covers all the main developments in this topic until that time and so the focus of this review is on new developments in the last 7 years. However, to make the review more self contained, I also repeat shortly some of the older developments. The review covers both design-based and model-dependent methods with emphasis on the prediction of the area target quantities and the assessment of the prediction error. The style of the paper is similar to the style of my previous review on SAE published in 2002, explaining the new problems investigated and describing the proposed solutions, but without dwelling on theoretical details, which can be found in the original articles. I am hoping that this paper will be useful both to researchers who like to learn more on the research carried out in SAE and to practitioners who might be interested in the application of the new methods.

New Important Developments in Small Area Estimation

Danny Pfeffermann

Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, SO17 1BJ, UK

Department of Statistics, Hebrew University of Jerusalem, Jerusalem, 91905, Israel

msdanny@soton.ac.il

Abstract

The purpose of this paper is to review and discuss some of the new important developments in small area estimation (SAE) methods. Rao (2003) wrote a very comprehensive book, which covers all the main developments in this topic until that time and so the focus of this review is on new developments in the last 7 years. However, to make the review more self contained, I also repeat shortly some of the older developments. The review covers both design-based and model-dependent methods with emphasis on the prediction of the area target quantities and the assessment of the prediction error. The style of the paper is similar to the style of my previous review on SAE published in 2002, explaining the new problems investigated and describing the proposed solutions, but without dwelling on theoretical details, which can be found in the original articles. I am hoping that this paper will be useful both to researchers who like to learn more on the research carried out in SAE and to practitioners who might be interested in the application of the new methods.

Key words: Benchmarking, Calibration, Confidence intervals, Errors in variables, Fence method, Informative sampling, matching priors, M-quantiles, Ordered means, Outliers, Prediction MSE, Spline, Two part model.

1. PREFACE

In 2002 I published a review paper in the *International Statistical Review* with a similar title. At that year small area estimation (SAE) was flourishing both in research and applications, but my own feeling in those days was that the topic has been more or less exhausted in terms of research and that it will just turn into a routine application in sample survey practice. As the past 7 years show, I was completely wrong and not only that research in this area is accelerating, it now involves some of the best known statisticians, who otherwise are not involved in survey sampling theory or applications. The diversity of new problems investigated is overwhelming, and the solutions proposed are not only elegant and innovative, but also very practical.

Rao (2003) published a comprehensive book on SAE that covers all the main developments in this topic until that time. Since SAE is so broadly applied, I thought that the time is ripe for a new critical review that focuses on some of the main developments in the last 7 years or so that I became aware of. The style of the paper is similar to the style of my previous review, explaining the new problems investigated and describing the proposed solutions, but without dwelling on theoretical details, which can be found in the original articles. For further clarity and to make this paper self contained, I provide a short background and overview some of the 'older' developments. I am hoping that this paper will be useful both to researchers who like to learn more on the research carried out in SAE and to practitioners who might be interested in applying the new methods.

2. SOME BACKGROUND

The problem of SAE is how to produce reliable estimates of characteristics of interest such as means, counts, quantiles, etc., for small areas or domains for which only small samples or no samples are available. The latter case of no samples requires the use of statistical models that define how to borrow information from neighboring areas or over time. Although the point estimators are usually of first priority, a related problem is how to assess the estimation (prediction) error. The great importance of SAE stems from the fact that many new policies such as fund allocation for needed areas, new educational or health programmes and environmental planning rely heavily on these estimates. Small area estimation techniques are also used in many countries to test and adjust the counts obtained from censuses that use administrative records.

The use of the term ‘SAE’ is a bit confusing, since it is the size of the sample in the area that creates the problem and not the size of the area. Also, the term ‘areas’ does not necessarily refer to geographical districts and may define another grouping, such as socio-demographic groups or types of industry, in which case they are often referred to as domains. Closely related terms in common use are ‘poverty mapping’ or ‘disease mapping’, which amount to SAE of poverty measures or disease incidence and then presenting the results on a map, with different colors defining different levels (categories) of the estimators. What is common to most small area estimation problems is that the point estimators and their error measures are required for every area separately, and not just as an average over all the areas under consideration.

SAE methods can be divided broadly into ‘design-based’ methods and ‘model- based’ methods. The latter methods use either the frequentist approach or a full Bayesian methodology, and in some cases combine the two, which is known in the SAE literature as ‘empirical Bayes’. Design-based methods often use a model for the construction of the estimators (known as ‘model assisted’), but the bias, variance and other properties of the estimators are evaluated under the randomization (design-based) distribution. The randomization distribution of an estimator is the distribution over all the samples that can possibly be selected from the target population of interest under the sampling design used to select the sample, with the population measurements considered as fixed values. Model-based methods on the other hand usually condition on the selected sample and the inference is with respect to the underlying model.

The common feature to design-based and model-based SAE is the use of covariate (auxiliary) information, as obtained from surveys and/or administrative records, such as censuses or registers. Some estimators only require knowledge of the covariates for the sampled units, and the true area means of these covariates. Other estimators require knowledge of the covariates for every unit in the population. The use of auxiliary information for SAE is vital because with the small sample sizes often encountered in practice, even the most elaborated model can be of little help if it does not involve a set of covariates that provide ample information on the small area quantities of interest.

3. NOTATION

Consider a population U of size N, divided into M exclusive and exhaustive areas $U_1 \cup ... \cup U_M$ with N_i units in area i, $i = 1,...,M$, such that $\sum_{i=1}^{M} N_i = N$. We assume the

availability of samples for $m \le M$ of the areas. Let $s = s_1 \cup ... \cup s_m$ define the sample, where s_i of size n_i defines the sample observed for area i, $\sum_{i=1}^{m} n_i = n$. Notice that n_i is random unless a planned sample of a fixed size is taken in that area. Let y define the characteristic of interest and denote by y_{ij} the response value for unit j belonging to area i, $i = 1,...,M$; $j = 1...N_i$, with sample means $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij} / n_i$. When information is available for p covariates, we denote by $\mathrm{x}_{ij} = (x_{1ij},...,x_{pij})'$ the covariates associated with unit (i, j) and by $\bar{\mathrm{x}}_i = \sum_{j=1}^{n_i} \mathrm{x}_{ij} / n_i$ the sample means. The corresponding true means are $\bar{X}_i = \sum_{j=1}^{N_i} \mathrm{x}_{ij} / N_i$. The small area target quantity is denoted by θ_i; for example, $\theta_i = \bar{Y}_i = \sum_{j=1}^{N_i} y_{ij} / N_i$, the true area mean. The common case of estimating a proportion is a special case where y_{ij} is binary. In other applications θ_i may represent a count or a quantile.

4. DESIGN-BASED METHODS

4.1 Design-Based Estimators in Common Use

A recent comprehensive review of design-based methods in SAE is provided by Lehtonen and Veijanen (2009). Here we only overview some of the basic ideas. Suppose that the sample is selected by simple random sampling without replacement (SRSWOR) and that the target quantities of interest are the area means, $\bar{Y}_i$. Estimation of a mean contains as special cases the estimation of a proportion, in which case y_{ij} is binary, and the estimation of the distribution $F_i(t) = \sum_{j \in U_i} v_{ij} / N_i$, in which case $v_{ij} = I(y_{ij} \le t)$, where $I(A)$ is the indicator function. Estimators of the percentiles of the distribution are commonly obtained from the estimated distribution.

If no covariates are available, the *direct* design-unbiased estimator of the area mean and its conditional design variance over the *randomization* distribution are given by,

(4.1) $$\bar{y}_i = \sum_{j=1}^{n_i} y_{ij} / n_i \;\; ; \;\; V_D[\bar{y}_i \mid n_i] = (S_i^2 / n_i)[1 - (n_i / N_i)],$$

where $S_i^2 = \sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_i)^2 / (N_i - 1)$. The term 'direct' is used to signify an estimator that only uses the data for the relevant area at the specific time of interest. The variance $V_D[\bar{y}_i \mid n_i]$ is $O(1/n_i)$ and for small n_i it is large, unless S_i^2 is sufficiently small.

Next suppose that covariates x_{ij} are also observed with $x_{1ij} \equiv 1$. An estimator in common use that utilizes the covariates is the *synthetic* regression estimator,

(4.2) $$\hat{\bar{Y}}_{reg,i}^{syn} = \bar{X}_i'\hat{B} = \frac{1}{N_i}\sum_{j=1}^{N_i}(x_{ij}'\hat{B}),$$

where $\hat{B} = [\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij}x_{ij}']^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij}y_{ij}$ is the ordinary least square (OLS) estimator. Notice that under SRSWOR, $\hat{B}$ is approximately design-unbiased for the vector B of regression coefficients computed from all the population values, irrespective of whether a linear relationship between y and x exists in the population. The term **"**synthetic" refers to the fact that an (approximately) design-unbiased estimator computed from all the areas ($\hat{B}$ in the present case) is used for every area separately, assuming that the areas are 'homogeneous' with respect to the quantity that is estimated. Thus, synthetic estimators borrow information from other 'similar areas' and they are therefore *indirect* estimators.

The obvious advantage of the synthetic estimator over the simple sample mean or any other direct estimator such as the regression estimator $\hat{\bar{Y}}_{reg,i} = \bar{y}_i + (\bar{X}_i - \bar{x}_i)'\hat{B}_i$, where $\hat{B}_i$ is computed only from the data observed for area i, is that $Var_D(\hat{\bar{Y}}_{reg,i}^{syn}) = O(1/n)$, and $n = \sum_{i=1}^{m} n_i$ is usually large. The use of the synthetic estimator is motivated ("assisted") by a linear regression model of y on x in the population with a common vector of coefficients. However, for $x_{1ij} \equiv 1$, $E_D(\hat{\bar{Y}}_{reg,i}^{syn} - \bar{Y}_i) \cong -\bar{X}_i'(B_i - B)$, where B_i is the OLS computed from all the population values in area i. Thus, if in fact different regression coefficients B_i operate in different areas, the synthetic estimator may have a large bias. When the sample is selected with unequal probabilities, the OLS estimator $\hat{B}$ in (4.2) is commonly replaced by the probability weighted estimator $\hat{B}_{pw} = [\sum_{i=1}^{m}\sum_{j=1}^{n_i} w_{ij}x_{ij}x_{ij}']^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n_i} w_{ij}x_{ij}y_{ij}$, where $\{w_{ij} = 1/\Pr[(i,j) \in s]\}$ are the base sampling weights.

In order to deal with the possible large bias of the synthetic estimator, it is common to estimate the bias and then subtract it from the synthetic estimator. The resulting *generalized regression* estimator (GREG) takes the form,

(4.3) $$\hat{\bar{Y}}_i^{GREG} = \bar{X}_i'\hat{B}_{pw} + \frac{1}{N_i}\sum_{j=1}^{n_i} w_{ij}(y_{ij} - x_{ij}'\hat{B}_{pw}) = \hat{\bar{Y}}_{i,H-T} + (\bar{X}_i - \hat{\bar{X}}_{i,H-T})'\hat{B}_{pw},$$

where $(\hat{\bar{Y}}_{i,H-T}, \hat{\bar{X}}_{i,H-T})$ are the Horvitz-Thompson estimators of $(\bar{Y}_i, \bar{X}_i)$. The GREG is approximately design-unbiased and performs well when the covariates have good predictive power, but the variance is back to order $O(1/n_i)$. The variance is often reduced by multiplying the bias correction, $\sum_{j=1}^{n_i} w_{ij}(y_{ij} - x'_{ij}\hat{B}_{pw})/N_i$, by $N_i / \hat{N}_i = N_i / \sum_{j=1}^{n_i} w_{ij}$.

A compromise between the possibly large bias of the synthetic estimator and the possibly large variance of the GREG is achieved by a linear combination of the two. The resulting *combined* (*composite*) estimator is defined as,

(4.4) $$\hat{\bar{Y}}_i^{COM} = \delta_i \hat{\bar{Y}}_i^{GREG} + (1-\delta_i)\hat{\bar{Y}}_{reg,i}^{syn}, \quad 0 \le \delta_i \le 1.$$

Ideally, the coefficient δ_i should be chosen such that it minimizes the mean square error (MSE) of $\hat{\bar{Y}}_i^{COM}$, but assessing sufficiently accurately the bias of the synthetic estimator for any given area is basically impossible. Consequently, it is common to let δ_i depend on the achieved sample size, n_i, in the area, such that the larger n_i, the larger is δ_i. See Rao (2003) for review of methods of specifying δ_i.

