ANNUAL REVIEWS

Annual Review of Statistics and Its Application

Robust Small Area Estimation: An Overview

Jiming Jiang[1] and J. Sunil Rao[2]

[1]Department of Statistics, University of California, Davis, California 95616, USA; email: jimjiang@ucdavis.edu

[2]Department of Public Health Sciences, University of Miami, Miami, Florida 33136, USA

Annu. Rev. Stat. Appl. 2020. 7:337–60

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

https://doi.org/10.1146/annurev-statistics-031219-041212

Keywords

borrowing strength, empirical best linear unbiased prediction, empirical best prediction, method of moments, model failure, model misspecification, model selection, mean squared prediction error, nonparametric, observed best prediction, outliers, robustness, small area estimation

Abstract

A small area typically refers to a subpopulation or domain of interest for which a reliable direct estimate, based only on the domain-specific sample, cannot be produced due to small sample size in the domain. While traditional small area methods and models are widely used nowadays, there have also been much work and interest in robust statistical inference for small area estimation (SAE). We survey this work and provide a comprehensive review here. We begin with a brief review of the traditional SAE methods. We then discuss SAE methods that are developed under weaker assumptions and SAE methods that are robust in certain ways, such as in terms of outliers or model failure. Our discussion also includes topics such as nonparametric SAE methods, Bayesian approaches, model selection and diagnostics, and missing data. A brief review of software packages available for implementing robust SAE methods is also given.

1. INTRODUCTION AND BRIEF OVERVIEW OF SMALL AREA ESTIMATION METHODS

Borrowing strength: using information from other sources, such as small areas, other variables, or additional knowledge, to improve SAE estimates

In recent years there has been substantial and growing interest in small area estimation (SAE), driven largely by practical demands. Here, the term small area typically refers to a subpopulation or domain of interest for which a reliable direct estimate, based only on the domain-specific sample, cannot be produced due to small sample size in the domain. Examples of small areas include geographical regions (e.g., state, county, municipality), demographic groups (e.g., specific age×sex×race groups), demographic groups within geographic regions, etc. Such small areas are often of primary interest, for example, in policy-making regarding allocation of resources to subgroups, or determination of subgroups with specific characteristics (e.g., in health and medical studies) in a population. It is desirable that the decisions regarding such policy-making be made based on reliable estimates, and the demand for and interest in SAE research have increased rapidly in recent years. For example, SAE is now routinely used for effective planning of health, social, and other services and for apportioning government funds in the United States, Canada, and many European countries. Since 2013, there has been an annual international conference on SAE and related topics in various locations worldwide. Reviews on SAE and related topics have been provided by, for example, Jiang & Lahiri (2006), Datta (2009), Pfeffermann (2013), and Rao & Molina (2015).

1.1. Simple Direct and Indirect Estimates

Direct estimates use summary statistics based on a given domain, or small area, to estimate a characteristic of interest associated with the small area. The typical characteristics of interest are small area means, proportions, and totals; the corresponding summary statistics are sample means, sample proportions, and products of sample mean and domain size (i.e., population size of the small area), assuming that the latter is known (see, e.g., chapter 2 of Rao & Molina 2015). The direct estimates do not borrow strength in the sense that the estimate for a given domain does not utilize information from other domains; it also does not make use of information from other sources or variables.

In contrast, indirect estimators can utilize information from other domains or sources, which is known as borrowing strength. Here we talk about indirect estimators without extensive use of statistical models, and leave the model-based methods to later discussion. Methods of indirect estimation include synthetic estimation, composite estimation, and shrinkage estimation (see Rao & Molina 2015, chapter 3, for details). To illustrate with a simple example, suppose that the domain sizes under poststratification (i.e., the strata in the population are formed after the samples are taken; e.g., Lohr 2010, section 4.4) are available, say, N_{ig}, $1 \le i \le m$, $1 \le g \le G$, where i represents the domain and g the poststratum. Also suppose that an estimate of the poststratum total, $\hat{Y}_{\cdot g}$, is available for $1 \le g \le G$. Let $N_{\cdot g} = \sum_{i=1}^{m} N_{ig}$ be the poststratum size. Then, a synthetic estimator of the domain total is given by $\hat{Y}_i = \sum_{g=1}^{G} (N_{ig}/N_{\cdot g})\hat{Y}_{\cdot g}$. It is clear that $\hat{Y}_i$ is a weighted average of estimators of the poststratum totals, where the weights depend on the domain. Furthermore, the same poststratum total estimators are used in all of the domain total estimators, the only difference being the weights; this way, different domains can borrow strength from each other.

1.2. Basic Small Area Estimation Models

According to Jiang & Lahiri (2006) and Pfeffermann (2013), there are three basic SAE models in the sense that other models may be viewed as extensions, or variations, of these models. The first is the Fay–Herriot model (Fay & Herriot 1979), also known as the area-level model; the second

is the nested error regression (NER) model (Battese et al. 1988), also known as the unit-level model; the third is the mixed logistic model (Jiang & Lahiri 2001), which is often used for binary outcomes or binomial proportions.

A Fay–Herriot model may be expressed as $y_i = x_i'\beta + v_i + e_i$, $i = 1, \ldots, m$, where m is the total number of small areas (for which data are available); y_i is a direct survey estimator for the ith small area; x_i is a vector of associated covariates, or predictors; β is a vector of unknown regression coefficients; v_i is an area-specific random effect that accounts for variation not explained by the predictors; and e_i is a sampling error. It is assumed that $v_i, e_i, i = 1, \ldots, m$ are independent such that $v_i \sim N(0, A)$, $e_i \sim N(0, D_i)$, where A is an unknown variance but D_i is assumed known, $1 \le i \le m$. In practice, D_i may not be exactly known but can be estimated with a high degree of accuracy. For example, typically, D_i can be expressed as an unknown variance divided by the sample size for the ith area, and the unknown variance can be estimated using a (much) larger data set; Rao & Molina (2015, chapter 6) provide further explanation. Thus, in the following, we use the notation D_i with the understanding that in most cases it is $\hat{D}_i$, an estimate of D_i.

An NER model can be expressed as $y_{ij} = x_{ij}'\beta + v_i + e_{ij}$, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$, where m is the same as above, n_i is the number of units sampled from the ith small area, y_{ij} is the jth sampled outcome measure from the ith area, and x_{ij} is a corresponding vector of auxiliary variables. The meanings of β and v_i are the same as in the Fay–Herriot model, and e_{ij} is an additional error. It is assumed that the v_is and e_{ij}s are independent with $v_i \sim N(0, \sigma_v^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$, where σ_v^2 and σ_e^2 are unknown variances.

As for the mixed logistic model, it is assumed that, given $v = (v_i)_{1 \le i \le m}$ where the v_is have the same meaning as above, binary responses $y_{ij}, i = 1, \ldots, m, j = 1, \ldots, n_i$ are conditionally independent such that $\text{logit}(p_{ij}) = x_{ij}'\beta + v_i$, with x_{ij}, β having the same meanings as in the NER model, $\text{logit}(p) = \log\{p/(1-p)\}$, and $p_{ij} = \text{P}(y_{ij} = 1|v)$. Note that the x_{ij}s are considered nonrandom here, so the latter conditional probability is the same as $\text{P}(y_{ij} = 1|x_{ij}'\beta + v_i)$ due to the independence of the v_is. Furthermore, it is assumed that v_i is distributed as $N(0, \sigma^2)$ with σ^2 unknown.

The assumptions underlying these models are considered strong in that they completely specify the underlying distribution of the data. Such assumptions would allow, for example, maximum likelihood (ML) or restricted maximum likelihood (REML) inference (e.g., Jiang 2007), but the latter may not be robust when the assumptions fail. For example, for computing the empirical best linear unbiased predictor (EBLUP; see below), one can use other types of consistent estimators of variance components than ML or REML estimators (e.g., Prasad & Rao 1990), but measures of uncertainty are more sensitive to the distributional assumptions; this is discussed further below.

1.3. Traditional Model-Based Inference

We refer to Rao & Molina (2015) for details of traditional methods of inference for small areas. A mainstream approach relies on using a statistical model in order to borrow strength. These models likely involve area-specific random effects, specification of the conditional mean and variance given the random effects, and normality assumptions about the random effects and other additional errors that are involved in the model. The model-based approach can be non-Bayesian or Bayesian. These approaches lead to the EBLUP, empirical best predictor (EBP), empirical Bayes (EB) and hierarchical Bayes (HB) estimators, including their variations. The EB approach is different from the HB approach in the way that hyperparameters at the bottom of the model hierarchy are estimated from the data in the former, while a prior will be assigned to the hyperparameters in the latter. Note that there are also design-based approaches that do not use any model in deriving the estimators, such as the direct survey estimators (see Section 1.1).

Consider, for example, the Fay–Herriot model. The small area means can be expressed as $\theta_i = x_i'\beta + v_i, 1 \le i \le m$. The EBLUP of θ_i can be expressed as

$$\hat{\theta}_i = (1 - \hat{B}_i)y_i + \hat{B}_i x_i'\hat{\beta}, \qquad 1.$$

where $\hat{B}_i = D_i/(\hat{A} + D_i)$, and $\hat{\beta}, \hat{A}$ are estimators of β, A, respectively. Expression 1 shows that the EBLUP is a weighted average of the direct estimator, y_i, and an indirect regression estimator, $x_i'\hat{\beta}$, with weights $(1 - \hat{B}_i)$ and $\hat{B}_i$, respectively. Alternatively, the EBLUP may be viewed as shrinking the direct estimator toward the regression estimator with the shrinkage factor $\hat{B}_i$. To understand the weights or shrinkage factor, note that if A is large compared with D_i, which means that the between-area variation is large, there is not much strength that the direct estimator can borrow from other areas through the regression estimator; as a result, the shrinkage factor is expected to be close to zero, meaning little or no shrinkage. In contrast, if A is small compared with D_i, the between-area variation is small; therefore, there is much strength that the direct estimator can borrow from other areas, and as a result, the shrinkage factor is close to one, meaning substantial shrinkage. The following example illustrates the small area model in action.

1.3.1. Example: hospital kidney transplant graft failure rates. Morris & Christiansen (1995) presented a data set involving 23 hospitals (out of a total of 219 hospitals) that had at least 50 kidney transplants during a 27-month period. Specifically, the y_is are graft failure rates for kidney transplant operations, that is, y_i = number of graft failures/n_i, where n_i is the number of kidney transplants at hospital i during the period of interest. The variance for graft failure rate, D_i, is approximated by $(0.2)(0.8)/n_i$, where 0.2 is the observed failure rate for all hospitals. Thus, the D_is are treated as known. A severity index, x_i, is considered as a covariate. Ganesh (2009) proposed a Fay–Herriot model as $y_i = \beta_0 + \beta_1 x_i + v_i + e_i$ to fit the data.

A graphic illustration of the model in action is shown in **Figure 1**. In panel *a*, the direct estimate and EBLUP are plotted against $\sqrt{D_i}$; in panel *b*, the absolute difference of the direct estimate and EBLUP is plotted against $\sqrt{D_i}$. A lowess smoother running through panel *b* shows that this absolute difference increases as D_i increases, or, in other words, as D_i increases, the EBLUP estimator is increasingly affected by underlying small area model. Also note that the estimators $\hat{\beta}$ and $\hat{A}$ are based on data from all of the areas; this is how the EBLUP for one area borrows strength from other areas. In fact, the weights in Equation 1 are optimal when the $\hat{A}$ is replaced by A, the true variance of the random effects. Therefore, assuming that $\hat{A}$ is a consistent estimator of A, EBLUP borrows strength in a way that is nearly optimal. The standard estimators of β and A include REML, ML (e.g., Jiang 2007), or ANOVA (e.g., Prasad & Rao 1990) estimators. As noted earlier, typically D_i is also estimated.

In addition to point estimates of small-area characteristics, researchers have also extensively studied measures of uncertainty, which routinely accompany the point estimates. A standard measure of uncertainty is the mean squared prediction error (MSPE). There are other measures of uncertainty such as prediction intervals. The standard methods of estimation of area-specific MSPE include the Prasad-Rao linearization method and resampling methods (see Rao & Molina 2015, chapters 6 and 7).

1.3.2. The need for robust methods. As is always said, there is no free lunch; if SAE is such a good thing, exactly what part of it is not free? Consider, for example, model-based SAE (e.g., Jiang & Lahiri 2006, Datta 2009), which typically borrows strength through a statistical model. What if the assumed model fails? Quite often, there is a consequence. This is a price that one has to pay for doing better when the assumed model holds, in the sense that one may actually

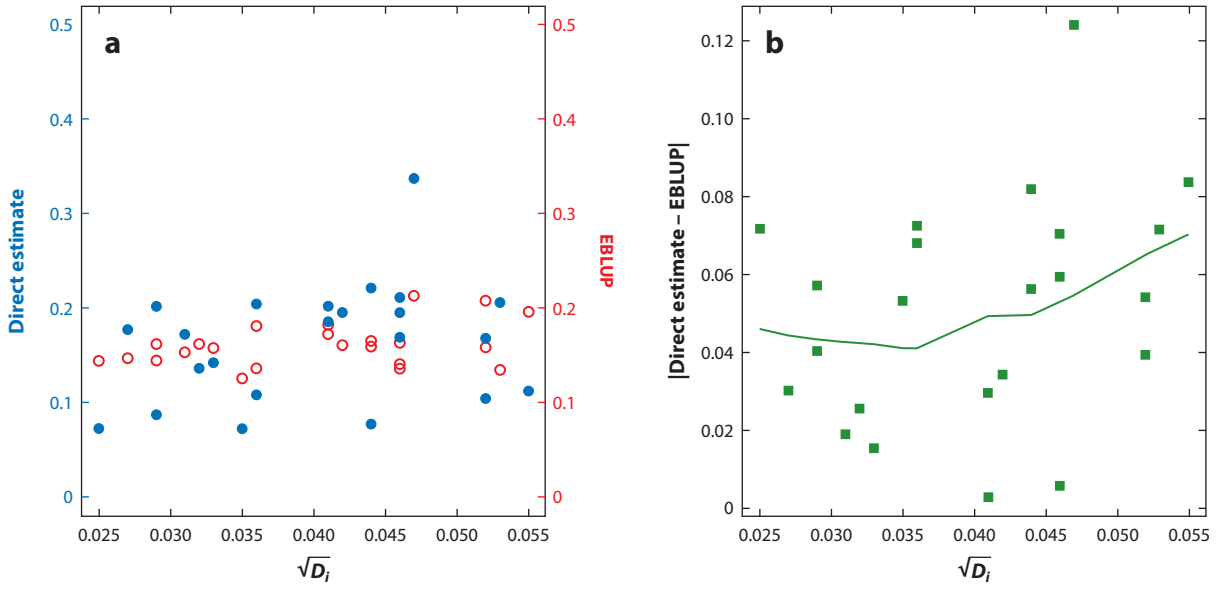

Figure 1

Illustration of a small area model in action. (*a*) Direct estimate, empirical best linear unbiased predictor (EBLUP) versus $\sqrt{D_i}$. (*b*) Absolute difference between direct estimate and EBLUP versus $\sqrt{D_i}$.

do worse than the direct estimates when the assumed model fails. Even if the assumed model holds, or in design-based SAE that is model free, one may still encounter the so-called outliers. Here, the term "design-based" refers to methods of inference based on random sampling from finite populations only (e.g., Rao & Molina 2015, section 2.2). Such problems may occur in every subject area of statistics, which has led to a well-developed field of robust statistics (e.g., Huber & Ronchetti 2009). This field is the main topic of the current review, with focus on robust methods in SAE. There are features of SAE unlike any other, and naturally, these are highlighted.

Below is a brief summary of the specific topics. In Section 2 we discuss SAE methods based on weaker model assumptions that either do not fully specify the underlying distribution of the data or consider a broader class of models that are more likely to hold. Methods developed under such weaker assumptions tend to be more robust against certain types of model failures. In Section 3 we discuss SAE methods that are more robust to surprises—unexpected events, such as model misspecification, outliers, or something else. We dedicate Section 4 to recent advances in nonparametric SAE, although there is some overlap between this and the topics discussed in other sections. In Section 5, we discuss other topics that are related, or potentially related, to robust SAE. These include model selection and diagnostics, Bayesian methods, and missing data. Some available software packages for implementing robust SAE methods are summarized in Section 6. We conclude with some remarks in Section 7.

2. METHODS DEVELOPED UNDER WEAKER ASSUMPTIONS

2.1. Relaxing Model Assumptions

Ghosh & Lahiri (1987) conducted early work on robust SAE. They were concerned about what they called mean robust estimators of strata means, with small area means being a special case. The authors derived the EB estimator under the posterior linearity assumption, meaning that under

the Bayesian framework, the posterior mean of the small area mean, θ_i, is a linear function of y_i, the vector of sampled responses from the ith small area. In a non-Bayesian setting, this simply means that the BP of θ_i is a linear function of y_i, assuming that all of the parameters are known (e.g., Rao & Molina 2015, section 9.9.2). This is weaker than assuming that the data are normal.

As the normality assumption is important, Jiang & Torabi (J. Jiang & M. Torabi 2018, unpublished technical report) developed a goodness-of-fit test for checking such an assumption. The method extends an earlier idea of Fisher (1922) and has a connection with the specification test in generalized method of moments (MoM) (e.g., Newey 1985). The test is guaranteed to have a χ^2 asymptotic null distribution. What is more, the method has a robust feature in the sense that it works correctly in testing a certain aspect of the model while some other aspect of the model may be misspecified.

In another recent development, Jiang and colleagues (J. Jiang, N.S. Matloff & T. Nguyen 2018, unpublished technical report) proposed a robust SAE method based on regression average (RA) (Matloff 1981). To describe the idea of RA, suppose that one has random samples $(x_i, y_i), i = 1, \ldots, n$ from a $(p+1)$-dimensional distribution, where $p = \dim(x_i)$. Suppose that the regression function (not necessarily linear), $\mathrm{E}(y_i|x_i) = g(x_i, \beta)$, is known except for the vector of unknown parameters, β. Suppose that $(x_i, y_i), i = 1, \ldots, n$ are independent and identically distributed (i.i.d.). The problem of interest is to estimate the mean of the outcome variable, y_i, that is, $\mu = \mathrm{E}(y_i)$. A well-known estimator of μ is the sample mean, $\bar{y} = n^{-1} \sum_{i=1}^n y_i$. However, Matloff (1981) showed that the estimator $\hat{\mu} = n^{-1} \sum_{i=1}^n g(x_i, \hat{\beta})$, where $\hat{\beta}$ is a weighted least squares estimator of β, is better than $\bar{y}$ in the sense that the asymptotic variance of $\hat{\mu}$ is smaller than that of $\bar{y}$. The improvement of $\hat{\mu}$ over $\bar{y}$ takes place as long as g is not linear; when g is linear, the asymptotic variances of the two estimators are the same. A notable feature of the RA method is that it relies on weaker assumptions.