4.2 Some New Developments in Design-Based Small Area Estimation

A general class of estimators is obtained by calibrating the base sampling weights $w_{ij} = 1/\Pr[(i,j) \in s]$. Suppose that the population can be partitioned into C calibration groups, $U = U_{(1)} \cup ... \cup U_{(C)}$ with known totals t_{cx} of auxiliary variables x in the groups, such that each area U_i belongs to one of the calibration groups. Let $s = s_{(1)} \cup ... \cup s_{(C)}$ define the corresponding partitioning of the sample. In a special case $C = 1$ and $U_{(1)} = U$. The *calibrated* estimator of the area mean $\bar{Y}_i$ is computed as,

(4.5) $$\hat{\bar{Y}}_i^{cal} = \sum_{j=1}^{n_i} w_{ij}^c y_{ij} / N_i \;;\; \sum_{j \in s_{(c)}} w_{ij}^c \mathrm{x}_{ij} = t_{c\mathrm{x}}.$$

The calibration weights $\{w_{ij}^c\}$ are chosen in such a way that they minimize an appropriate distance from the base weights $\{w_{ij}\}$, subject to satisfying the constraints $\sum_{j \in s_{(c)}} w_{ij}^c \mathrm{x}_{ij} = t_{c\mathrm{x}}$. For example, when using the distance $\chi^2 = \sum_{j \in s_{(c)}} (w_{ij}^c - w_{ij})^2 / w_{ij}$ and $x_{1ij} \equiv 1$, the calibrated weights are,

(4.6) $$w_{ij}^c = w_{ij} g_{ij}; \quad g_{ij} = \{1 + (t_{c\mathrm{x}} - \hat{t}_{c\mathrm{x},\mathrm{H\text{-}T}})'[\sum_{i,j \in S_{(c)}} w_{ij} \mathrm{x}_{ij} \mathrm{x}'_{ij}]^{-1} \mathrm{x}_{ij}\},$$

where $\hat{t}_{\mathrm{cx,H\text{-}T}}$ is the H-T estimator of the total and $\hat{B}_{c,pw}$ is the probability weighted estimator of the regression coefficients in the group. When $U_c = U_i$ (the calibration group is the domain), $\hat{\bar{Y}}_i^{cal}$ is the direct GREG in the domain.

Calibration of the sampling weights is in broad use in sample survey practice not only for SAE. See Kott (2009) for a recent comprehensive review and discussion. The basic idea behind the use of calibrated estimators in SAE is that if y is approximately a linear combination of x in the domains belonging to U_c, then $\bar{Y}_i \cong \bar{X}_i' B_c$ for domain $i \in U_c$, and since $\sum_{j \in s_{(c)}} w_{ij}^c \mathrm{x}_{ij} = t_{\mathrm{cx}}$, $\hat{\bar{Y}}_i^{cal} = \sum_{k \in s_i} w_{ik}^c y_{ik} / N_i$ will be a good estimator of $\bar{Y}_i$. Indeed, the advantage of the use of (4.5) over (4.3) is that it is assisted by a model that only assumes common regression coefficients within the groups U_c, and not for all the domains, as implicitly assumed by the use of (4.3). The estimator (4.5) is approximately design-unbiased irrespective of any model, but $Var_D(\hat{\bar{Y}}_i^{cal} \mid n_i) = O(1/n_i)$, which may still be large.

Another way of calibrating the weights is by use of *instrumental variables* (Estevao and Särndal, 2004, 2006). Denote the vector of instrument values for unit (i,j) by h_{ij}. The calibrated weights are defined as,

$$w_{ij}^{ins} = w_{ij}(1 + g_c' h_{ij}) \;;\; g_c' = (t_{\mathrm{cx}} - \hat{t}_{\mathrm{cx,H\text{-}T}})' [\sum_{i,j \in S_{(c)}} w_{ij} h_{ij} \mathrm{x}_{ij}']^{-1} . \tag{4.7}$$

Notice that the instrument values need only be known for the sampled units in $s_{(c)}$ and that $\sum_{j \in s_{(c)}} w_{ij}^{ins} \mathrm{x}_{ij} = t_{\mathrm{cx}}$, thus satisfying the same constraints on the auxiliary variables as before. The calibrated estimator of $\bar{Y}_i$ is now $\hat{\bar{Y}}_{i,h}^{cal} = \sum_{j=1}^{n_i} w_{ij}^{ins} y_{ij} / N_i$. The instruments may include some of the variables in x. When $\mathrm{h}=\mathrm{x}$, $w_{ij}^{ins} = w_{ij}^c$.

The synthetic estimator (4.2), the GREG (4.3) and the various calibration-based estimators considered above are all assisted by models that assume a linear relationship between y and x. These estimators only require knowledge of the covariates for the sampled units, and the area (or group) totals of these covariates. Lehtonen *et al*. (2003, 2005) consider the use of generalized linear models (GLM) and even generalized linear mixed models (GLMM) as the assisting models, which requires knowledge of the covariates for every element in the population. Suppose that $E_M(y_{ij}) \cong f(\mathrm{x}_{ij}; \psi)$ for some nonlinear function $f(\cdot)$ with an unknown vector parameter ψ. A simple important example is where $f(\mathrm{x}_{ij}; \psi)$ is the logistic

function. Estimating ψ by the pseudo-likelihood approach yields the estimator $\hat{\psi}_{pl}$ and the predicted values $\{\hat{y}_{ij} = f(\mathrm{x}_{ij};\hat{\psi}_{pl})\}$. The pseudo-likelihood approach consists of estimating the likelihood equations that would be obtained in the case of a census by the corresponding H-T estimators (or alternatively weighting each sample observation by its sampling weight), and then maximizing the resulting estimated equations. The synthetic and the GREG estimators are computed as,

$$(4.8) \qquad \hat{\bar{Y}}^{syn}_{GLM,i} = \frac{1}{N_i}\sum_{j=1}^{N_i} f(\mathrm{x}_{ij};\hat{\psi}_{pl}); \quad \hat{\bar{Y}}^{GREG}_{GLM,i} = \hat{\bar{Y}}^{syn}_{GLM,i} + \frac{1}{N_i}\sum_{j=1}^{n_i} w_{ij}[y_{ij} - f(\mathrm{x}_{ij};\hat{\psi}_{pl})].$$

A further extension consists of adding random area effects to the assisting model, that is, assuming $E_M(y_{ij} \mid \mathrm{x}_{ij}, u_i) \cong f(\mathrm{x}_{ij}, u_i; \psi^*)$, with $E_M(u_i) = 0$, $Var_M(u_i) = \sigma_u^2$. Estimation of the model parameters ψ^*, σ_u^2 is now under the model, ignoring the sampling weights. The synthetic and GREG estimators are defined similarly to (4.8) but with $f(\mathrm{x}_{ij};\hat{\psi}_{pl})$ replaced by $f(\mathrm{x}_{ij}, \hat{u}_i; \hat{\psi}^*)$. For a sufficiently large sample size within the area, the extended GREG is approximately design unbiased for the true area mean but it is not clear how to estimate the design (randomization) variance in this case in a way that accounts for the prediction of the random effects (see Section 6). Torabi and Rao (2008) compare the MSE of model-based predictors and a GREG assisted by a GLMM.

Jiang and Lahiri (2006a) propose the use of model-dependent estimators that are design consistent under the randomization distribution as the area sample sizes increase. The basic idea is to model the direct estimators $\hat{\bar{Y}}_{iw} = \sum_{j=1}^{n_i} w_{ij} y_{ij} / \sum_{j=1}^{n_i} w_{ij}$ instead of the individual observations y_{ij}, and then obtain the empirical best predictor of the true mean in each area under the model. The authors consider the general two-level model $E[\hat{\bar{Y}}_{iw} \mid u_i] = \xi_i$ $= \xi(u_i, \hat{\bar{X}}_{iw}; \psi)$, where the u_i s are independent random effects with zero mean and variance σ_u^2, and $\xi(\cdot)$ is some known function governed by the unknown vector parameter ψ. The empirical best predictor is the best predictor under the model but with the unknown parameters replaced by model consistent estimators; $\hat{\bar{Y}}_i^{EBP} = E_M(\xi_i \mid \hat{\bar{Y}}_{iw}; \hat{\psi})$. The estimator is shown to be model-consistent under the correct model and design-consistent for large n_i even if the model is misspecified, thus robustifying the estimation. Note, however, that the problem of SAE is the small sample sizes in some or all of the areas, such that design consistency for large sample size within the area is not a very appealing property in this case.

The authors develop estimators of the prediction mean squared error (PMSE) for bounded sample sizes, with bias of desired order $o(1/m)$, where m is the number of sampled areas. The PMSE is computed with respect to the model holding for the individual observations and over the randomization distribution.

Chandra and Chambers (2009) propose the use of model-design based estimators, which they call model-based direct estimators (MBDE). The idea here is to fit a model holding for the population values, obtain the EBLUP weights for predicting the population total and then apply the same weights in each small area using classical design-based direct estimators. The model fitted in this study is the general linear model,

$$Y_U = X_U\beta + \varepsilon_U;\ E(\varepsilon_U) = 0,\ E(\varepsilon_U\varepsilon_U') = \Sigma = \begin{bmatrix} \Sigma_{ss} & \Sigma_{sr} \\ \Sigma_{rs} & \Sigma_{rr} \end{bmatrix}, \tag{4.9}$$

where the index U signifies that this is a population level model, s signifies the sample of size n and r signifies the sample complement of size $(N-n)$. (N is the population size). Notice that the small area models defined by (5.1) and (5.3) are special cases of (4.9). For known model parameters, the BLUP of the population total $t_y = \sum_{k=1}^{N} y_k$ under this model is,

$$\hat{t}_y^{BLUP} = 1_n' y_s + 1_{N-n}'[X_r\hat{\beta}_{GLS} + \Sigma_{rs}\Sigma_{ss}^{-1}(y_s - X_s\hat{\beta}_{GLS})] = \sum_{i\in s} w_i^{BLUP} y_i\ . \tag{4.10}$$

The EBLUP predictor is $\hat{t}_y^{EBLUP} = \sum_{i\in s} w_i^{EBLUP} y_i$, where the EBLUP weights are the same as in (4.10) but with estimated parameters. The MBDE of the true mean in area i is,

$$\hat{\bar{Y}}_i^{MBD} = \sum_{j=1}^{n_i} w_j^{EBLUP} y_i \Big/ \sum_{j=1}^{n_i} w_j^{EBLUP}\ . \tag{4.11}$$

The authors derive estimators for the bias and variance of the small area estimators and illustrate the robustness of the proposed estimators to certain model misspecifications.

All the estimators considered so far assume a given sampling design with random sample sizes within the areas. When the areas of interest are known in advance, considerable gains in efficiency can be achieved by changing the sampling design and in particular, by controlling the sample sizes within the areas. Ideally, the preferred sampling scheme in such a case would be stratified sampling with the strata defined by the areas. In practice, however, this sampling scheme may not be feasible when estimates are required for different non-nested partitions of the population into small areas since it requires the use of cross-classified strata and there may be too many of them for a given overall sample size. In a recent article, Falrosi and Righi (2008) propose a general sampling strategy for multivariate multi-domain estimation that guarantees that the sampling errors of the domain estimators are lower than

pre-specified thresholds. The strategy combines the use of a balanced sampling technique and GREG estimation, but extensions to the use of synthetic estimators and model-based estimation are also considered. A successful application of this strategy requires good predictions of weighted sums of residuals featuring in the variance expressions, and it may happen that the resulting overall sample size is far too large, but this is a promising avenue that should be studied further. The article contains several empirical illustrations.

4.3 Pros and Cons of Design-Based Small Area Estimation

The apparent advantage of design-based methods is that the estimation is less dependent on an assumed model, although models are used (assisted) for the construction of the estimators. The estimators are aimed to be approximately design unbiased and consistent under the randomization distribution for large sample sizes within the areas, which are desirable properties that add some protection against possible model misspecification. However, as noted above, in practical applications the area sample sizes are often very small, such that these properties should be judged with caution.

Against this advantage stand many disadvantages. Direct estimators generally have large variance due to the small sample sizes. GREG type estimators are approximately unbiased but may likewise be too variable with small sample sizes. Synthetic estimators are generally biased, with limited possibilities to assess the bias. Composite estimators have a smaller bias than the corresponding synthetic estimator but larger variance, and it is not obvious how to best choose the weights attached to the synthetic estimator and the GREG (or other direct estimators). Computation of randomization-based confidence intervals generally requires large sample normality assumptions, but the sample sizes in at least some of the areas may be very small.

Another limitation of design-based inference (not restricted to SAE) is that it does not lend itself to conditional inference, for example, conditioning on the sampled values of the covariates or the sampled clusters in a two-stage sampling design. This again inflates the variance of the estimators. Conditional inference is in the heart of classical statistical inference under both the frequentist and the Bayesian approaches. Last, but not the least important limitation of design-based SAE is that there is no founded theory for estimation in areas with no samples. The use of the randomization distribution does not extend to prediction problems, such as the prediction of the dependent variable for given covariates under a regression model, or the prediction of small area means for areas with no samples. Design-based theory is restricted to estimation of population quantities from a sample taken

from this population. As stated in the introduction, it is often the case that samples are available for only a minority of the areas but estimators and MSE estimators are required for each of the areas, whether sampled or not.

5. MODEL-BASED METHODS

5.1 General Formulation

Model-based methods assume a model for the sample data and use the optimal, or approximately optimal predictor of the area characteristic of interest under the model. The MSE of the prediction error is likewise defined and estimated under the model. Note that I now use the term 'prediction' rather than estimation because the target characteristics are generally random under the model. The use of models overcomes the problems underlying the use of design-based methods discussed in Section 4.3, but it is important to emphasize again that even the most elaborate model cannot produce predictors with acceptable precision when the area sample sizes are too small and no covariates with good predictive power are available. The use of models raises the question of the robustness of the inference to possible model misspecification, and I later review various studies that deal with this problem.

Denote by θ_i the quantity of interest in area i (mean, proportion, quantile,…). Let y_i define the observed responses for area i (when the area is sampled) and x_i define the corresponding values of the covariates (when available). As becomes evident below, y_i is either a scalar, in which case x_i is a vector, or y_i is a vector, in which case x_i is usually a matrix. A typical small area model consists of two parts: The first part models the distribution (moments) of $y_i \mid \theta_i;\psi_{(1)}$. The second part models the distribution (moments) of $\theta_i \mid \mathrm{x}_i;\psi_{(2)}$, linking the θ_i's to known covariates and to each other. This is achieved by including in the model random effects that account for the variability of the θ_i's not explained by the covariates. The hyper-parameters $\psi = (\psi_{(1)}, \psi_{(2)})$ are typically unknown and are estimated either under the frequentist approach or under the Bayesian approach, after setting appropriate prior distributions. In some applications the index i may define time, in which case the model for $\theta_i \mid \mathrm{x}_i;\psi_2$ is a time series model.

5.2 Models in Common Use

In this section I review briefly three models in common use, as most of the recent developments in SAE relate to these models or extensions of them. For more details see Pfeffermann (2002), Rao (2003) and Datta (2009), and the references therein.

5.2.1 *Area level model*

This model is in broad use when the covariate information is only at the area level, so that x_i is a vector of known area characteristics. The model, studied originally for SAE by Fay and Herriot (1979) is defined as,

(5.1) $$\tilde{y}_i = \theta_i + e_i \;\; ; \;\; \theta_i = \mathrm{x}_i'\beta + u_i ,$$

where $\tilde{y}_i$ denotes the direct sample estimator of θ_i (for example, the sample mean $\overline{y}_i$) and e_i represents the sampling error, assumed to have zero mean and known design (randomization) variance, $Var_D(e_i) = \sigma_{Di}^2$. The random effects u_i are assumed to be independent with zero mean and common variance σ_u^2. For known σ_u^2, the best linear unbiased predictor (BLUP) of θ_i under this model is,

(5.2) $$\hat{\theta}_i = \gamma_i \tilde{y}_i + (1-\gamma_i)\mathrm{x}_i'\hat{\beta}_{GLS} = \mathrm{x}_i'\hat{\beta}_{GLS} + \gamma_i(\tilde{y}_i - \mathrm{x}_i'\hat{\beta}_{GLS}) ,$$

where $\hat{\beta}_{GLS}$ is the generalized least square (GLS) estimator of β under the model. The BLUP $\hat{\theta}_i$ is in the form of a composite estimate (Eq. 4.4) but with a tuning (shrinkage) coefficient $\gamma_i = \sigma_u^2 / (\sigma_u^2 + \sigma_{Di}^2)$, which depends optimally on the ratio $\sigma_u^2 / \sigma_{Di}^2$ of the variances of the prediction errors of $\mathrm{x}_i'\beta$ and $\tilde{y}_i$ respectively. Under normality of the error terms and a diffuse uniform prior for β, $\hat{\theta}_i$ is also the Bayesian predictor (posterior mean) of θ_i. For a nonsampled area k (but known x_k), the BLUP is now obtained optimally as $\mathrm{x}_k'\hat{\beta}_{GLS}$. In practice, the variance σ_u^2 is seldom known and it is replaced in γ_i and $\hat{\beta}_{GLS}$ by a sample estimate, yielding what is known as the empirical BLUP (EBLUP) under the frequentist approach, or the empirical Bayes (EB) predictor. Alternatively, one may assume a prior distribution for σ_u^2 and compute the posterior distribution of θ_i given the available data, which is then used for the computation of the point predictor and credibility intervals.

Remark 1. The synthetic estimator $\mathrm{x}_i'\hat{\beta}_{GLS}$ in (5.2), and hence also the BLUP $\hat{\theta}_i$ are unbiased under the joint distribution of y_i and θ_i, but are biased predictors of θ_i when conditioning on

u_i, similarly to what we had under the randomization distribution. Conditioning on u_i amounts to assuming different fixed intercepts in different areas. The unbiasedness of $\hat{\theta}_i$ under the model is achieved by viewing the intercepts as random. The same applies for the BLUP (5.4) below.

5.2.2 *Nested error unit level model*

This model uses individual observations y_{ij} such that y_i is now a vector and x_i is generally a matrix. As with design-based methods, the use of this model for SAE requires that the area means, $\overline{\mathrm{X}}_i = \sum_{j=1}^{N_i} \mathrm{x}_{ij} / N_i$ are known. The model, first proposed for SAE by Battese *et al.* (1988) has the form,

(5.3) $$y_{ij} = \mathrm{x}'_{ij}\beta + u_i + \varepsilon_{ij},$$

where the u_i's (random effects) and the ε_{ij} s (residual terms) are mutually independent with zero means and variances σ_u^2 and σ_ε^2 respectively. Under the model, the true small area means are $\overline{Y}_i = \overline{X}_i'\beta + u_i + \overline{\varepsilon}_i$, but since $\overline{\varepsilon}_i = \sum_{j=1}^{N_i} \varepsilon_{ij} / N_i \cong 0$ for large N_i, the target means are often defined as $\theta_i = \overline{X}_i'\beta + u_i = E(\overline{Y}_i \mid u_i)$. For known variances $(\sigma_u^2, \sigma_\varepsilon^2)$, the BLUP of θ_i is,

(5.4) $$\hat{\theta}_i = \gamma_i[\overline{y}_i + (\overline{X}_i - \overline{\mathrm{x}}_i)'\hat{\beta}_{GLS}] + (1-\gamma_i)\overline{X}_i'\hat{\beta}_{GLS},$$

where $\hat{\beta}_{GLS}$ is the GLS of β computed from all the observations, $\overline{\mathrm{x}}_i = \sum_{j=1}^{n_i} \mathrm{x}_{ij} / n_i$ and $\gamma_i = \sigma_u^2 / (\sigma_u^2 + \sigma_\varepsilon^2 / n_i)$. For area k with no sample (but known $\overline{X}_k$), the BLUP is $\hat{\theta}_k = \overline{\mathrm{X}}_k'\hat{\beta}_{GLS}$. See Rao (2003) for the BLUP of the means $\overline{Y}_i$ in sampled areas.

As with the area level model, the BLUP (5.4) is also the Bayesian predictor (posterior mean) under normality of the error terms and a diffuse uniform prior for β. Replacing the unknown variances σ_u^2 and σ_ε^2 in γ_i and $\hat{\beta}_{GLS}$ by sample estimates yields the corresponding EBLUP or EB predictors. Hierarchical Bayes (HB) predictors are obtained by specifying prior distributions for β and the two variances and computing the posterior distribution of θ_i (or $\overline{Y}_i$) given all the observations in all the areas.

5.2.3 *Mixed logistic model*

The previous two models assume continuous response values. Suppose now that y_{ij} is binary taking the values 1 and 0, in which case the small area quantities of interest are usually proportions or counts (say, the proportion or total of unemployed persons in the area). The

following generalized linear mixed model (GLMM) considered originally by MacGibbon and Tomberlin (1989) for SAE is in broad use for this kind of problems:

(5.5) $\Pr(y_{ij}=1 \mid p_{ij}) = p_{ij}\ ;\ logit(p_{ij}) = \mathrm{x}'_{ij}\beta + u_i\ ;\ u_i \sim N(0,\sigma_u^2)$.

The responses y_{ij} are assumed to be conditionally independent given the random effects u_i, and likewise for the random effects. The purpose is to predict the true area proportions $p_i = \sum_{j=1}^{N_i} y_{ij} / N_i$. Let $\psi = (\beta, \sigma_u^2)$ denote the model parameters. For this model there is no explicit expression for the best predictor (BP), $E(p_i \mid y_i, \mathrm{x}_i; \psi)$, but as shown in Jiang and Lahiri (2006b), for known ψ the BP can be computed (approximated) numerically as the ratio of two one-dimensional integrals. The authors review methods of estimating ψ, yielding the empirical BP (EBP) $\hat{p}_i^{EBP} = E(p_i \mid y_i, \mathrm{x}_i; \hat{\psi})$. Alternatively, the predictors $\hat{p}_i$ can be obtained by application of the EB approach (setting $\psi = \hat{\psi}$), as in MacGibbon and Tomberlin (1989), or by application of the full HB approach, as in Malec *et al.* (1997). The application of the HB approach consists of the following basic steps:

1. Specify prior distributions for σ_u^2 and β;

2. Compute the posterior distributions of β and $u_1, ..., u_m$ by say, MCMC simulations and draw a large number of realizations $(\hat{\beta}^{(r)}, \{\hat{u}_i^{(r)}\})$, $r = 1, ..., R$, $i = 1, ..., m$, and hence realizations, $y_{ik}^{(r)} \sim p_{ik}^{(r)} = \dfrac{\exp(x'_{ik}\beta^{(r)} + u_i^{(r)})}{1 + \exp(x'_{ik}\beta^{(r)} + u_i^{(r)})}$ for $k \notin s_i$;

3. Predict: $\hat{p}_i = (\sum_{j \in s_i} y_{ij} + \sum_{k \notin s_i} \hat{y}_{ik}) / N_i\ ;\ \hat{y}_{ik} = \sum_{r=1}^{R} y_{ik}^{(r)} / R,\ k \notin s_i$.

Writing $\hat{p}_i = \dfrac{1}{R}\sum_{r=1}^{R}(\sum_{j \in s_i} y_{ij} + \sum_{k \notin s_i} \hat{y}_{ik}^{(r)}) / N_i = \dfrac{1}{R}\sum_{r=1}^{R} \hat{p}_i^{(r)}$, the posterior variance is approximated as, $\hat{V}_{post}(\hat{p}_i) = \dfrac{1}{R(R-1)}\sum_{r=1}^{R}(\hat{p}_i^{(r)} - \hat{p}_i)^2$.

6. NEW DEVELOPMENTS IN MODEL-BASED SAE

6.1 Estimation of Prediction MSE

As stated in the introduction, an important aspect of SAE is the assessment of the prediction accuracy of the predictors. This problem is solved 'automatically' under the full Bayesian paradigm, which produces realizations of the posterior distribution of the target quantities around their predictors (the posterior means). However, estimation of the

prediction MSE (PMSE) and the computation of confidence intervals (C.I.) under the frequentist approach is complicated because of the added variability induced by the estimation of the model hyper-parameters. Prasad and Rao (1990) develop PMSE estimators with bias of order $o(1/m)$, (m is the number of sampled areas), under the linear mixed models 5.1.1 and 5.1.2 for the case where the random errors have a normal distribution and the model variances are estimated by the ANOVA type method of moments estimators. Lahiri and Rao (1995) show that the PMSE estimator in the case of the model 5.1.1 is robust to departures from normality of the random area effects. Datta and Lahiri (2000) extend the estimation of Prasad and Rao to general linear mixed models of the form,

(6.1) $$y_i = X_i\beta + Z_i u_i + e_i, \quad i = 1...m,$$

where Z_i is a fixed matrix of order $n_i \times d$, and u_i and e_i are independent normally distributed random effects and residual terms of orders $d \times 1$ and $n_i \times 1$ respectively; $u_i \sim N_d(0, Q_i)$, $e_i \sim N_{n_i}(0, R_i)$. The authors develop MSE estimators with bias of order $o(1/m)$ for the EBLUP obtained when estimating Q_i and R_i by MLE or REML. Datta *et al.* (2005) show that for the area level model (5.1), if σ_u^2 is estimated by the method proposed by Fay and Herriot (1979), then it is required to add an extra term to the PMSE estimator to achieve the desired order of bias of $o(1/m)$. See Datta (2009) for review of methods for estimating the PMSE of the EBLUP and EB under mixed linear models.