Bell & Huang (2006) assumed, instead of normality, nonstandardized t-distributions for the random effects or sampling errors in the area-level model. A random variable, ξ, has a nonstandardized t-distribution if there are constants $\mu \in R$ and $\sigma > 0$ such that $(\xi - \mu)/\sigma$ has a t-distribution. Note that the t-distribution may be viewed as a weaker assumption than the normal distribution in that the latter requires the degrees of freedom of the t-distribution to be large (or infinity). The authors' main concern was outliers, and they proposed a Bayesian approach under the t-distribution. Gershunskaya & Lahiri (2018) noted that the NER model is sensitive to outliers in that a small portion of extreme observations may cause problems in estimation of the model parameters as well as in prediction of the mixed effects. The authors took an approach in modeling the outlying process via a mixture distribution. The underlying idea is that these outliers occur due to some distributional reasons; this is different from some outlier situations that are treated in the next section. The authors also proposed a test for the outlying area; based on the test result, different SAE strategies are used for different areas. The mixture model may be viewed as an extension of the NER model by relaxing the constant-variance assumption for the errors (i.e., e_{ij}). The authors also made comparisons with several other methods that were developed based on robustness considerations (see Section 4 for further discussion).

In addition to relaxing the constant-variance assumption in the NER model, Jiang & Nguyen (2012) considered a heteroscedastic nested error regression (HNER) model with completely unknown within-area variances. The authors noted that the NER model can be reparametrized as $\mathrm{var}(e_{ij}) = \sigma^2$ and $\mathrm{var}(v_i)/\mathrm{var}(e_{ij}) = \gamma$, where σ^2 and γ are unknown variance components, and that $\rho = \mathrm{cor}(y_{ij}, y_{ik}) = \gamma/(1+\gamma)$. The HNER model allows σ^2 to be area-specific, that is, $\sigma^2 = \sigma_i^2$ for area i, so that the standardized observations, defined as $z_{ij} = (y_{ij} - x_{ij}'\beta)/\sigma_i$, satisfy an NER model with mean 0. The latter fact is an interpretation of the assumption that γ does not depend on i. The nonstandardized observations then satisfy the HNER model with constant γ but

area-specific σ^2. Further details are provided in section 7.6.1 of Rao & Molina (2015). The unknown variances σ_i^2, $1 \le i \le m$, cannot be consistently estimated if the n_is are bounded; however, Jiang & Nguyen (2012) showed that the maximum likelihood estimators (MLEs) of β and γ (ρ) are consistent under the HNER model. Note that β and γ (or β and ρ) are necessary for computing the EBLUP. In contrast, even in a very simple case with unbalanced data and heteroscedasticity, the MLE of γ, or ρ, obtained under the NER model is inconsistent. Note that HNER is weaker than NER. In simulation studies, the authors showed that on the one hand, even if the NER model actually holds, as long as the sample size n_i is not too small, there is not much difference in performance between the HNER and NER based EBLUPs; on the other hand, if the NER model does not hold but the HNER model holds, there is a substantial difference between the two in favor of the HNER based EBLUP. Sugasawa & Kubokawa (2017) considered a variation of the HNER model. Instead of assuming the σ_i^2s are completely unknown, the authors considered modeling the σ_i^2s; they also relaxed the normality assumption about the random effects and errors.

Prasad-Rao method: linearization method for estimating MSPE of the EBLUP of a mixed effect, such as a small area mean

2.2. Robust Methods for Mean Squared Prediction Error Estimation

In the Prasad-Rao method (Prasad & Rao 1990), the variance components involved in a linear mixed model (LMM) (e.g., Jiang 2007, chapter 1) are estimated using the MoM, also known as the analysis of variance (ANOVA) method (these terms are used interchangeably hereafter). The EBLUP is then obtained with the Prasad-Rao estimators of the variance components. The initial consideration of Prasad & Rao (1990) seemed to be simplicity—the ANOVA estimators have closed-form expressions, unlike the ML or REML estimators, which makes it easier to derive and justify a second-order unbiased MSPE estimator. The bias of the MSPE estimator is therefore $o(m^{-1})$, where m is the total number of small areas from which data are available. However, there is a bonus to the Prasad-Rao method, that is, the ANOVA method requires specification of only up to the second moments of the data; in particular, the normality assumption is not needed. In fact, Lahiri & Rao (1995) showed that, under the Fay–Herriot model, the Prasad-Rao MSPE estimator is robust to nonnormality of the small area-specific random effects, although the normality of the sampling errors is still needed.

In terms of resampling methods for MSPE estimation, Jiang et al. (2002) proposed a jackknife (Quenouille 1949) method for estimating the MSPE of EBP. The method is especially convenient to implement under the posterior linearity assumption, which is weaker than assuming that the data are normal, as noted earlier (Ghosh & Lahiri 1987). It should be noted that the Jiang et al. (2002) jackknife is not restricted to work under the posterior linearity assumption; the latter just makes the computation easier, because the MSPE of the BP then has an analytic expression. Later, Jiang et al. (2018) extended the Jiang et al. (2002) jackknife to cases where the MSPE of the BP does not have a analytic form. A different approach was taken by Hall & Maiti (2006), who proposed a bootstrap method for the NER model that does not require complete specification of the distribution of the random effects and errors; more detail is provided in Section 4.

In many cases the data for the outcome variable are binary or counts. In such cases an LMM may not be appropriate; instead, a generalized linear mixed model (GLMM) (e.g., Jiang 2007, chapter 3) may be used (e.g., Ghosh et al. 1998). In particular, Jiang & Lahiri (2001) proposed an EBP approach and extended the Prasad-Rao method of MSPE estimation to SAE with binary data; see also Jiang (2003). In those papers, the authors used the MoM estimators of the parameters under a GLMM, instead of the ML estimators, which are computationally challenging to obtain (e.g., Jiang 2007, section 4.1; Torabi 2012). Furthermore, the MoM relies on weaker assumptions than GLMM. In fact, for the MoM estimators to be consistent, one only requires that the conditional mean function of the response, given the random effects, is correctly

specified; an MoM estimator with improved efficiency can be obtained by correctly specifying the conditional first two moments (Jiang & Zhang 2001). It should be noted, however, that the MoM estimator is not robust to misspecification in the distribution of the random effects, unless the GLMM is a LMM (e.g., Jiang & Nguyen 2009).

3. SMALL AREA ESTIMATION METHODS THAT ARE MORE ROBUST IN A CERTAIN WAY

In practice, a particular statistical method is chosen for various reasons, which may be scientific, economic, or political; once a method is adopted, it is often not easy to make substantial changes because of the original considerations. Thus, a practitioner often has little alternative but to proceed with using the adopted method.

What if an unexpected situation—one that is not what the adopted method was designed for—is encountered? For example, a linear regression model is designed to be used when the scatter plot of the data follows a linear trend. Occasionally, such a linear trend is ruined by a few outliers. Clearly, the outliers are unexpected, or unwelcomed. In order to protect the method from falling apart when facing the unexpected, a method should be chosen, before it is adopted, that is relatively robust to the unexpected.

3.1. Robustness to Outliers

In the statistics literature, quantiles, such as the median, have been used as alternatives to the mean as measures of location. Chambers & Tzavidis (2006) proposed a quantile-based approach to SAE. Their intention was to offer an alternative to the modeling of between-area variation using the random effects. The approach was motivated by the quantile regression (Koenker & Bassett 1978). A key element is called the M-quantile, defined as $Q = Q_q(x; \psi)$, that satisfies the integral identity $\int \psi_q(y - Q)f(y|x)\mathrm{d}y = 0$, where q is a given number in (0, 1), $\psi_q(r) = 2\psi^{-1}(r/s)\{q1_{(r>0)} + (1-q)1_{(r\le 0)}\}$, $\psi(\cdot)$ is an influence function, and s is a robust estimator of scale. Here, y and x denote the response and covariates, respectively. The use of M-quantile instead of standard quantile regression is mainly due to some practical considerations, as the authors argued. Namely, the M-quantile regression is easier to fit by utilizing an iteratively reweighted least squares algorithm. In order to apply the M-quantiles to SAE, the authors introduce a unit-level M-quantile coefficient, q_j, defined through the equation $Q_{q_j}(x_j; \psi) = y_j$, where x_j and y_j are x and y for the jth unit. Furthermore, the area-specific M-quantile coefficient is defined as the average of the unit-level ones. The area M-quantile coefficients are estimated by fitting the model with sample M-quantiles. The authors argue that a main advantage of the M-quantile model is that it allows for outlier-robust inference using widely available M-estimation software. They also discuss a link between their area-specific M-quantile coefficients and the random effects. For example, if all area M-quantile coefficients are equal to 0.5, one may conclude that there is no between-area variation beyond that explained by the model covariates.

Regarding the measure of uncertainty, Chambers & Tzavidis (2006) suggested using the mean squared error (MSE). This seems a bit unnatural as the point seemed to be to avoid using the mean-based approach, which leads to the BP under the MSE. Perhaps something like the inter-quantile range should be considered instead. Another difference from mean-based approaches such as the EBLUP is that, unlike the latter, there was no optimality consideration under the M-quantile framework. Apparently, there is a lack of theoretical foundation regarding the M-quantile.

Chambers (1986, p. 1063) defines a representative outlier as a "sample element with a value that has been correctly recorded and cannot be regarded as unique," and for which "there is

no reason to assume that there are no more similar outliers in the non-sampled part of the population." Sinha & Rao (2009) studied the impact of the representative outliers on the normality-based EBLUP. Although the EBLUPs are efficient under the assumed Gaussian mixed model, they are sensitive to outliers that deviate from the assumed model. Such outliers exist practically, because the Gaussian distribution assumption may never hold exactly. In order to make the EBLUP robust to such outliers, Sinha & Rao (2009) consider the likelihood equation that is derived under the normality assumption. They then modify certain terms involved in the likelihood equation by down-weighting contributions due to the outliers. In particular, the authors used Huber's ψ-function to modify certain "residual" terms, that is, standardized versions of $y_i - X_i\beta$. The robustified ML estimator may be viewed as an M-estimator (e.g., Huber & Ronchetti 2009). Sinha & Rao (2009) established asymptotic normality of their robustified ML estimator. For the measure of uncertainty, the authors adopted a parametric bootstrap approach and studied its empirical performance.

3.2. Robustness to Model Failure

One advantage of the design-based direct estimates is that they are free of model assumptions and therefore not affected by model failure. The design-based approach may be inefficient, of course, when sample size for the direct estimate is small, which is the main concern of SAE. It would be nice to combine the advantages of model-based and design-based approaches. Jiang & Lahiri (2006) attempted to do this by proposing a model-assisted EBP approach using an assumed mixed-effects model, which may be linear or nonlinear, to derive the EBP. Then they justified that the EBP is design-consistent in the sense that, when the sample size is large, the EBP is close to the design-based estimator, and this is true whether the assumed model holds or not (the technical definition is that, when the sample size n_i for the ith small area goes to ∞, the difference between the EBP and the design-based estimator of the finite population domain mean goes to 0 in probability). Since the design-based estimator is known to be accurate when the sample size is large, the EBP is protected from model failure at least for areas with large sample sizes. In practice, due to inhomogeneous subpopulation (small area) sizes, some areas do end up with relatively large sample sizes. For example, in a US national survey, the sample sizes for California, New York, or Texas are often quite large. Model-assisted methods had been previously used in SAE; readers are directed to articles by Särndal (1984), Kott (1989), Prasad & Rao (1999), and You & Rao (2003), among others. An advantage of the Jiang-Lahiri EBP is that an explicit model assumption is not needed for the unobserved units of the finite population. This is in sharp contrast to the pseudo-EBLUP method of Prasad & Rao (1999), who derived a design-consistent estimator assuming that a (superpopulation) LMM holds for all of the units in the finite population, observed or unobserved. In practice, model assumptions are difficult to check for unobserved units.

However, the model-assisted methods mentioned above are not protected from model failure for areas with small sample sizes. For example, a linear model may be so oversimplified that it misrepresents the true small area mean. Of course, one may avoid such model misspecification by carefully choosing the assumed model via a statistical model selection process (see Section 5.2). However, there are practical, sometimes even political, reasons that a simple model like LMM is preferred. For example, such a model is simple to use and interpret, and it easily utilizes auxiliary information. Note that the auxiliary data are often collected using taxpayers' money, so it might be politically incorrect not to use them, even if that is a result of the model selection.

Equation 1 shows that area direct estimate with relatively high sampling variance (or relatively small sample size when the area-level estimate is aggregated from samples within the area) is moved more toward the regression estimator and so is not protected from model misspecification.

Observed best prediction (OBP): uses an assumed model and a broader model to consider how the parameters of the assumed model should be estimated in order to reduce the impact of model misspecification in prediction of mixed effects

Best predictive estimator (BPE): gives more weight to areas with larger sampling variance D_i, in contrast to MLE, which gives more weight to areas with smaller sampling variance

Observed best prediction (OBP), proposed by Jiang et al. (2011), intends to minimize the impact of such a misspecification. Essentially, OBP entertains two models: One is the assumed model and the other is a broader model, usually under very mild or no assumptions. The broader model is always, or almost always, correct, yet it is useless in terms of utilizing the auxiliary information. The assumed model is used to derive the best prediction (BP) of the small area mean, which is no longer the BP when the assumed model fails. The broader model, in contrast, is only used to derive a criterion for estimating the parameters under the assumed model, and the criterion is not model dependent. Note that some of the well-known criteria for parameter estimation, such as ML or REML, are model dependent. The OBP criterion for parameter estimation is derived as an observed MSPE under the true model, which is unknown (but luckily this does not matter, as far as parameter estimation is concerned). The parameter estimator obtained by minimizing the observed MSPE is called the best predictive estimator (BPE), which is different from the ML or REML estimators, in general. Because OBP estimates the parameters under an objective criterion that is unaffected by model misspecification, it is not surprising that OBP is more robust against model failure than EBLUP in terms of predictive performance. The latter was demonstrated both theoretically and empirically by Jiang et al. (2011) and in subsequent work (Chen et al. 2015, Jiang et al. 2015a, Bandyopadhyay 2017).

Example 1 (OBP for hospital data). Recall the hospital data discussed in Section 1.3. Ganesh (2009) proposed a Fay–Herriot model as $y_i = \beta_0 + \beta_1 x_i + v_i + e_i$ to fit the data. However, an inspection of the scatter plot suggests some nonlinear trend. In fact, it appears that a quadratic model would fit the data well except for a potential outlier at the upper right corner.

The question is what to do in this situation. One option would be to look for a more complex model that fits the data better. This approach will be explored later in Section 5.2. Another option is to stay with the relatively simple quadratic model but take into account the potential model misspecification. This would also avoid overfitting, especially given the small sample size. Following the latter approach, the quadratic model, expressed as

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + v_i + e_i, \quad i = 1, \ldots, 23, \tag{2.}$$

is fitted using the OBP method, which is known to be more robust to model misspecification than the EBLUP method. The OBP is based on the BPE of the model parameters, rather than being based on the ML, REML, or Prasad-Rao estimators. To illustrate the difference, let us assume, for now, that A is known. Under the general expression of the Fay–Herriot model (see Section 1.2), the BPE of β has the expression

$$\hat{\beta} = \left\{ \sum_{i=1}^{m} \left(\frac{D_i}{A + D_i} \right)^2 x_i x_i' \right\}^{-1} \sum_{i=1}^{m} \left(\frac{D_i}{A + D_i} \right)^2 x_i y_i. \tag{3.}$$

In comparison, the MLE of β has the expression

$$\tilde{\beta} = \left(\sum_{i=1}^{m} \frac{x_i x_i'}{A + D_i} \right)^{-1} \sum_{i=1}^{m} \frac{x_i y_i}{A + D_i}. \tag{4.}$$

Comparing Equations 3 and 4, we can see that both estimators are weighted averages of the data; the only difference is how the weights are assigned.

This seems to be intuitive when recalling the expression of the BP, which is Equation 1 with $\hat{\beta}$ and $\hat{A}$ replaced by the true β and A, respectively. In other words, the BP is a weighted average of the direct estimator, y_i, and model-based estimator, $x_i'\beta$ (assuming that β is known), and the model part is more relevant to area with larger D_i. Jiang (2017, p. 45) wrote the following on interpreting the difference between the BPE and MLE: "Imagine that there is a meeting of representatives from the different small areas to discuss what estimate of β is to be used in the BP. The areas with larger D_i think that their 'voice' should be heard more (i.e., they should receive more weights), because the BP is more relevant to their business. Their request is reasonable (although, politically, this may not work. . .)." It should be noted that the BP, not the OBP, determines how to assign weights to the direct and model-based estimators; the OBP only finds a way to estimate the parameters involved in the BP, namely β and A, to minimize the potential damage in case the model is misspecified. The EBLUP, in contrast, finds a different way to estimate the parameters, typically via ML, REML, or MoM estimators, which differ from the BPE. For the hospital data, in particular, the BPE under the current model (see Equation 2) are given by $\hat{\beta}_0 = 0.280$, $\hat{\beta}_1 = -0.989$, $\hat{\beta}_2 = 3.261$, and $\hat{A} = 2.099 \times 10^{-4}$. A trio of plots of the direct estimates and resulting OBPs and EBLUPs are presented in **Figure 2**. The effect of differential weighting of hospitals is clearly evident in **Figure 2*c***, which compares the OBPs to the EBLUPs (the line of identity is overlaid). In this case, using the BPE estimates forces the OBPs to be systematically larger in magnitude than their EBLUP counterparts.