Estimation of the PMSE under GLMM is more involved and following we review resampling procedures that can be used in such cases. For convenience, we consider the mixed logistic model (5.5) but the procedures are applicable to other models belonging to this class. The first procedure, proposed by Jiang *et. al* (2002) uses the Jackknife method. Let $\lambda_i = E(\hat{p}_i^{EBP} - p_i)^2$ denote the PMSE, where $p_i = \sum_{j=1}^{N_i} y_{ij} / N_i$ is the true proportion and $\hat{p}_i^{EBP} = E(p_i \mid y_i, \mathrm{x}_i; \hat{\psi})$ is the empirical best predictor. The following decomposition holds,

(6.2) $$\lambda_i = E(\hat{p}_i^{(BP)} - p_i)^2 + E(\hat{p}_i^{(EBP)} - \hat{p}_i^{(BP)})^2 = M_{1i} + M_{2i},$$

where M_{1i} is the PMSE of the BP (assumes known parameter values) and M_{2i} is the contribution to the PMSE from estimating the model parameters. Denote by $\hat{\lambda}_i^{BP}(\hat{\psi})$ the '*naive*' estimator of M_{1i}, obtained by setting $\psi = \hat{\psi}$. Let $\hat{\lambda}_i^{BP}(\hat{\psi}_{-l})$ denote the naive estimator when estimating ψ from all the areas except for area l, and $\hat{p}_i^{EBP}(\hat{\psi}_{-l})$ denote the corresponding EBP. The Jackknife estimator is:

(6.3) $$\hat{\lambda}_i^{JK} = \hat{M}_{1i} + \hat{M}_{2i}; \quad \begin{matrix} \hat{M}_{1i} = \hat{\lambda}_i^{BP}(\hat{\psi}) - \frac{m-1}{m}\sum_{l=1}^{m}[\hat{\lambda}_i^{BP}(\hat{\psi}_{-l}) - \hat{\lambda}_i^{BP}(\hat{\psi})] \\ \hat{M}_{2i} = \frac{m-1}{m}\sum_{l=1}^{m}[\hat{p}_i^{EBP}(\hat{\psi}_{-l}) - \hat{p}_i^{EBP}(\hat{\psi})]^2 \end{matrix} .$$

Under some regularity conditions $E(\hat{\lambda}_i^{JK}) - \lambda_i = o(1/m)$, as desired.

The jackknife estimator estimates the unconditional PMSE over the joint distribution of the random effects and the responses. Lohr and Rao (2009) proposed a modification of the jackknife estimator that is computationally simpler and estimates the conditional PMSE, $E[(\hat{p}_i^{(EBP)} - p_i)^2 \mid y_i]$. Denoting $q_i(\psi, y_i) = Var(\theta_i \mid y_i; \psi)$, the modification consists of replacing $\hat{M}_{1i}$ in (6.3) by $\hat{M}_{1i,c} = q_i(\hat{\psi}, y_i) - \sum_{l\neq i}^{m}[q_i(\hat{\psi}_{-l}, y_i) - q_i(\hat{\psi}, y_i)]$. The modified jackknife estimator, $\hat{\lambda}_{i,c}^{JK} = \hat{M}_{1i,c} + \hat{M}_{2i}$, is shown to have bias of order $o_p(1/m)$ in estimating the conditional PMSE and a bias of order $o(1/m)$ in estimating the unconditional PMSE.

Hall and Maiti (2006) propose estimating the PMSE based on double-bootstrap. For the model (5.5) the procedure consists of the following steps:

1. Generate a new population from the model (5.5) with parameters $\hat{\psi}$ and compute the 'true' area proportions for this population. Compute the EBPs based on new sample data and newly estimated parameters. The new population and sample use the same covariates as the original population and sample. Repeat the same process independently B_1 times, with B_1 sufficiently large. Denote by $p_{i,b_1}(\hat{\psi})$ and $\hat{p}_{i,b_1}^{(EBP)}(\hat{\psi}_{b_1})$ the 'true' proportions and corresponding EBPs for population and sample b_1, $b_1 = 1,..., B_1$. Compute the first-step bootstrap MSE estimator,

(6.4) $$\hat{\lambda}_{i,1}^{BS} = \frac{1}{B_1}\sum_{b_1=1}^{B_1}[\hat{p}_{i,b_1}^{(EBP)}(\hat{\psi}_{b_1}) - p_{i,b_1}(\hat{\psi})]^2 .$$

2. For each sample drawn in Step 1, repeat the computations of Step 1 B_2 times with B_2 sufficiently large, yielding new 'true' proportions $p_{i,b_2}(\hat{\psi}_{b_1})$ and EBPs $\hat{p}_{i,b_2}^{(EBP)}(\hat{\psi}_{b_2})$, $b_2 = 1,..., B_2$. Compute the second-step bootstrap MSE estimator,

(6.5) $$\hat{\lambda}_{i,2}^{BS} = \frac{1}{B_1}\sum_{b_1}^{B_1}\frac{1}{B_2}\sum_{b_2=1}^{B_2}[\hat{p}_{i,b_2}^{(EBP)}(\hat{\psi}_{b_2}) - p_{i,b_2}(\hat{\psi}_{b_1})]^2 .$$

The double-bootstrap MSE estimator is obtained by computing one of the classical bias corrected estimators. For example,

(6.6) $$\hat{\lambda}_i^{D-BS} = \begin{cases} \hat{\lambda}_{i,1}^{BS} + (\hat{\lambda}_{i,1}^{BS} - \hat{\lambda}_{i,2}^{BS}), & \text{if } \hat{\lambda}_{i,1}^{BS} \geq \hat{\lambda}_{i,2}^{BS} \\ \hat{\lambda}_{i,1}^{BS} \exp[(\hat{\lambda}_{i,1}^{BS} - \hat{\lambda}_{i,2}^{BS}) / \hat{\lambda}_{i,2}^{BS}], & \text{if } \hat{\lambda}_{i,1}^{BS} < \hat{\lambda}_{i,2}^{BS} \end{cases}.$$

The estimator has bias of order $o(1/m)$ under some regularity conditions.

Remark 2. All the above procedures although being resampling methods are actually parametric methods that rely on the underlying model.

6.2 Construction of Confidence Intervals

As in other statistical applications, very often analysts are interested in prediction intervals for the unknown area characteristics, which can be used also for comparing between the areas by appropriate hypothesis testing. Construction of prediction intervals under the full Bayesian approach, known as *credibility intervals* is straightforward via the posterior distribution of the predictor. A natural prediction interval under the EB and the frequentist approaches with desired coverage rate $(1-\alpha)$ is $\hat{\theta}_i^{(\cdot)} \pm z_{\alpha/2}[\hat{Var}(\hat{\theta}_i^{(\cdot)} - \theta_i)]^{1/2}$, where $\hat{\theta}_i^{(\cdot)}$ is the EB, EBP or EBLUP predictor and $\hat{Var}(\hat{\theta}_i^{(\cdot)} - \theta_i)$ is an appropriate estimate of the prediction error variance. However, even under asymptotic normality of the prediction error, the use of this prediction interval has coverage error of order $O(1/n)$, which is not accurate enough for many small area applications. Recent work on small area prediction intervals focuses, therefore, on reducing the coverage error via parametric bootstrap.

Hall and Maiti (2006) consider the following general model: For a suitable smooth function $f_i(\beta)$ of the covariates $X_i = (X_{i1},...,X_{in_i})$ in area i and a vector parameter β, random variables $\Theta_i = f_i(\beta) + u_i;\ E(u_i) = 0$ are drawn from a distribution $Q\{f_i(\beta), \xi\}$. The observations Y_{ij} are drawn independently from the distribution $R\{\psi(\Theta_i), \eta_i\}$, where $\psi(\cdot)$ is a known link function and η_i is either known or is the same for every area i. For given covariates X_{i0} and sample size n_{i0} and known parameters, a α-level prediction interval for the corresponding realization Θ_{i0} is,

(6.7) $$I_\alpha(\beta, \xi) = [q_{(1-\alpha)/2}(\beta, \xi), q_{(1+\alpha)/2}(\beta, \xi)],$$

where $q_\alpha(\beta, \xi)$ defines the α-level quantile of the distribution $Q\{f_i(\beta), \xi\}$. Notice that this interval does not make use of the area-specific direct estimator. For the case of unknown parameters a naive prediction interval is $\hat{I}_\alpha(\hat{\beta}, \hat{\xi})$, but this interval has coverage error of $O(1/m)$. To reduce the error, $\hat{I}_\alpha(\hat{\beta}, \hat{\xi})$ is calibrated on α. This is implemented by generating

parametric bootstrap samples and re-estimating β and ξ, similarly to the first step of the double-bootstrap procedure for PMSE estimation described above. Denote by $\hat{I}^*_\alpha = I_\alpha(\hat{\beta}^*, \hat{\xi}^*)$ the bootstrap interval and let $\hat{\alpha}$ denote the solution of the equation $\Pr(\theta_i^* \in \hat{I}^*_{\hat{\alpha}}) = \alpha$, where $\theta_i^* \sim Q\{f_i(\hat{\beta}), \hat{\xi}\}$. The bootstrap-calibrated prediction interval is $\hat{I}_{\hat{\alpha}}(\hat{\beta}, \hat{\xi})$ and it has coverage error of $O(m^{-2})$. By applying a double-bootstrap procedure, the coverage error is $O(m^{-3})$.

Chatterjee *et al*. (2008) consider the general linear mixed model $Y = X\beta + Zu + e$, where Y (of dimension n) signifies all the observations in all the areas, $X_{n\times p}$ and $Z_{n\times q}$ are known matrices and u and e are independent normal variates of random effects and residual terms with dispersion matrices $D(\psi)$ and $R(\psi)$ of dimensions q and n respectively, which depend on a vector parameter ψ of dimension k. Notice that this model and the model considered by Hall and Maiti (2006) include as special cases the mixed linear models defined by (5.1) and (5.3). The present model cannot handle nonlinear mixed effects (for example, the GLMM 5.5), which the Hall and Maiti model can, but it does not require independence of the observations as under the Hall and Maiti model. The (parametric bootstrap) prediction interval of Chatterjee *et al*. (2008) for a univariate linear combination $t = c'(X\beta + Zu)$ is obtained by the following steps. (Prediction intervals for vectors T can be obtained following similar steps). First compute the conditional mean, μ_t and variance σ_t^2 of $t \mid Y; \beta, \psi$. Next generate new observations $Y^* = X\hat{\beta} + Zu^* + e^*$, where $u^* \sim N(0, D(\hat{\psi}))$, $e^* \sim N(0, R(\hat{\psi}))$. From Y^* estimate $\hat{\beta}^*$ and $\hat{\psi}^*$ using the same method as for $\hat{\beta}$ and $\hat{\psi}$, and compute $\hat{\mu}_t^*$ and $\hat{\sigma}_t^*$ (the counterparts of μ_t and σ_t with estimated parameters $\hat{\beta}^*$ and $\hat{\psi}^*$). Denote by L_n^* the bootstrap distribution of $\hat{\sigma}_t^{-1*}(t^* - \hat{\mu}_t^*)$ and let $d = (q + k)$ be the dimension of the parameter space. Then, if $d^2/n \to 0$ and under some regularity conditions, if q_1, q_2 satisfy $L_n^*(q_2) - L_n^*(q_1) = 1 - \alpha$,

(6.8) $$\Pr(\hat{\mu}_t + q_1\hat{\sigma}_t \le t \le \hat{\mu}_t + q_2\hat{\sigma}_t) = 1 - \alpha + O(d^3 n^{-3/2}).$$

Notice that this theory allows the parameter dimension, d, to grow with the total sample size, n, which is often the case in small area models, and that the coverage error is defined in terms of n, rather than m, the number of sampled areas, as under the Hall and Maiti (2006) theory. The total sample size increases also as the sample sizes within the areas increase, and not just by increasing m.

6.3 Choice of Matching Priors for Bayesian Applications

For the area level model (5.1), Ganesh and Lahiri (2008) develop a class of priors $p(\sigma_u^2)$, which for a given set of weights $\{\omega_i\}$ satisfy,

(6.9) $$\sum_{i=1}^{m} \omega_i E\{Var(\theta_i \mid y) - \text{PMSE}[\hat{\theta}_i(\hat{\sigma}_u^2)]\} = o(1/m),$$

where $Var(\theta_i \mid y)$ is the posterior variance under the desired prior, $\hat{\theta}_i(\hat{\sigma}_u^2)$ is the EBLUP of θ_i obtained by substituting $\hat{\sigma}_u^2$ for σ_u^2 in (5.2) and the expectation and PMSE are computed under the joint distribution of θ and y. The prior $p(\sigma_u^2)$ satisfying (6.9) is shown to be,

(6.10) $$P(\sigma_u^2) \propto \sum_{i=1}^{m} 1/(\sigma_{Di}^2 + \sigma_u^2)^2 \Big/ \sum_{i=1}^{m} \omega_i [\sigma_{Di}^2/(\sigma_{Di}^2 + \sigma_u^2)]^2.$$

The motivation for using the prior (6.10) is to warrant some "frequency validity" to the Bayesian inference by guaranteeing that the weighted average of the expected difference between the posterior variance and the PMSE of the EBLUP is sufficiently close. Having satisfied (6.9), the analyst may then take advantage of the flexibility of Bayesian inference resulting from the ability to draw observations from the posterior distribution. By appropriate choice of the weights $\{\omega_i\}$, the prior (6.10) contains as special cases the flat prior $p(\sigma_u^2) = U(0,\infty)$, the prior developed by Datta *et al*. (2005) for a given area, satisfying $E[Var(\theta_i \mid y)] = PMSE[\hat{\theta}_i(\hat{\sigma}_u^2)] + o(1/m)$ (different prior for different areas), and the average moment matching prior (obtained by setting $\omega_i \equiv 1$).

6.4 Benchmarking

Model-based SAE depends on models that can be hard to validate and if the model is misspecified, the resulting predictors may perform poorly. Benchmarking is another way of trying to robustify the inference by forcing the model-based predictors to agree with the design-based estimator for an aggregation of the areas for which it can be trusted. Assuming for convenience that this aggregation contains all the areas, the benchmarking equation takes the general form,

(6.11) $$\sum_{i=1}^{m} b_i \hat{\theta}_{i,model} = \sum_{i=1}^{m} b_i \hat{\theta}_{i,design}.$$

The coefficients $\{b_i\}$ are fixed weights, assumed without loss of generality to sum to 1 (e.g., relative area sizes). The modification (6.11) has the further advantage of guaranteeing consistency of publication between the model-based small area predictors and the design-based estimator for the aggregated area, which is often required by statistical bureaus. For

example, the model-based predictors of total unemployment in counties should add up to the design-based estimate of total unemployment in the country, which is deemed accurate.

A benchmarking procedure in common use, often referred to as ratio or pro-rata adjustment is defined as,

(6.12) $$\hat{\theta}_i^{bench} = (\sum_{i=1}^{m} b_i \hat{\theta}_{i,design} / \sum_{i=1}^{m} b_i \hat{\theta}_{i,model}) \times \hat{\theta}_{i,model} .$$

The use of this procedure, however, applies the same ratio correction for all the areas, irrespective of the relative precision of the model-based predictors before benchmarking. In addition, estimation of the PMSE of the prorated predictors is not straightforward. Consequently, other procedures have been proposed in the literature.

Wang *et al*. (2008) derive a benchmarked BLUP (BBLUP) under the area level model (5.1) as the predictor minimizing $\sum_{i=1}^{m} \varphi_i E(\theta_i - \hat{\theta}_i^{bench})^2$ subject to (6.11), where the φ_i s are chosen positive weights. The BBLUP is,

(6.13) $$\hat{\theta}_{i,BLUP}^{bench} = \hat{\theta}_{i,model}^{BLUP} + \delta_i \sum_{j=1}^{m} b_j (\theta_{j,design} - \hat{\theta}_{j,model}^{BLUP}); \ \delta_i = (\sum_{j=1}^{m} \varphi_j^{-1} b_j^2)^{-1} \varphi_i^{-1} b_i .$$

When the variance σ_u^2 is unknown, it is replaced by its estimator everywhere in (6.13), yielding the empirical BBLUP. The PMSE of the latter predictor can be estimated by a method developed by Isaki *et al*. (2000), or by one of the resampling procedures described in Section 6.1. You & Rao (2002) achieve "automatic benchmarking" for the unit level model (5.3) by changing the estimator of β. Wang *et al*. (2008) consider a similar procedure for the area level model. The approach is further extended by augmenting the covariates x_i to $\tilde{\mathrm{x}}_i' = [\mathrm{x}_i', b_i \sigma_{Di}^2]$. (The variances σ_{Di}^2 are considered known under the area level model.) The use of the augmented model yields a BLUP that likewise satisfies the benchmark constraint (6.11) and is more robust to model misspecification.

Pfeffermann & Tiller (2006) add monthly benchmark constraints of the form (6.11) to the measurement (observation) equation of a time series state-space model fitted jointly to the direct estimates in several areas. Adding benchmark constraints to time series models used for the production of model-dependent small area predictors is particularly important since time series models are slow to adapt to abrupt changes. By adding the constraints to the model, the use of this approach permits estimating the variance of the benchmarked estimators as part of the model fitting. The variance accounts for the variances of the model error terms, the variances and autocovariances of the sampling errors of the direct estimators

and of the benchmark, $\sum_{i=1}^{m} b_i \hat{\theta}_{i,direct}$, and the cross-covariances and autocovariances between the sampling errors of the direct estimators and the benchmark.

Datta *et al*. (2010) develop Bayesian benchmarking by minimizing,

(6.14) $$\sum_{i=1}^{m} \varphi_i E[(\theta_i - \hat{\theta}_i^{bench})^2 \mid \hat{\theta}_{design}] \text{ s.t. } \sum_{i=1}^{m} b_i \hat{\theta}_i^{bench} = \sum_{i=1}^{m} b_i \hat{\theta}_{i,design},$$

where $\hat{\theta}_{design} = (\hat{\theta}_{1,design}, ..., \hat{\theta}_{m,design})'$. The solution of this minimization problem is the same as (6.13), but with $\hat{\theta}_{k,model}^{BLUP}$ replaced everywhere by the posterior mean $\hat{\theta}_{k,Bayes}$. Denote the resulting predictors by $\hat{\theta}_{i,Bayes}^{bench,1}$. The use of these predictors has the drawback of 'over shrinkage' in the sense that $\sum_{i=1}^{m} b_i (\hat{\theta}_{i,Bayes}^{bench,1} - \bar{\hat{\theta}}_{b,Bayes}^{bench,1})^2 < \sum_{i=1}^{m} b_i E[(\theta_i - \bar{\theta}_b)^2 \mid \hat{\theta}_{design}]$, where $\bar{\hat{\theta}}_{b,Bayes}^{bench,1} = \sum_{i=1}^{m} b_i \hat{\theta}_{i,Bayes}^{bench,1}$ and $\bar{\theta}_b = \sum_{i=1}^{m} b_i \theta_i$. To deal with this problem, Datta *et al*. (2010) propose to consider instead the predictors $\hat{\theta}_{i,Bayes}^{bench,2}$ satisfying,

(6.15) $$\sum_{i=1}^{m} b_i \hat{\theta}_{i,Bayes}^{bench,2} = \sum_{i=1}^{m} b_i \hat{\theta}_{i,design} \; ; \; \sum_{i=1}^{m} b_i (\hat{\theta}_{i,Bayes}^{bench,2} - \sum_{i=1}^{m} b_i \hat{\theta}_{i,design})^2 = H,$$

where $H = \sum_{i=1}^{m} b_i E(\theta_i - \bar{\theta}_b)^2 \mid \hat{\theta}_{design}]$. The desired predictors have now the form,

(6.16) $$\hat{\theta}_{i,Bayes}^{bench,2} = \sum_{i=1}^{m} b_i \hat{\theta}_{i,design} + A_{CB}(\hat{\theta}_{i,Bayes} - \bar{\hat{\theta}}_{Bayes}) \; ; \; A_{CB}^2 = H / \sum_{i=1}^{m} b_i (\hat{\theta}_{i,Bayes} - \bar{\hat{\theta}}_{Bayes})^2 .$$

Notice that the development of the Bayesian benchmarked predictors by Datta *et al*. (2010) is general and not restricted to any particular model.

6.5 Accounting for Measurement Errors in the Covariates

Ybarra and Lohr (2008) consider the case where some or all of the true covariates x_i in the area level model (5.1) are unknown and one uses instead an estimator $\hat{\mathrm{x}}_i$, with $MSE(\hat{\mathrm{x}}_i) = C_i$, in the expression for the BLUP $\hat{\theta}_i$ defined by (5.2). (For known covariates x_{ki}, $C_{ki} = 0$.) Estimates for the missing covariates may be obtained from another survey. Denoting the resulting predictor by $\hat{\theta}_i^{Err}$, it follows that,

(6.17) $$PMSE(\hat{\theta}_i^{Err}) = PMSE(\hat{\theta}_i) + (1 - \gamma_i)^2 \beta' C_i \beta ,$$

where $PMSE(\hat{\theta}_i)$ is the PMSE when using the true x_i. Thus, reporting $PMSE(\hat{\theta}_i)$ as the PMSE in this case results in under-reporting of the true PMSE. Moreover, if $\beta' C_i \beta > \sigma_u^2 + \sigma_{Di}^2$, $MSE(\hat{\theta}_i^{Err}) > \sigma_{Di}^2$, the variance of the direct estimator $\tilde{y}_i$. The authors propose therefore using instead the predictor,

(6.18) $\hat{\theta}_i^{Me} = \tilde{\gamma}_i \tilde{y}_i + (1-\tilde{\gamma}_i)\hat{\mathrm{x}}_i'\beta$; $\tilde{\gamma}_i = (\sigma_u^2 + \beta' C_i \beta)/(\sigma_{Di}^2 + \sigma_u^2 + \beta' C_i \beta)$.

The predictor $\hat{\theta}_i^{Me}$ minimizes the MSE of linear combinations of $\tilde{y}_i$ and $\hat{\mathrm{x}}_i'\beta$. $E(\hat{\theta}_i^{Me} - \theta_i) = (1-\tilde{\gamma}_i)[E(\hat{\mathrm{x}}_i) - \mathrm{x}_i]'\beta$, implying that the bias vanishes when $\hat{\mathrm{x}}_i$ is unbiased for x_i, and $E(\hat{\theta}_i^{Me} - \theta_i)^2 = \tilde{\gamma}_i \sigma_{Di}^2 \le \sigma_{Di}^2$. The authors develop estimators for the unknown parameters σ_u^2 and β, which are then substituted in (6.18) to obtain the corresponding empirical predictor. The PMSE of the empirical predictor is estimated using the jackknife procedure of Jiang *et al*. (2002) described in Section 6.1.

Torabi *et al*. (2009) consider the unit level model (5.3) but where in truth $\mathrm{x}_{ij}'\beta = \beta_0 + \beta_1 x_i$, a single covariate common to all the units in the same area. It is assumed that $x_{ij} = x_i + \eta_{ij}$; $x_i \sim N(\mu_x, \sigma_x^2)$, and that $(u_i, \varepsilon_{ij}, \eta_{ij})$ are mutually independent normally distributed random errors with zero means and variances $\sigma_u^2, \sigma_\varepsilon^2$ and σ_η^2 respectively. The sample observations consist of $\{y_{ij}, x_{ij}; i = 1,...,m,\ j = 1,...,n_i\}$. An example giving rise to such a model is where x_i defines the true level of air pollution in the area and the x_{ij} s represent pollution measures at different sites in the area, with the response values, y_{ij}, measuring a certain health indicator at the same sites. The authors follow the EB approach and show that when all the model parameters are known, the posterior distribution of the unobserved *y*-values in a given area i is multivariate normal, which yields the following predictor for $\overline{Y}_i$:

(6.19) $\hat{\overline{Y}}_{i,Bayes} = (1 - f_i A_i)\overline{y}_i + f_i A_i(\beta_0 + \beta_1 \mu_x) + f_i A_i \gamma_{xi} \beta_1 (\overline{X}_i - \mu_x)$,

where $f_i = 1 - (n_i / N_i)$ is the finite population correction in the area, $\gamma_{xi} = n_i \sigma_x^2 (\sigma_\eta^2 + n_i \sigma_x^2)^{-1}$ and $A_i = [n_i \beta_1^2 \sigma_x^2 \sigma_\eta^2 + (n_i \sigma_u^2 + \sigma_\varepsilon^2) v_i]^{-1} \sigma_\varepsilon^2 v_i$ with $v_i = (\sigma_\eta^2 + n_i \sigma_x^2)$. For large N_i and small (n_i / N_i), the PMSE of $\hat{\overline{Y}}_{i,Bayes}$ is $E[(\hat{\overline{Y}}_{i,Bayes} - \overline{Y}_i)^2 \mid \{y_{ij}, x_{ij}\}] = A_i[\beta_1^2 \sigma_x^2 + \sigma_u^2 - n_i \beta_1^2 \sigma_x^4 v_i^{-1}]$. The unknown model parameters, $\psi = (\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2)$ are estimated by a method of moments. Replacing the unknown model parameters by their estimates yields the corresponding EB estimate, which is shown to be asymptotically optimal as the number of sampled areas increases. The PMSE of the EB predictor is estimated by use of a weighted jackknife procedure proposed by Chen and Lahiri (2002).