3.3. Benchmarking

Pfeffermann (2013, p. 50) wrote in his review article on SAE, "Benchmarking robustifies the inference by forcing the model-based predictors to agree with a design-based estimator for an aggregate of the areas for which the design-based estimator is reliable." A benchmarking equation typically looks like the following:

$$\sum_{i=1}^{m} w_i \hat{\theta}_i = \sum_{i=1}^{m} w_i y_i, \qquad 5.$$

where $\hat{\theta}_i$ is a model-based predictor of the small area mean for the ith small area, y_i is a design-based estimator for the same small area mean, and the w_is are known weights. The right side of Equation 5 is typically a reliable estimator for the aggregated population mean. For example, if the small areas are states, the aggregated population mean may correspond to the national mean. Also, because the design-based estimators have nothing to do with the model, the right side of Equation 5 is unaffected by model failures. If Equation 5 is forced to satisfy when developing the model-based predictors, $\hat{\theta}_i$, there is, at least, protection from model failure at the aggregated level of interest. Sometimes there are multiple benchmarking requirements, so one may have more than one equation like Equation 5 that need to be satisfied. The traditional EBLUPs do not, in general, satisfy the benchmarking requirement(s). For the most part, there have been two approaches to benchmark the EBLUP. The first is to make a suitable adjustment to the traditional EBLUP by adding or multiplying a term to the EBLUP, so that the adjusted EBLUPs satisfy the benchmarking equation. For more information, readers are referred to Pfeffermann & Barnard (1991), Wang et al. (2008), and Rao & Molina (2015, section 6.4.6). The second approach is to modify the way that EBLUP is derived so that the resulting small area predictors are self-benchmarking, that is, they automatically satisfy the benchmarking requirement(s). For example, You & Rao (2002)

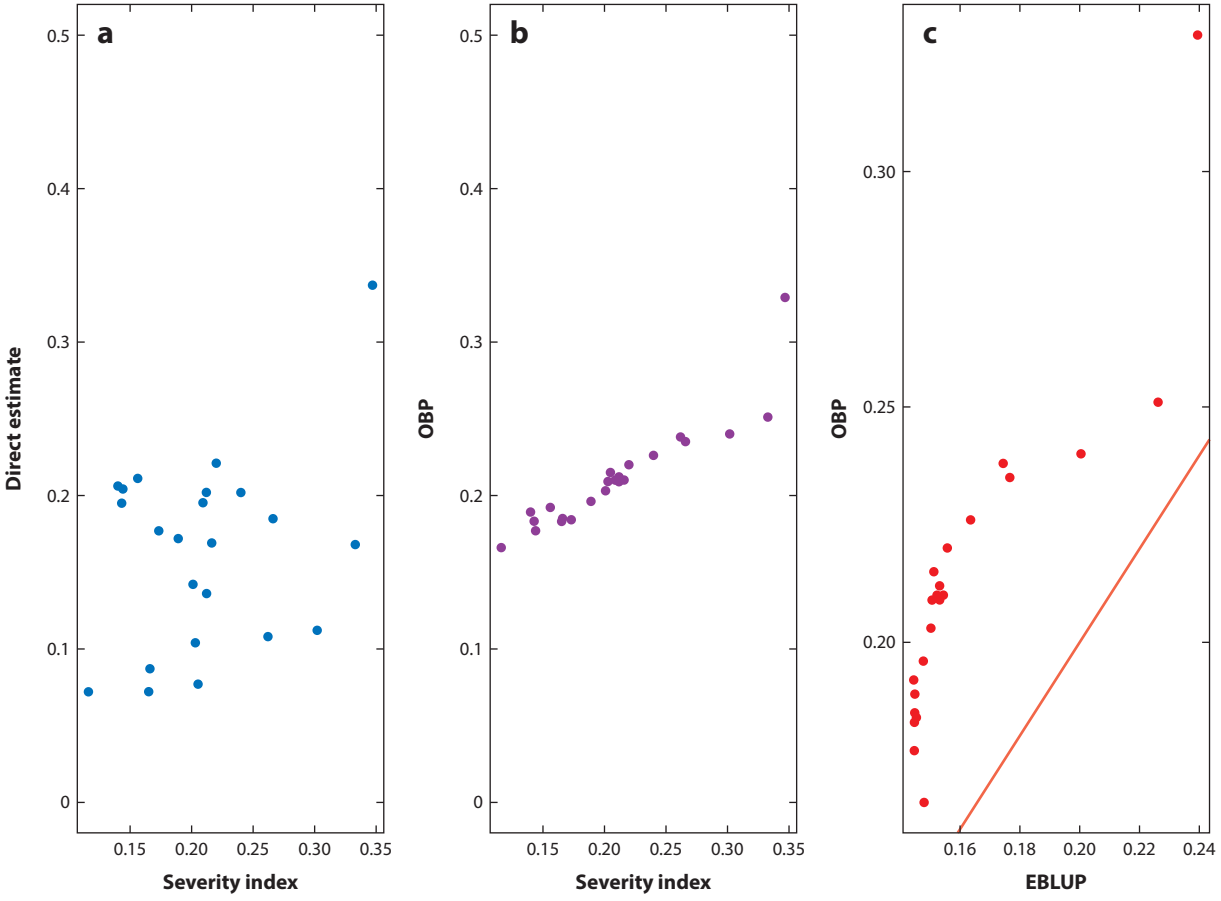

Figure 2

Trio of plots from the hospital data set (Morris & Christiansen 1995) showing (*a*) the direct estimates versus the severity index, (*b*) the observed best prediction (OBP) versus the severity index (notice the smoothing effect produced), and (*c*) the OBP versus the empirical best linear unbiased predictor (EBLUP) estimates. Panel *c* plots the OBP against the EBLUP; the line of identity is shown.

proposed a pseudo-EBLUP using survey weights, which has the self-benchmarking property. In an alternative approach, Wang et al. (2008) introduced an augmented model by adding the weights w_i (see Equation 5), as an additional covariate, to the model. The EBLUPs derived under the augmented model are then self-benchmarking.

Bandyopadhyay (2017) extended both methods, the adjustment method and the augmented model method, to OBP to benchmark the latter under the Fay–Herriot model. Suppose that the original Fay–Herriot model can be expressed as $y_i = x_i'\beta_1 + v_i + e_i,\ i = 1, \ldots, m$, where everything is as described in the second paragraph of Section 1.2, with β replaced by β_1. The augmented Fay–Herriot model is $y_i = x_i'\beta_1 + \beta_2 w_i(1-\gamma_i)^{-1} + v_i + e_i$, which is obtained by adding the term $\beta_2 w_i(1-\gamma_i)^{-1}$ to the original Fay–Herriot model, where $\gamma_i = A/(A + D_i)$. If A is known, define $X_1 = (x_i')_{1\le i\le m}$, $X_2 = X_2(A) = [w_i(1-\gamma_i)^{-1}]_{1\le i\le m}$, and $X = X(A) = [X_1\ X_2(A)]$. Then, the augmented Fay–Herriot model can be written as $y = X(A)\beta + v + e$, where $y = (y_i)_{1\le i\le m}$, $\beta = (\beta_1', \beta_2)'$, $v = (v_i)_{1\le i\le m}$, and $e = (e_i)_{1\le i\le m}$. The BPE of β is given by $\tilde{\beta}(A) = \{X(A)'\Gamma^2(A)X(A)\}^{-1}X(A)'\Gamma^2(A)y$, where $\Gamma(A) = \mathrm{diag}(1-\gamma_i, 1 \le i \le m)$.

When A is unknown, it is replaced by its BPE, $\hat{A}$, which is the minimizer of $\{y - X(A)\tilde{\beta}(A)\}'\Gamma^2(A)\{y - X(A)\tilde{\beta}(A)\} + 2A\mathrm{tr}\{\Gamma(A)\}$, leading to $\hat{\beta} = \tilde{\beta}(\hat{A})$. The OBP of the vector of small area means, $\theta = (\theta_i)_{1\le i\le m}$, is given by $\hat{\theta} = \hat{A}\hat{V}^{-1}y + D\hat{V}^{-1}X(\hat{A})\hat{\beta}$, where $\hat{V} = \hat{A}I_m + D$ and $D = \mathrm{diag}(D_i, 1 \le i \le m)$. A few lines of algebra can then show that the OBP satisfies the benchmark equation, Equation 5, which can be written as $w'\hat{\theta} = w'y$ with $w = (w_i)_{1\le i\le m}$. To explain the procedure intuitively, note that, by Equation 1, one has $y_i - \hat{\theta}_i = (1 - \hat{\gamma}_i)(y_i - x_i'\hat{\beta}) \approx (1 - \gamma_i)(y_i - x_i'\hat{\beta})$, so Equation 5 implies that $\sum_{i=1}^m w_i(1 - \gamma_i)^{-1}(y_i - x_i'\hat{\beta}) \approx 0$. The latter suggests that, after fitting the linear regression of y_i on x_i, the residual, $y_i - x_i'\hat{\beta}$, needs to be uncorrelated with $w_i(1 - \gamma_i)^{-1}$; to make sure that this happens, $w_i(1 - \gamma_i)^{-1}$ needs to be included as a covariate when fitting the regression model, which leads to the augmented Fay–Herriot model.

4. NONPARAMETRIC/SEMIPARAMETRIC SMALL AREA ESTIMATION

One way to achieve robustness is to make the underlying model less restrictive. For example, instead of assuming a linear model, one may include higher-order terms, such as quadratic or cubic functions of the covariates. The higher-order model includes the linear model as a special case (when the coefficients of the higher-order terms are zero) and thus is less restrictive than the linear model in that, even if the linear model fails, the higher-order model may still be valid. More generally, one may model the mean function nonparametrically or semiparametrically.

A nonparametric area-level model, extending the Fay–Herriot model, may be written as

$$y_i = f(x_i) + v_i + e_i, \quad i = 1, \ldots, m, \tag{6.}$$

where the assumptions about v_i and e_i are the same as in the Fay–Herriot model (see Section 1.2) but $f(\cdot)$ is an unknown function. In order to make inference about $f(\cdot)$ trackable, Opsomer et al. (2008) used a P-spline approximation to $f(\cdot)$. Here, P-spline refers to penalized spline, which uses penalization methods to ensure smoothness of the fitting function and avoid overfitting. Specifically, the P-spline is in the form of

$$\tilde{f}(x) = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \gamma_1(x - \kappa_1)_+^p + \cdots + \gamma_q(x - \kappa_q)_+^p, \tag{7.}$$

where p is the degree of the spline, q is the number of knots, κ_j, $1 \le j \le q$ are the knots, and $x_+ = x1_{(x>0)}$. The idea can be extended to other types of small area models as well. In fitting the P-spline-based model, Opsomer et al. (2008) assumed that the γ coefficients are random; this leads to a connection between P-spline fitting and LMM (e.g., Wand 2003) that is used to determine the penalty parameter. Jiang (2010, section 13.4) noted that the LMM connection is asymptotically unjustified in estimating the unknown function $f(\cdot)$. Jiang et al. (2010) considered model selection in choosing the degree of the spline, p, and number of knots, q, using fence methods, a class of strategies for model selection that is particularly suitable for nonconventional problems (e.g., Jiang 2014; also see Section 5.2). Rao et al. (2014) considered a similar P-spline approach to SAE under semiparametric mixed models, extending the work of Sinha & Rao (2009). Lombardía & Sperlich (2008) considered an extension of GLMM in which the conditional mean of the response given the random effects is assumed to satisfy

$$g(\mathrm{E}(y_{ij}|\alpha_i, T_{ij}, X_{ij})) = \lambda(T_{ij}) + x_{ij}'\beta + z_{ij}'\alpha_i,$$

where $g(\cdot)$ is a known link function; α_i is a vector-valued random effect; x_{ij}, t_{ij} are observed vectors of regressors; and z_{ij} is a subvector of $(1, x_{ij}')'$. Furthermore, $\lambda(\cdot)$ is an unknown function. The

authors combined a likelihood approach for mixed-effects models with kernel methods. As for measure of uncertainty, the authors proposed a bootstrap procedure and provided a theoretical justification. They also discussed application to SAE.

Lohr and Mendez (S. Lohr & G. Mendez 2011, unpublished technical report) noted that for spline-based approaches, one needed to have the x_{ij} for population units, not just $\bar{X}_i$, the population mean of the xs. They also raised the question about spline smoothing with several continuous predictors that would then have issues with the curse of dimensionality (Hastie & Tibshirani 1990). In addition, many of the auxiliary data are categorical instead of continuous; as a result, the mean function may not be smooth. They took the approach of subsetting and interactions instead, that is, there may be different relationships between the response and the auxiliary variables in different subgroups. Building upon the work of Mendez (2008), who developed tree-based approaches to model dependent data, Lohr and Mendez extended the proposed tree-based models to SAE. This recognized earlier work on using classification and regression trees for survey data (see Goksel et al. 1992, Schouten & De Nooij 2005, Toth & Eltinge 2011). Lohr and Mendez developed tree growing and pruning approaches. They then went a step further and proposed random forests (RFs) (Brieman 2001) for SAE, which they called mixed RFs to allow random effects. RFs are appealing and are now widely used because they are excellent predictors, nonparametric, resistant to overfitting, and able to model complex interactions. As part of their mixed-RF algorithm, Lohr and Mendez used their proximity-based estimators of residual variance (Mendez & Lohr 2011) for fitting. It should be noted that Bilton (2016) independently studied tree-based methods for SAE when looking at modeling strategies for poverty estimation used to estimate levels of deprivation across small geographical domains, using data from the World Food Programme.

Datta et al. (2018), in one of the last contributions of Peter Hall to SAE before he sadly passed away in 2016, noted that measurements of auxiliary variables used in SAE are often subject to measurement errors. Ignoring such error-in-variable can lead to estimators that perform even worse than the direct survey estimators. The authors proposed a semiparametric approach based on the Fay–Herriot model to produce reliable prediction intervals for small-area characteristics of interest. The approach is semiparametric because it is assumed that the distribution of the auxiliary variable without error, X, which is unobserved, is completely unknown; other parts of random variables, such as the area-specific random effects, the sampling errors, and the measurement errors corresponding to X, follow either parametric or known distributions. The unknown probability density function of X is estimated using the kernel deconvolution estimator of Carroll & Hall (1988), based on which the prediction intervals are produced.

Resampling methods in SAE have received much attention since the beginning of the century (see, e.g., Gershunskaya et al. 2009 for a review). In particular, Hall & Maiti (2006) proposed a nonparametric bootstrap method for estimating the MSPE of EBLUP under an NER model. We note that it is not obvious, in general, how to bootstrap nonparametrically under a mixed-effects model. Efron's (1979) bootstrap, which is based on the i.i.d. assumption, clearly cannot be directly applied. Hall & Maiti (2006) had a clever idea about how to bootstrap without making specific distributional assumptions about the random effects and errors when the interest is in estimating the MSPE. They considered an extended version of the NER model (Battese et al. 1988). A key observation is that the MSPE is a second/fourth-moment quantity in the sense that, up to the order of $o(m^{-1})$, the MSPE of EBLUP only depends on the second and fourth moments of the random effects and errors. Thus, if one can generate random effects and errors such that their second and fourth moments match those of the true random effects and errors, the MSPE of the EBLUP under the generated data distribution would be the same, up to a difference of $o(m^{-1})$,

as that under the actual data distribution. The second and fourth moments of the random effects and errors are estimated from the data, which the authors called moment matching. It should be noted that this method is designed for the area means. It does not work for complex small area parameters, such as small area proportions under a GLMM (e.g., Jiang 2007). In the latter context, Diallo & Rao (2018) developed a skew-normal distribution approach for estimating complex parameters.

5. OTHER TOPICS

There are various contributions that are related to robust SAE, in one way or another, but the topics may not belong to one of the earlier sections, or have special features of their own. We review a few such topics in this section.

5.1. Bayesian Approaches

There are HB and EB approaches to producing robust estimators using the Dirichlet process prior. The latter is a standard tool in nonparametric or semiparametric Bayesian analysis (see, e.g., Ghosh et al. 1989, Polettini 2017). In the EB case, parameters of the Dirichlet process prior are estimated from the data.

Datta & Lahiri (1995) proposed a robust HB approach for SAE in the presence of covariates and outliers. They suggested a way to achieve robustness to second-level model failure in a multivariate Fay–Herriot model by replacing normality with a scale mixture of normal distributions with unknown parameters, which are handled through assignment of priors. They derived some interesting theoretical properties. For example, it was shown that if the model fails for a given area due to an outlier, one still retains the benefit of shrinking for all areas except for the outlying area, which does not happen if one had the normal prior (in which case the HB estimators for all areas converge to the direct estimators). For the outlying area, the HB estimator converges to the direct estimator.

Datta & Ghosh (1991) proposed an HB small area predictor. Chakraborty et al. (2018) showed that this method is not robust to outliers in a way similar to EBLUP (see Section 3.1). Following Sinha & Rao's (2009) method of robustifying the EBLUP, Chakraborty et al. (2018) proposed a robust Bayesian method based on a normal mixture distribution. The set-up may be viewed as a Bayesian version of the NER model (Battese et al. 1988; see Section 1.2), the only difference being that the distribution of the unit-level errors is assumed to follow a two-component normal mixture instead of following the normal distribution. Both components have zero means, but the variances are different. Here, the larger variance corresponds to source of the outliers. The prior for the parameters is noninformative, but sufficient conditions are given to ensure that the resulting posterior is proper. Furthermore, the prior is chosen carefully so that the conditional distributions are simple. Note that the conditional distributions are approximated using the Markov Chain Monte-Carlo (MCMC) method; thus, simplicity of the conditional distributions is important from a computational point of view. Chakraborty et al. (2018) carried out an empirical study of the frequentist properties of their proposed Bayesian predictors, and they showed that the latter performed similarly to the robust EBLUP of Sinha & Rao (2009) in the presence of outliers; both methods outperformed the HB method of Datta & Ghosh (1991) and the M-quantile method of Chambers & Tzavidis (2006) (see Section 3.1). In the absence of outliers, the Chakraborty et al. method performed similarly to that of Datta & Ghosh (1991). An application to the Iowa crops data (Battese et al. 1988) was discussed.

5.2. Model Selection and Diagnostics

Model selection: helps determine the robustness of a model-based method; for example, in some cases a higher-order model provides a better fit to the data than a linear one

Fence methods: construct a statistical barrier to eliminate incorrect models; the optimal model is within the fence and is selected by a criterion that can incorporate practical interest

The importance of model selection in SAE was noted earlier by Battese et al. (1988) and Ghosh & Rao (1994), among others. Datta & Lahiri (2001) discussed the use of the Bayes factor in choosing between a fixed effects model and a random effects model. Meza & Lahiri (2005) demonstrated limitation of Mallow's C_p in selecting the fixed covariates in a NER model. Vaida & Blanchard (2005) proposed a conditional AIC (Akaike information criterion) method that is applicable to selection among NER models. Under a general framework, Jiang et al. (2008) proposed a class of strategies for model selection, known as the fence methods, that is especially suitable to non-conventional model selection problems such as mixed model selection (see Jiang 2014 for more details).