Remark 3. In an earlier study, Ghosh *et al*. (2006) used the same model and inference procedure as Torabi *et al*. (2009), but they only used the observed y-values (and not the observed covariates) for estimating the area means under the model.

6.6 Treatment of Outliers

Bell and Huang (2006) consider the area level model (5.1) from a Bayesian perspective, but assume that the random effect or the sampling error (but not both) have a Student's $t_{(k)}$ distribution. The *t* distribution is often used in statistical modeling to account for possible outliers because of its long tails. One of the models considered by the authors is,

(6.20) $$e_i \sim N(0,\sigma_{Di}^2)\,,\; u_i \mid \delta_i,\sigma_u^2 \sim N(0,\delta_i\sigma_u^2);\; \delta_i^{-1} \sim Gamma[k/2,(k-2)/2],$$

which implies $E(\delta_i)=1$ and $u_i \mid \sigma_u^2 \sim t_{(k)}(0,\sigma_u^2(k-2)/k)$ with $Var(u_i \mid \sigma_u^2)=\sigma_u^2$. The coefficient δ_i can thus be viewed as a multiplicative random effect distributed around 1, which inflates or deflates the variance of $u_i=\theta_i-\mathrm{x}_i'\beta$. A large value δ_i signals the existence of an outlying area mean θ_i. The degrees of freedom parameter, *k*, is taken as known. Setting $k=\infty$ is equivalent to assuming the model (5.1). The authors consider several possible (small) values for *k* in their application, but the choice of an appropriate value is largely the result of data exploration. Alternatively, the authors assume the model (6.20) for the sampling error e_i, (with σ_u^2 replaced by σ_{Di}^2), in which case $u_i \sim N(0,\sigma_u^2)$. The effect of assuming the model for the random effects is to push the small area predictor (the posterior mean) towards the direct estimator, whereas the effect of assuming the model for the sampling errors is to push the predictor towards the synthetic part. The use of either model is shown empirically to perform well in identifying outlying areas but at present it is not clear how to choose between the two models. Huang and Bell (2006) extend the approach to a bivariate area level model where two direct estimates are available for every area, with uncorrelated sampling errors but correlated random effects. This model handles a situation where estimates are obtained from two different surveys.

Ghosh *et al*. (2008) likewise consider the model (5.1) and follow the EB approach. The starting point in this study is that an outlying direct estimate may arise either from a large sampling error or from an outlying random effect. The authors propose therefore to replace the EB predictor (5.2) by the robust EB predictor,

(6.21) $$\hat{\theta}_i^{Rob}=\tilde{y}_i-(1-\hat{\gamma}_i)\hat{V}_i\Psi_G[(\tilde{y}_i-\mathrm{x}_i'\hat{\beta}_{G\hat{L}S})\hat{V}_i^{-1}]\;;\; \hat{V}_i^2=\hat{Var}(\tilde{y}_i-\mathrm{x}_i'\hat{\beta}_{GLS}),$$

where $\hat{\beta}_{G\hat{L}S}$ is the empirical GLS estimator of β under the model with estimated variance $\hat{\sigma}_u^2$, and Ψ_G is the Huber influence function $\Psi_G(t) = sign(t)\min(G, |t|)$, for some given value $G > 0$. Thus, for large positive standardized residuals $(\tilde{y}_i - \mathrm{x}_i'\hat{\beta}_{G\hat{L}S})\hat{V}_i^{-1}$, the EB $\hat{\theta}_i = \tilde{y}_i - (1-\hat{\gamma}_i)\hat{V}_i(\tilde{y}_i - \mathrm{x}_i'\hat{\beta}_{GLS})\hat{V}_i^{-1}$ under the model is replaced by $\hat{\theta}_i^{Rob} = \tilde{y}_i - (1-\hat{\gamma}_i)\hat{V}_i G$, and similarly for large negative standardized residuals, whereas in other cases the EB predictor is unchanged. The value G may actually change from one area to the other and chosen adaptively in such a way that the excess Bayes risk under the model from using the predictor (6.21) is bounded by some percentage point. Alternatively, G may be set to some constant $1 \le G_0 \le 2$ as often found in the robustness literature. The authors derive the PMSE of $\hat{\theta}_i^{Rob}$ under the model with bias of order $o(1/m)$ for the case where σ_u^2 is estimated by MLE, and develop an estimator for the PMSE that is correct up to the order of $O_p(1/m)$.

Under the approach of Ghosh *et al*. (2008), the EB small area predictor (5.2) is replaced by the robust predictor (6.21) but the estimators of the unknown model parameters remain intact. Sinha and Rao (2009) propose to robustify also the estimation of the model parameters. The authors consider the general mixed linear model (6.1), which when written compactly for all the observations $y = (y_1', ..., y_m')'$ has the form,

(6.22) $\quad y = X\beta + Zu + e,\ E(u) = 0,\ E(uu') = Q;\ \ E(e) = 0,\ \ E(ee') = R,$

where $X = [\mathrm{x}_1, ..., \mathrm{x}_m]'$, u is the vector of random effects and e is the vector of residuals (or sampling errors). The matrices Q and R are block diagonal with elements that are functions of a vector parameter $\zeta = (\zeta_1, ..., \zeta_L)$ of variance components, such that, $V(y) = V = ZQZ' + R = V(\zeta)$. The model contains as special cases the models defined by (5.1) and (5.3). The target is to predict the linear combination $\tau = l'\beta + h'u$ by $\hat{\tau} = l'\hat{\beta} + h'\hat{u}$.

The MLE of β and ζ are obtained by solving the normal equations, $X'V^{-1}(y - X\beta) = 0$; $(y - X\beta)'V^{-1}\dfrac{\partial V}{\partial \zeta_l}V^{-1}(y - X\beta) - tr(V^{-1}\dfrac{\partial V}{\partial \zeta_l}) = 0,\ l = 1, ..., L$. In order to deal with possible outliers, the authors propose solving instead,

(6.23) $\ X'V^{-1}U^{1/2}\Psi_G(r) = 0;\ \Psi_G'(r)U^{1/2}V^{-1}\dfrac{\partial V}{\partial \zeta_l}V^{-1}U^{1/2}\Psi_G(r) - tr(V^{-1}\dfrac{\partial V}{\partial \zeta_l}c\mathrm{I}_n) = 0,\ l = 1, ..., L,$

where $r = U^{-1/2}(y - X\beta)$, $U = Diag[V]$, $\Psi_G(r) = [\Psi_G(r_1), \Psi_G(r_2), ...]'$ with $\Psi_G(r_k)$ defining the Huber influence function, I_n is the identity matrix of order n (the total sample size) and

$c = E[\Psi_G^2(r_k)]$ ($r_k \sim N(0,1)$). Notice that since Q and R are block diagonal, the normal equations and the robust estimating equations can be written as sums over the *m* areas. Denote by $\hat{\beta}_{Rob}, \hat{\zeta}_{Rob}$ the solutions of (6.21). The random effects are predicted by solving,

(6.24) $$Z'\hat{R}^{-1/2}\Psi_G[\hat{R}^{-1/2}(y - X\hat{\beta}_{Rob} - Zu)] - \hat{Q}^{-1/2}\Psi_G(\hat{Q}^{-1/2}u) = 0,$$

where $\hat{R} = R(\hat{\zeta}_{Rob}), \hat{Q} = Q(\hat{\zeta}_{Rob})$. Sinha and Rao (2009) estimate the PMSE of the robust predictors by application of the first step of the double-bootstrap procedure of Hall and Maiti (2006) described in Section 6.1 (Equation 6.4). All the parameter estimates and predictors of the random effects needed for the application of this procedure are computed by the robust estimating equations defined by (6.23) and (6.24).

6.7 M-quantile Estimation

Classical SAE methods under the frequentist approach model the expectations $E(y_i \mid \mathrm{x}_i, \mathrm{u}_i)$ and $E(u_i)$. Tzavidis and Chambers (2005) and Chambers and Tzavidis (2006) propose modelling instead the quantiles of the conditional distribution $f(y_i \mid \mathrm{x}_i)$, where for now y_i is a scalar and x_i is a vector of covariates. Assuming a linear model for the quantiles, this leads to a family of models indexed by the coefficient $q \in (0,1)$; $q = \Pr[y_i \le \mathrm{x}_i'\beta_q]$. In quantile regression the vector β_q is estimated by the vector $\tilde{\beta}_q$ minimizing,

(6.25) $$\sum_{i=1}^{n}\{|\, y_i - \mathrm{x}_i'\tilde{\beta}_q \,|[(1-q)\mathrm{I}(y_i - \mathrm{x}_i'\tilde{\beta}_q \le 0) + q\mathrm{I}(y_i - \mathrm{x}_i'\tilde{\beta}_q > 0)]\},$$

where $\mathrm{I}(\cdot)$ is the indicator function. *M-quantile* regression uses influence functions for the estimation of β_q by solving the estimating equations,

(6.26) $$\sum_{i=1}^{n}\Psi_q(r_{iq})\mathrm{x}_i = 0;\ r_{iq} = (y_i - \mathrm{x}_i'\beta_q),\ \Psi_q(r_{iq}) = 2\Psi(s^{-1}r_{iq})[(1-q)\mathrm{I}(r_{iq} \le 0) + q\mathrm{I}(r_{iq} > 0)],$$

where s is a robust estimate of scale, and Ψ is an appropriate influence function. The solution of (6.26) is obtained by an iterative reweighted least square algorithm (assuming a continuous monotone influence function). Denote by $\hat{\beta}_q$ the (unique) estimate solving (6.26). Notice that each sample value (y_i, x_i) lies on one and only one of the M-quantiles $m_q(\mathrm{x}_i) = \mathrm{x}_i'\beta_q$ (follows from the fact that the quantiles are continuous in q).

How is the M-quantile theory used for SAE? Suppose that the sample consists of unit level observations $\{y_{ij}, \mathrm{x}_{ij};\ i = 1,...,m,\ j = 1,...,n_i\}$. Define for unit (i, j) the value q_{ij} such that

$\mathrm{x}'_{ij}\hat{\beta}_{q_{ij}} = y_{ij}$. A small area predictor of the mean θ_i is obtained by averaging the quantiles q_{ij} over the sampled units $j \in s_i$, and then computing,

(6.27) $$\hat{\theta}_i^M = N_i^{-1}(\sum_{j\in s_i} y_{ij} + \sum_{k\notin s_i} \mathrm{x}'_{ik}\hat{\beta}_{\bar{q}_i}); \; \bar{q}_i = \sum_{j=1}^{n_i} q_{ij} / n_i \, .$$

Alternatively, one can average the vector coefficients $\beta_{q_{ij}}$ and replace the vector $\hat{\beta}_{\bar{q}_i}$ in (6.27) by the mean $\bar{\hat{\beta}}_i = \sum_{j=1}^{n_i} \hat{\beta}_{q_{ij}} / n_i$. The vectors $\hat{\beta}_{\bar{q}_i}$ or $\bar{\hat{\beta}}_i$ account for differences between the areas, similarly to the random effects under the unit level model (5.3).

The use of this approach is not restricted to the estimation of means although it does assume continuous *y*-values. Thus, the distribution function in Area i can be estimated as,

(6.28) $$\hat{F}_i(t) = N_i^{-1}[\sum_{j\in s_i} \mathrm{I}(y_{ij} \le t) + \sum_{k\notin s_i} \mathrm{I}(\mathrm{x}'_{ik}\bar{\hat{\beta}}_i \le t)] .$$

The M-quantile approach does not assume a parametric model although in the current applications it assumes that the quantiles are linear in the covariates. Clearly, if the unit level model (5.3) holds the use of the model for SAE is more efficient, but the authors illustrate that the use of the M-quantiles can be more robust to model misspecification. Notice in this regard that the approach is not restricted to any specific definition of the small areas. It accounts also for possible outliers by choosing an appropriate influence function in the estimating equation (6.26). On the other hand, there seems to be no obvious way of how to predict the means or other target quantities for nonsampled areas. A possible simple solution would be to set $q = 0.5$ for such areas, or weight the *q*-values of neighboring sampled areas, but it raises the question of how to estimate the corresponding PMSE, unless under a model.

6.8 SAE Using Penalized Spline Regression

Another way of robustifying the general mixed linear regression model defined by (6.22) or other parametric models is by use of penalized spline (P-spline) regression. The idea is not to assume apriori a functional form for the expectation of the response variable. Suppose that there is a single covariate x. The simple P-spline model assumes that $y = m_0(x) + \varepsilon$, $E(\varepsilon) = 0, Var(\varepsilon) = \sigma_\varepsilon^2$. The mean $m_0(x)$ is taken as unknown and is approximated by,

(6.29) $$m(x;\beta,\gamma) = \beta_0 + \beta_1 x + ... + \beta_p x^p + \sum_{k=1}^{K} \gamma_k (x - \mathcal{K}_k)_+^p \; ; \; (x - \mathcal{K}_k)_+^p = \max[0, (x - \mathcal{K}_k)^p] ,$$

where p is the degree of the spline and $\mathcal{K}_1 < ... < \mathcal{K}_K$ are fixed knots. For large K and a good spread of the knots over the range of *x*, the spline function approximates well most smooth

functions. The spline (6.29) uses the basis $[1, x, ..., x^p, (x-K_1)_+^p, ..., (x-K_K)_+^p]$ to approximate the mean $m_0(x)$, but other bases can be used, particularly when there are more covariates.

Opsomer *et al*. (2008) use P-spline regression for SAE by treating the γ-coefficients in (6.29) as additional random effects. Suppose as under the unit level model that the data consist of the *n* observations $\{y_{ij}, \mathrm{x}_{ij};\ i=1,...,m,\ j=1,...,n_i\}$. For unit j in area i, the model considered is,

(6.30) $$y_{ij} = \beta_0 + \beta_1 x_{ij} + ... + \beta_p x_{ij}^p + \sum_{k=1}^K \gamma_k (x_{ij} - K_k)_+^p + u_i + \varepsilon_{ij},$$

where the u_i's are the usual area random effects and the ε_{ij}'s are the residuals. Let $u = (u_1, ..., u_m)'$ and $\gamma = (\gamma_1, ..., \gamma_K)'$. Defining $d_{ij} = 1(0)$ if unit j is (is not) in area i and denoting $d_j = (d_{1j}, ..., d_{mj})'$, $D = [d_1, ..., d_n]'$, the model for the vector y of all the sample response values can be written compactly as,

(6.31) $$y = X\beta + Z\gamma + Du + \varepsilon;\ \gamma \sim (0, \sigma_\gamma^2 \mathrm{I}_k),\ u \sim (0, \sigma_u^2 \mathrm{I}_m),\ \varepsilon \sim (0, \sigma_\varepsilon^2 \mathrm{I}_n),$$

where $X = [x_1^{(p)}, ..., x_n^{(p)}]'$ with $x_l^{(p)} = (1, x_l, ..., x_l^p)'$, and $Z = [z_1, ..., z_n]'$ with $z_l = [(x_l - K_1)_+^p, ..., (x_l - K_K)_+^p]'$. Written this way the model (6.29) looks similar to the model (6.22), but notice that under (6.31) the responses y_{ij} are not independent between the areas because of the common random effects γ. Nonetheless, the BLUP and EBLUP predictors of (β, u, γ) are obtained using standard results. See the paper for the appropriate expressions. The small area predictors are then,

(6.32) $$\hat{\theta}_{i,EBLUP}^{\text{P-spline}} = \hat{\beta}' \bar{X}_i^{(p)} + \hat{\gamma}' \bar{Z}_i + \hat{u}_i;\ \bar{X}_i^{(p)} = \sum_{l \in U_i} x_l^{(p)} / N_i,\ \bar{Z}_i = \sum_{l \in U_i} z_l / N_i.$$

The use of this approach requires that the covariates are known for every element in the population. Opsomer *et al*. (2008) derive the PMSE of the EBLUP (6.32), correct to the second order, when estimating the unknown variances by restricted MLE, and an estimator of the PMSE with bias that is also correct to the second order. The authors propose also a nonparametric bootstrap algorithm for estimating the PMSE and for testing the hypotheses $\sigma_u^2 = 0$, $\sigma_\gamma^2 = 0$. Rao *et al*. (2009) use a similar model to (6.31), but rather than using the EBLUP predictors under the model, the authors propose estimators that are robust to outliers, similarly (but not the same) to the methodology developed by Sinha and Rao (2009) for the mixed linear model (6.22) described in Section 6.5. Jiang *et al*. (2010) show how to select an appropriate spline model by use of the fence method. See Section 6.9 below.

6.9. Mixed Model Selection

Model selection is one of the major problems in SAE because the models usually involve unobserved random effects with limited or no information on their distribution. Classical model selection criteria such as the AIC and BIC don't apply straightforwardly in the case of mixed models because they require the use of the likelihood, which in turn requires specification of the distribution of the random effects, and because of difficulties in determining the effective sample size. Also, for generalized linear mixed models (GLMM), the likelihood involves high dimensional integrals which makes the use of the AIC and BIC criterions unattractive.

In order to deal with this problem, Jiang *et al*. (2008) propose a class of strategies for mixed model selection called *fence methods*, which apply to linear and generalized linear mixed models. The strategies involve a procedure to isolate a subgroup of correct models, and then select the optimal model from this subgroup according to some specified criterion. Let $Q_M = Q_M(y, \psi_M)$ define a measure of 'lack of fit' of a candidate model M with parameters ψ_M, such that $E(Q_M)$ is minimized when M is the true model. Examples for Q_M are the negative of the loglikelihood and the residual sum of squares. Define, $\hat{Q}_M = Q_M(y, \hat{\psi}_M) = \inf_{\psi_M \in \Psi_M} Q_M(y, \psi_M)$, where Ψ_M is the parameter space under M, and let $\tilde{M} \in \mathrm{M}$ be such that $Q_{\tilde{M}} = \min_{M \in \mathrm{M}} \hat{Q}_M$, where M represents the set of candidate models. The authors prove that under certain conditions $\tilde{M}$ is a correct model with probability tending to one.

In practice, there can be more than one correct model and a second step of the proposed procedure is to select the optimal model among the models that are within a fence around $Q_{\tilde{M}}$. Examples of optimality criteria are minimal dimension of the model or minimum PMSE. The fence is defined as $\hat{Q}_M \leq \hat{Q}_{\tilde{M}} + c_n \hat{\sigma}_{M,\tilde{M}}$, where $\hat{\sigma}_{M,\tilde{M}}$ is an estimate of the standard deviation of $\hat{Q}_M - \hat{Q}_{\tilde{M}}$ and c_n is a tuning coefficient that increases with the total sample size, n. Jiang *et al.* (2008) discuss alternative possibilities of computing $\hat{\sigma}_{M,\tilde{M}}$ and propose an adaptive procedure for choosing the tuning coefficient c_n. The procedure consists of parametric bootstrapping new samples from the 'full' model, computing for every candidate model M the empirical proportion $p^*(M, c_n)$ that it is selected by the fence

method with a given c_n, computing $p^*(c_n) = \max_{M \in \mathrm{M}} p^*(M, c_n)$ and then choosing c_n that maximizes $p^*(c_n)$.

Jiang *et al*. (2008) illustrate the application of the fence method for selecting the covariates in the area level model (5.1), and in the unit level model (5.2) with correlated observations. Jiang *et al*. (2010) apply the fence method for selecting nonparametric P-spline models of the form (6.31). Selecting a correct model in this case requires selecting the degree of the spline, p, the number of knots, K, and a smoothing parameter, λ, used for the estimation of the model parameters.

6.10 Prediction of Ordered Area Means

Malinovsky and Rinott (2010) consider the following (hard) problem: predict the ordered area means $(\theta_{(1)} \le \theta_{(2)} \le ... \le \theta_{(m)})$ under the area level model $\tilde{y}_i = \mu + u_i + e_i = \theta_i + e_i$ (special case of 5.1), with $u_i \overset{iid}{\sim} H(0, \sigma_u^2)$, $e_i \overset{iid}{\sim} G(0, \sigma_e^2)$. H and G are general distributions with zero means and variances σ_u^2 and σ_e^2. In order to illustrate the difference between the prediction of the unordered means and the ordered means, consider the prediction of $\theta_{(m)} = \max\{\theta_i\}$. If $E(\hat{\theta}_i \mid \theta_i) = \theta_i$, then $E(\max_i \hat{\theta}_i \mid \{\theta_j\}) > \theta_{(m)}$ so that the largest estimator among the estimators of the area means overestimates the true largest mean. On the other hand, the Bayesian predictors $\theta_i^* = E[\theta_i \mid \{\hat{\theta}_j\}]$ satisfy, $\max_i\{\theta_i^*\} < E(\theta_{(m)})$, an underestimation in expectation.

Wright, Stern and Cressie (2003) considered the prediction of ordered means from a Bayesian perspective but their approach requires heavy numerical calculations and is sensitive to the choice of priors. Malinovsky and Rinott (2010) compare three simple predictors of the ordered means under the frequentist approach, using the loss function $L(\tilde{\theta}_{()}, \theta_{()}) = \sum_{i=1}^{m} (\tilde{\theta}_{(i)} - \theta_{(i)})^2$ and the Bayes risk $E[L(\tilde{\theta}_{()}, \theta_{()})]$. Let $\hat{\theta}_i$ define the direct estimator of θ_i ($\tilde{y}_i$ in the notation of Section 5.2) and $\hat{\theta}_{(i)}$ the i-th ordered estimator (statistic). The predictors compared are:

(6.33) $\tilde{\theta}_{(i)}^{(1)} = \hat{\theta}_{(i)}$; $\tilde{\theta}_{(i)}^{(2)}(\delta) = \delta\hat{\theta}_{(i)} + (1-\delta)\bar{\hat{\theta}}$, $\bar{\hat{\theta}} = \sum_{i=1}^{m} \hat{\theta}_i / m$; $\tilde{\theta}_{(i)}^{(3)} = E(\theta_{(i)} \mid \hat{\theta})$, $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_m)'$.

The mean μ is taken as unknown and estimated by $\bar{\hat{\theta}}$.

Denote by $\tilde{\theta}^{[k]}$ the predictor of the ordered true means when using the predictors $\tilde{\theta}_{(i)}^{(k)}$; $i = 1, ..., m$, $k = 1, 2, 3$, and let $\gamma = \sigma_u^2(\sigma_u^2 + \sigma_e^2)^{-1}$ be the optimal shrinkage coefficient

under the model for predicting the unordered means (Eq. 5.2). The authors derive several theoretical comparisons for the case of known variances σ_u^2 and σ_e^2. For example,

(6.34) If $\gamma \le (m-1)^2/(m+1)^2$ then $E[L(\tilde{\theta}_{()}^{[2]}(\delta),\theta_{()})] \le E[L(\tilde{\theta}_{()}^{[1]},\theta_{()})]$ for all $\gamma \le \delta \le 1$.

Noting that $\lim_{m\to\infty}[(m-1)^2/(m+1)^2] = 1$, it follows that (6.34) holds asymptotically for all γ, and the inequality $\gamma \le \delta \le 1$ implies less shrinkage of the direct estimators towards the mean. In particular, the optimal choice δ^{opt} of δ for $\tilde{\theta}^{[2]}(\delta)$ satisfies, $\lim_{m\to\infty} \delta^{opt} = \gamma^{1/2}$.

The results so far assume general distributions *H* and *G* of the random effects and the residual terms. When these distributions are normal, then for $m = 2$, $E[L(\tilde{\theta}_{()}^{[3]},\theta_{()})] \le E[L(\tilde{\theta}_{()}^{[2]}(\delta),\theta_{()})]$ for all δ. A conjecture supported by simulations is that this relationship holds also for $m > 2$ and the optimal choice δ^{opt}. However, the simulations suggest that for sufficiently large m (e.g., $m \ge 25$), $\tilde{\theta}^{[3]}$ is efficiently replaced by $\tilde{\theta}^{[2]}(\gamma^{1/2})$. The last two conclusions are shown empirically to hold also in the case where the variances σ_u^2 and σ_e^2 are unknown and replaced by method of moments variance estimators.