Jiang et al. (2009) applied the fence methods to selection of NER models. They, in particular, applied the method to the Iowa crop data of Battese et al. (1988) and came up with the optimal models of corn for corn, and soybeans for soybeans. Jiang et al. (2010) applied the fence methods to nonparametric model selection using spline approximation (see Section 4), which led to the cubic model mentioned above for the special case of hospital data (Morris & Christiansen 1995, Ganesh 2009). Datta et al. (2011) considered model selection for the Fay–Herriot model by testing the presence of the area-specific random effects. This is equivalent to testing $H_0 : A = 0$, where A is the variance of the area-specific random effect. If H_0 is rejected, the EBLUP of the small area mean is used with the corresponding MSPE as measure of uncertainty; if H_0 is accepted, the regression predictor is used with the standard regression variance estimator as measure of uncertainty. Molina et al. (2015) showed that the method of Datta et al. (2011) (hereafter referred to as the DHM method) can be improved; in particular, EBLUP is retained because it automatically converges to the regression predictor when A is small (see the discussion below Equation 1). Preliminary testing is used to construct the MSPE estimator of EBLUP (see also section 6.4.3 of Rao & Molina 2015). Jiang et al. (2018) noted that the DHM method's measure of uncertainty does not take into account the additional uncertainty in model selection (here via hypothesis testing) and proposed a Monte Carlo jackknife method to assess uncertainty in post-model-selection SAE. Jiang & Torabi (2019) proposed a Sumca (which is an abbreviation for "simple, unified, Monte Carlo assisted") method that can also capture the additional uncertainty due to model selection.

Another type of model selection under the Bayesian framework regards the choice of the prior. Datta et al. (2005) suggested choosing a prior such that measures of uncertainty under Bayesian and frequentist frameworks approximately agree with each other. Here the Bayesian measure of uncertainty is the expected posterior variance, while a frequentist measure of uncertainty is the MSPE. Ganesh & Lahiri (2008) extended the work of Datta et al. (2005) to the weighted average of the expected posterior variances and MSPEs over different small areas.

Example (model selection for the hospital data). The hospital data were first analyzed by Ganesh (2009) in an SAE context. The author proposed a Fay-Herriot model for the graft failure rates. The model is the same as that described in the second paragraph of Section 1.2 with $x_i'\beta = \beta_0 + \beta_1 x_i$, where x_i is the severity index, that is, the average fraction of females, blacks, children, and extremely ill kidney recipients at hospital i. An inspection of the raw data suggests one potential outlier (at the upper right corner of **Figure 3*b***), which corresponds to hospital #5. Note that the case is outlying when a linear model is fitted; when a more complex model is fitted, it may no longer be outlying.

In fact, Jiang et al. (2010) used the fence method to identify the optimal model for the hospital data. They found that the optimal model is a cubic model, which corresponds to the smooth curve in the **Figure 3*b***. Specifically, we consider a class of spline-based non-parametric area-level models. The fence method is used to select the degree of polynomial,

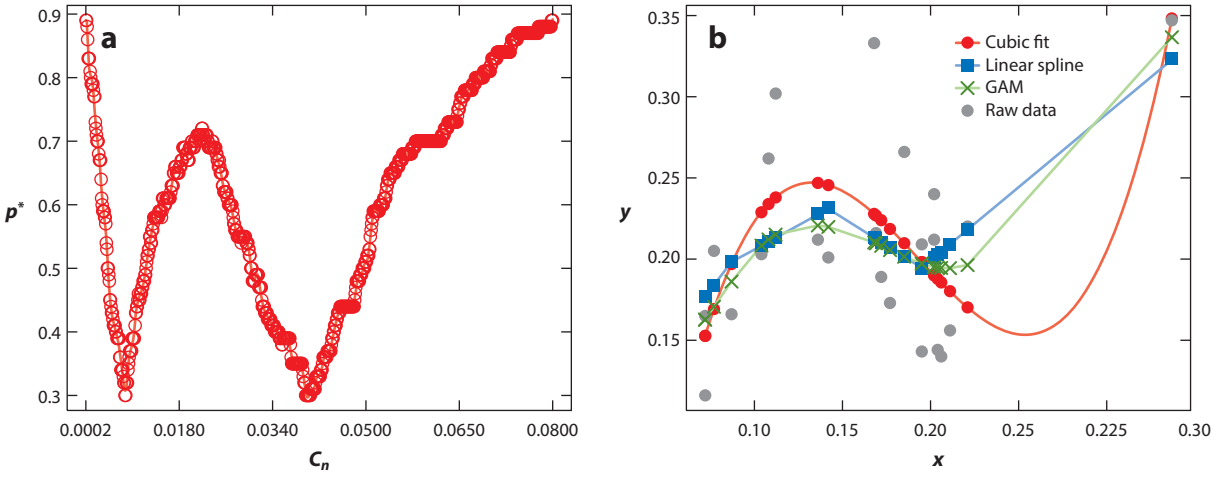

Figure 3

(*a*) Plot of p^* against c_n from the search over the full model space. (*b*) Raw data and the fitted values (*red dots*) and curves—the cubic function resulted from the full model search. Blue squares and lines show the linear spline with four knots from the restricted model search; green crosses and lines show the generalized additive model (GAM) fit.

p, and the number of knots, q, for the spline. The procedure has selected the model with $p = 3$ and $q = 0$, that is, a cubic function with no knots, as the optimal model. It should be noted that the fence method involved bootstrap evaluation of the empirical selection probability. To take into account of the chance error involved in the bootstrapping, the fence procedure was repeated 100 times, each giving the same optimal model. **Figure 3*a*** shows a plot of the empirical selection probability, p^*, against c ($= c_n$), which is a tuning parameter whose choice is a key of the fence method.

Jiang et al. (2010) compared the fence method to a few other methods. The first comparison is with the fence method restricted to the space of the linear splines (i.e., $p = 1$). In this case, the fence method selected a linear spline with four knots (i.e., $q = 4$). The second comparison is with a generalized cross-validation (GCV)–based smoothing method. Here the BRUTO procedure of Hastie & Tibshirani (1990) was used, which augments the class of models to look at a null fit and a linear fit for the spline function. The resulting model selection (i.e., null, linear, or smooth fits) was embedded into a weighted back-fitting algorithm, using GCV for computational efficiency. In this case, the BRUTO finds an overall linear fit for the fixed effects mean function. The models selected by different methods are plotted in **Figure 3*b***.

Closely related to model selection is model checking or diagnostics. Such techniques are often used in conjunction with model selection. For example, an initial check of the proposed model may suggest that the model is a poor fit to the data, a model selection procedure is then carried out to choose a suitable model, model diagnostics are then applied again to make sure that the new model is appropriate, and so on. Broadly speaking, many diagnostic problems in SAE have to do with mixed model diagnostics, a topic that has been discussed in the literature, though not extensively. Several authors have used EBLUP or EB estimators for diagnostic plots, especially regarding the random effects. Examples include articles by Dempster & Ryan (1985), Lange & Ryan (1989), and Calvin & Sedransk (1991). There have also been goodness-of-fit tests for mixed model diagnostics. Jiang (2001) proposed a χ^2-type goodness-of-fit test for LMM diagnostics. Pan & Lin (2005)

proposed a goodness-of-fit test for GLMM based on cumulative sums of residuals. Claeskens & Hart (2009) proposed an alternative approach to that of Jiang (2001) for checking the normality assumption in LMM. It is a likelihood-ratio test that compares the estimated distribution with the null distribution (i.e., normal); model selection via the information criteria is used to determine the larger class of distributions, to which the normal distribution is embedded. Gu (2008) extended the method of Jiang (2001) to mixed logistic models, which are a special case of GLMM (e.g., Jiang 2007, chapter 3). Tang (2010) proposed a different χ^2-type goodness-of-fit test that, unlike that of Jiang (2001), is not based on cell frequencies. More recently, Dao & Jiang (2016) proposed a modified Pearson's χ^2-test for GLMM diagnostics that is guaranteed to have a χ^2 asymptotic null distribution. The χ^2-type tests of Jiang (2001), Gu (2008), and Tang (2010) have weighted χ^2 asymptotic null distributions, whose evaluation requires Monte-Carlo simulations, similar to the Claeskens & Hart (2009) test.

In terms of diagnostics for Bayesian models, Yan & Sedransk (2007, 2010) proposed three methods for checking a missing hierarchical structure in a Bayesian model. The first two methods use *p*-values defined in different ways; one is posterior predictive *p*-value derived from the predictive posterior distribution, and the other is associated with a diagnostic statistic. The third method uses Q-Q plots of the predictive standardized residuals in a way similar to Calvin & Sedransk (1991). Rao & Molina (2015, p. 344) discussed some limitation of the posterior predictive values due to double use of the same data.

5.3. Missing Data

Survey data are often incomplete in the sense that some of the responses, or covariates, are missing. The standard approaches to handling missing data are based on the E-M algorithm and multiple imputation (e.g., Carpenter & Kenward 2013). However, such procedures often require strong, untestable assumptions.

Plass et al. (2017) adapted a cautious likelihood approach (CLA) (see also Plass et al. 2015) to nonresponses in SAE problems. They considered the case of binary responses, of which some are missing. As an example, the authors discussed a study on the area-specific ratio of people at risk of poverty based on data from the German General Social Survey. Here the binary response corresponds to the status of "poor" or "rich" according to a certain definition. Out of a total of 3,466 intended responses, 454 were missing. Instead of assuming the standard missing at random (or not missing at random) missing-data mechanism (MDM) (e.g., Little & Rubin 2014), the CLA makes either no assumption or weak assumptions regarding the MDM.[1] In the case of binary responses with no MDM assumption, the responses are assumed to be either 0, or 1, or "NA" (no answer), hence in three categories. A categorical (multinomial) likelihood approach is then used for inference. In the case of weak MDM assumption, the weak assumption is in terms of setting up constraints on the categorical likelihood. The CLA is, for the most part, applied to the design-based logistic generalized regression–synthetic estimator (Lehtonen & Veijanen 1998). Application to model-based methods is limited, according to Plass et al. (2017), due to some technical and computational difficulties.