Remark 4. The problem of predicting the ordered means is different from ranking them, one of the famous triple-goal estimation objectives in SAE put forward by Shen and Louis (1998). The triple-goal estimation consists of producing 'good' area specific estimates, 'good' estimates of the histogram (distribution) and 'good' estimates of the ranks. See also Rao (2003). Judkins and Liu (2000) considered another related problem of estimating the range of the area means. The authors show theoretically and by simulations that the range of the direct estimators overestimates the true range, whereas the range of the empirical Bayes estimators underestimates the true range, in line with what is stated at the beginning of this section. The bias is much reduced by the use of a constrained empirical Bayes estimator. For the model considered by Malinovsky and Rinott (2010), the constrained estimator is obtained by replacing the shrinkage factor $\gamma = \sigma_u^2(\sigma_u^2 + \sigma_e^2)^{-1}$ in (5.2) by $\tilde{\gamma} \cong \gamma^{-1/2}$, which again shrinkages less the direct estimator.

7. DEVELOPMENTS FOR SPECIFIC APPLICATIONS

Below I review briefly a few other important developments for specific applications.

1. *Small area estimation with varying boundaries*: Moura *et al*. (2005) investigate the use of hierarchical models for SAE with varying area boundaries. The paper shows how area estimates and corresponding PMSE estimates can be obtained at a variety of nested and

intersecting boundary systems by fitting the model at the lowest possible level. See also Section 6.7 on M-quantile estimation.

2. *Accounting for Spatial correlations between the area means*: The models considered in this article assume that the area random effects are independent between the areas. Noting that the random effects account for the unexplained variation by the existing covariates, it is often reasonable to assume that the random effects in neighbouring areas are correlated. See Pratesi and Salvati (2008) for the spatial EBLUP predictors and estimation of their PMSE when the spatial relationships are modelled by a simultaneously autoregressive process with a known weight matrix and an unknown autoregression parameter. Souza *et al*. (2009) account for spatial relationships when fitting hierarchical exponential growth models, which are used for small area population predictions by application of the Bayesian methodology.

3. *SAE estimation under informative sampling*: All the studies reviewed in this paper assume implicitly that the selection of the sampled areas is noninformative, and similarly for the sampling within the selected areas, implying that the population model applies to the observed sample data with no selection bias. This, however, may not be the case and ignoring the effects of informative sampling may bias the inference very severely. This problem is considered by Malec *et al*. (1999) under the Bayesian approach and by Pfeffermann and Sverchkov (2007) under the frequentist approach.

4. *The use of a two-part model for SAE*: It is sometimes the case that the outcome value is either zero or an observation from a continuous distribution. A typical example is the assessment of literacy with the possible response being either zero, indicating illiteracy, or a positive score measuring the level of literacy. Another example is the consumption of illicit drugs. Pfeffermann *et al*. (2008) consider the estimation of the average response and the proportion of positive responses in small areas for this kind of problem by joint modelling the probability of a positive response and the distribution of the positive responses and applying the Bayesian methodology.

8. CONCLUDING REMARKS

In this article I reviewed some of the main important developments in SAE in recent years. Looking at these developments gives a clear indication that analysts involved in SAE problems have now much richer and versatile tools for their analysis, notably in terms of the robustness of the inference to possible model misspecifications, measurement errors and outliers. Research on SAE continues all over the world, both in terms of new theories and in

applications to new intriguing problems, and it will not take long before another book or review paper will come out discussing this new research, not covered in the present paper.

REFERENCES

BATTESE, G. E., HARTER, R. M. and FULLER, W. A. (1988). An error components model for prediction of county crop area using survey and satellite data. *Journal of the American Statistical Association* **83** 28–36.

BELL, W. R. and HUANG, E. T. (2006). Using the *t*-distribution to deal with outliers in small area estimation. Proceedings of Statistics Canada Symposium on Methodological Issues in Measuring Population Health.

CHAMBERS, R. and TZAVIDIS, N. (2006). M-quantile models for small area estimation. *Biometrika* **93** 255-268.

CHANDRA, H. and CHAMBERS, R. (2009). Multipurpose small area estimation. *Journal of Official Statistics* **25** 379–395.

CHATTERJEE, S., LAHIRI, P. and Li, H. (2008). Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models. *The Annals of Statistics* **36** 1221-1245.

CHEN, S. and LAHIRI, P. (2002). On mean squared prediction error estimation in small area estimation problems. In Proceedings of the Survey Research Methods Section, American Statistical Association. pp. 473-477.

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.

DATTA, G. S., RAO, J. N. K. and SMITH, D. D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika* **92** 183–196.

DATTA, G. S. (2009). Model-Based Approach to Small Area Estimation. In: Handbook of Statistics 29B; *Sample Surveys*: *Inference and Analysis*. Eds. D. Pfeffermann and C.R. Rao. North Holland. pp. 251-288.

DATTA, G. S., GHOSH, M., STEORTS, R. and MAPLES, J. (2010). Bayesian Benchmarking with Applications to Small Area Estimation. *The Canadian Journal of Statistics* **38**. To appear.

ESTEVAO, V. M. and SÄRNDAL, C. E. (2004). Borrowing strength is not the best technique within a wide class of design-consistent domain estimators. *Journal of Official Statistics* **20** 645–669.

ESTEVAO, V. M. and SÄRNDAL, C. E. (2004). Survey estimates by calibration on complex auxiliary information. *International Statistical Review* **74** 127-147.

FALORSI, P. D. and RIGHI, P. (2008). A balanced sampling approach for multi-way stratification designs for small area estimation. *Survey Methodology* **34** 223-234.

FAY, R. E. and HERRIOT, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association* **74** 269–277.

GANESH, N. and LAHIRI, P. (2008). A new class of average moment matching prior. *Biometrika* **95** 514-520.

GHOSH, M., SINHA, K. and KIM, D. (2006). Empirical and hierarchical Bayesian estimation in finite population sampling under structural measurement error models. *Scandinavian Journal of Statistics* **33** 591-608.

GHOSH, M., MAITI, T. and ROY, A. (2008). Influence functions and robust Bayes and empirical Bayes small area estimation. *Biometrika* **95** 573 -585.

HALL, P. and MAITI, T. (2006b). On parametric bootstrap methods for small area prediction. *Journal of the Royal Statistical Society Series B* **68** 221- 238.

HUANG, E. T. and BELL, W. R. (2006). Using the *t*-distribution in small area estimation: an application to SAIPE state poverty models. In Proceedings of the Survey Research Methods Section, American Statistical Association. pp. 3142-3149.

ISAKI, C. T., TSAY, J. H. and FULLER, W. A. (2000). Estimation of census adjustment factors. *Survey Methodology* **26** 31-42.

JIANG, J., LAHIRI, P. S. and WAN, S. M. (2002). A unified jackknife theory for empirical best prediction with M-estimation. *The Annals of Statistics* **30** 1782–1810.

JIANG, J. and LAHIRI, P. S. (2006a). Estimation of finite population domain means: a model-assisted empirical best prediction approach. *Journal of the American Statistical Association* **101** 301–311.

JIANG, J. and LAHIRI, P. S. (2006b). Mixed model prediction and small area estimation. *Test* **15** 1–96.

JIANG, J., RAO, J. S., GU, Z. and NGUYEN, T. (2008). Fence methods for mixed model selection. *The Annals of Statistics* **36** 1669-1692.

JIANG, J., NGUYEN, T. and RAO, J. S. (2010). Fence method for nonparametric small area estimation. *Survey Methodology* **36**. To appear.

KOTT, P. S. (2009). Calibration Weighting: Combining Probability Samples and Linear Prediction Models. In: Handbook of Statistics 29B; *Sample Surveys*: *Inference and Analysis*. Eds. D. Pfeffermann and C.R. Rao. North Holland. pp. 55-82.

LAHIRI, P. S. and RAO, J. N. K. (1995). Robust estimation of mean squared error of small area estimators. *Journal of the American Statistical Association* **90** 758–766.

LEHTONEN, R., SÄRNDAL, C. E. and VEIJANEN, A. (2003). The effect of model choice in estimation for domains, including small domains. *Survey Methodology* **29** 33–44.

LEHTONEN, R., SÄRNDAL, C. E. and VEIJANEN, A. (2005). Does the model matter? comparing model-assisted and model-dependent estimators of class frequencies for domains. *Statistics in Transition* **7** 649–673.

LEHTONEN, R. and VEIJANEN, A. (2009). Design-based Methods of Estimation for Domains and Small Areas. In: Handbook of Statistics 29B; *Sample Surveys*: *Inference and Analysis*. Eds. D. Pfeffermann and C.R. Rao. North Holland. pp. 219-249.

LOHR, S. L. and RAO, J. N. K. (2009). Jackknife estimation of mean squared error of small area predictors in nonlinear mixed models. *Biometrika* **96** 457-468.

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. *Journal of the American Statistical Association* **92** 815–826.

MALEC, D., DAVIS, W. W. and CAO, X. (1999). Model-based small area estimates of overweight prevalence using sample selection adjustment. *Statistics in Medicine* **18** 3189–3200.

MALINOVSKY, Y. and RINOTT, Y. (2010). Prediction of ordered random effects in a simple small area model. *Statistica Sinica*

MOURA, F., CLARKE, P. and PFEFFERMANN, D. (2005). Small area estimation under varying area boundaries using the synthetic estimator. *Statistics in Transition* 7 625-636.

OPSOMER, J. D., CLAESKENS, G., RANALLI, M. G., KAUEMANN, G. and BREIDT, F. J. (2008). Non-parametric small area estimation using penalized spline regression. *Journal of the Royal Statistical Society*, series B **70** 265-286.

PFEFFERMANN, D. (2002). Small Area Estimation- New Developments and Directions. I*nternational Statistical Review* **7**0 125-143.

PFEFFERMANN, D. and TILLER, R. (2006). Small Area Estimation with State-Space Models Subject to Benchmark Constraints. *Journal of the American Statistical Association* **101** 1387-1397.

PFEFFERMANN, D. and SVERCHKOV, M. (2007). Small-area estimation under informative probability sampling of areas and within the selected areas. *Journal of the American Statistical Association* **102** 1427-1439.

PFEFFERMANN, D., TERRYN B. and MOURA, F. A. S. (2008). Small area estimation under a two-part random effects model with application to estimation of literacy in developing countries. *Survey Methodology* **34** 235-249.

PRASAD, N. G. N. and RAO, J. N. K. (1990). The estimation of mean squared errors of small area estimators. *Journal of the American Statistical Association* **85** 163–171.

PRATESI, M. and SALVATI, N. (2008). Small area estimation: the EBLUP estimator based on spatially correlated random area effects. *Statistical Methods and Application* 17 113-141.

RAO, J. N. K. (2003). *Small area estimation*. John Wiley & Sons, New York.

RAO, J. N. K., SINHA, S. K. and ROKNOSSADATI M. (2009). Robust small area estimation using penalized spline mixed models. In: Proceedings of the Survey Research Methods Section, American Statistical Association. pp. 145-153.

SHEN, W. and LOUIS, T. A. (1998). Triple-goal estimates in two-stage hierarchical models. *Journal of the Royal Statistical Society*, Series B **60** 455-471.

SINHA, S. K. and RAO, J. N. K (2009). Robust small area estimation. *The Canadian Journal of Statistics* **37** 381-399.

SOUZA, D. F., MOURA, F. A. S. and MIGON, H. S. (2009). Small Area Population Prediction via Hierarchical Models. *Survey Methodology* **35** 203-214.

TZAVIDIS, N. and CHAMBERS, R. (2005). Bias adjusted estimation for small areas with M-quantile models. *Statistics in Transition* **7** 707-713.

TORABI, M. and Rao, J. N. K. (2008). Small area estimation under a two-level model. *Survey Methodology* **34** 11–17.

TORABI, M., DATTA, G. S. and RAO, J. N. K. (2009). Empirical Bayes estimation of small area means under a nested error linear regression model with measurement errors in the covariates. *Scandinavian Journal of Statistics* **36** 355-368.

WANG, J., FULLER, W. A. and QU, Y. (2008). Small area estimation under a restriction. *Survey Methodology* **34** 29-36.

WRIGHT, D. L., STERN, H. S. and CRESSIE, N. (2003). Loss function for estimation of extreme with an application to disease mapping. *The Canadian Journal of Statistics* **31** 251-266.

YBARRA, L. M. R. and LOHR, S. L. (2008). Small area estimation when auxiliary information is measured with error. *Biometrika* **95** 919-931.

YOU, Y. and RAO, J. N. K. (2002). A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *The Canadian Journal of Statistics* **30** 431-439.