Jiang et al. (2015b) developed the expectation-model selection (E-MS) algorithm for model selection in the presence of incomplete data. The basic idea is to extend the concept of "parameter" to the model plus the parameters under the model. This way, the idea of the expectation-maximization algorithm (Dempster et al. 1977) is extended to model selection problems. As

[1] Sensitivity analysis is a subject area that has been developed to address robustness of the results of statistical analysis to various MDMs. While these procedures are fairly standard in medical studies (e.g., Molenberghs & Kenward 2007), applications to SAE problems so far has been limited.

missing or incomplete data are often encountered in surveys, and in view of the importance of model selection in SAE (see Section 5.2), the E-MS algorithm is expected to become a useful tool in SAE.

6. SOFTWARE FOR IMPLEMENTING ROBUST SMALL AREA ESTIMATION METHODS

It is informative to know what software is currently in the public domain for fitting some of the methods we have described. While this information is certainly going to change over time, to date, this is what is currently available in R:

1. The `sae` package (Molina & Marhuenda 2018) produces the Prasad & Rao (1990) MSPE estimator for the Fay–Herriot model. The package will also return resampling-based (parametric and nonparametric bootstrap) MSPE estimates for various forms of the Fay–Herriot model as well as unit-level models.
2. The `robustsae` package (Ghosh et al. 2016) does fully nonsubjective Bayesian analysis for general area-level models.
3. Bandyopadhyay (2017) developed an R package, `OBPSAE`, which returns OBPs for both the area-level and unit-level models. Also implemented in the package is the benchmarked OBP method (see Section 3.3), which the author developed.
4. The BIAS project provides an R package called `ecoreg` along with `WinBUGS` resources for fitting a variety of area-level and unit-level models.
5. The `emdi` R package (Kreutzmann et al. 2018) estimates and maps regional disaggregated indicators. It employs either direct estimation or the EBP approach proposed by Molina & Rao (2010), who developed the approach for estimating nonlinear small area population parameters (but the approach is applicable to general nonlinear parameters). Estimates of the MSPE for the EBLUPs are done using the parametric bootstrap technique developed by González-Manteiga et al. (2008).

7. CONCLUDING REMARKS

One issue related to robustness of SAE methods has to do with measurement errors. In fact, outliers in the data are often due to measurement or recording errors. Most of the existing literature on measurement errors is not very useful from a practical standpoint. This is because most, if not all, of the papers deal with random measurement error models rather than systematic measurement error models, which are the most challenging case. Even for a random measurement error model, it seems that the only case where something may be implemented is when it arises from sampling consideration (e.g., covariate subject to measurement errors). In order to estimate the sampling variances, one would need a large sample, but then the result would likely be almost the same if one simply ignores the measurement errors. In fact, if the measurement error is not due to sampling, it is not even clear how to define an uncertainty due to measurement error (some coverage of this topic is provided by A.L. Erciulescu, C. Franco, and P. Lahiri in their 2018 unpublished book chapter on use of administrative records in small area estimation).

There have been notable trends in the current development of SAE, including both research and applications. These include SAE with big data (e.g., Marchetti et al. 2015) and linked data (e.g., Lahiri 2017; Y. Han & P. Lahiri, unpublished manuscript), as well as interaction of SAE with other fields. In the latter trend, the interaction is in both ways. On the one hand, modern statistical methods that have been mostly used in other fields of statistics, such as nonparametric

methods (see Section 4), are increasingly being used in SAE. On the other hand, SAE methods have received increasing attention in other fields of statistics, such as health science statistics (e.g., Sun et al. 2018). There is also growing interest in exploring more complex data structure, such as spatial correlation (e.g., J. Jiang & M. Torabi, unpublished manuscript).

Historically, it did not take long for modern technologies from other fields of statistics to make their ways to SAE, but doing so must incorporate the special and practical needs of the latter. For example, one reason that robustness to model misspecification is important is because SAE methods are often used by practitioners who are not technically prepared to enjoy using sophisticated models. In fact, this is why linear models are popular among SAE practitioners. Clearly, a simpler model, such as a linear model, is more likely to be misspecified than a complex one, such as a nonparametric model. This was illustrated in our example in Section 5.2. However, a simple model is easier to understand and interpret. For example, an EBLUP is, perhaps, much easier to explain to a government official than something fitted via a machine learning method in spite of the fact that the latter may provide a better fit. This is a reality that one must face in practical SAE.

As is seen throughout the review, robust methods often have an advantage in dealing with complex data structure, working under complex models, and potentially in working with big data. One reason is that robust methods tend to be simpler than methods that rely on strong assumptions. For example, MoM type methods (e.g., Section 2.2) tend to be simpler than likelihood-based methods. Such simplicity often leads to computational advantage, which is important in the case of big data or complex models.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors are grateful to Editorial Committee, Professor Thomas Louis, and Professors P. Lahiri and J.N.K. Rao for their comments and discussions that have helped greatly in improving the manuscript.

LITERATURE CITED

Bandyopadhyay R. 2017. *Benchmarking the observed best predictor*. PhD Diss., Univ. Calif., Davis

Battese GE, Harter RM, Fuller WA. 1988. An error-components model for prediction of county crop areas using survey and satellite data. *J. Am. Stat. Assoc.* 80:28–36

Bell WR, Huang ET. 2006. Using the *t*-distribution to deal with outliers in small area estimation. In *Proceedings of Statistics Canada Symposium on Methodological Issues in Measuring Population Health*. Ottawa: Stat. Can.

Bilton PA. 2016. *Tree-based models for poverty estimation*. PhD Thesis, Massey Univ., Palmerston North, N.Z.

Brieman L. 2001. Random forests. *Mach. Learn.* 45:5–32

Calvin JA, Sedransk J. 1991. Bayesian and frequentist predictive inference for the patterns of care studies. *J. Am. Stat. Assoc.* 86:36–48

Carpenter JR, Kenward MG. 2013. *Multiple Imputation and Its Application*. New York: Wiley

Carroll RJ, Hall P. 1988. Optimal rates of convergence for deconvolving a density. *J. Am. Stat. Assoc.* 83:1184–86

Chakraborty A, Datta GK, Mandal A. 2018. Robust hierarchical Bayes small area estimation for nested error regression model. arXiv:1702.05832v2 [stat.ME]

Chambers R, Tzavidis N. 2006. M-quantile models for small area estimation. *Biometrika* 93:255–68

Chambers RL. 1986. Outlier robust finite population estimation. *J. Am. Stat. Assoc.* 81:1063–69
Chen S, Jiang J, Nguyen T. 2015. Observed best prediction for small area counts. *J. Surv. Stat. Methodol.* 3:136–61
Claeskens G, Hart JD. 2009. Goodness-of-fit tests in mixed models (with discussion). *TEST* 18:213–39
Dao C, Jiang J. 2016. A modified Pearson's χ^2 test with application to generalized linear mixed model diagnostics. *Ann. Math. Sci. Appl.* 1:195–215
Datta GS. 2009. Model-based approach to small area estimation. In *Handbook of Statistics*, Vol. 29B, *Sample Surveys: Inference and Analysis*, ed. D Pfeffermann, CR Rao, pp. 251–88. Amsterdam: North-Holland
Datta GS, Delaigle A, Hall P, Wang Li. 2018. Semi-parametric prediction intervals in small areas when auxiliary data are measured with error. *Stat. Sin.* 28:2309–36
Datta GS, Ghosh M. 1991. Bayesian prediction in linear models: applications to small area estimation. *Ann. Stat.* 19:1748–70
Datta GS, Hall P, Mandal A. 2011. Model selection by testing for the presence of small-area effects, and applications to area-level data. *J. Am. Stat. Assoc.* 106:361–74
Datta GS, Lahiri P. 1995. Robust hierarchical Bayes estimation of small area characteristics in the presence of covariates. *J. Multivar. Anal.* 54:310–28
Datta GS, Lahiri P. 2001. Discussions on a paper by Efron & Gous. In *Model Selection*, ed. P Lahiri, pp. 249–54. Beachwood, OH: Inst. Math. Stat.
Datta GS, Rao JNK, Smith DD. 2005. On measuring the variability of small area estimators under a basic area level model. *Biometrika* 92:183–96
Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* 39:1–38
Dempster AP, Ryan LM. 1985. Weighted normal plots. *J. Am. Stat. Assoc.* 80:845–50
Diallo MS, Rao JNK. 2018. Small area estimation for complex parameters under unit-level model with skew-normal error. *Scand. J. Stat.* 45:1092–1116
Efron B. 1979. Bootstrap method: another look at the jackknife. *Ann. Stat.* 7:1–26
Fay RE, Herriot RA. 1979. Estimates of income for small places: an application of James-Stein procedures to census data. *J. Am. Stat. Assoc.* 74:269–77
Fisher RA. 1922. On the interpretation of chi-square from contingency tables, and the calculation of P. *J. R. Stat. Soc.* 85:87–94
Ganesh N. 2009. Simultaneous credible intervals for small area estimation problems. *J. Multivar. Anal.* 100:1610–21
Ganesh N, Lahiri P. 2008. A new class of average moment matching priors. *Biometrika* 95:514–20
Gershunskaya J, Jiang J, Lahiri P. 2009. Resampling methods in surveys. In *Sample Surveys: Theory, Methods and Inference*, ed. D Pfeffermann, CR Rao, pp. 121–51. Amsterdam: Elsevier
Gershunskaya J, Lahiri P. 2018. Robust empirical best small area finite population mean estimation using a mixture model. *Calcutta Stat. Assoc. Bull.* 69:183–204
Ghosh M, Lahiri P. 1987. Robust empirical Bayes estimation of means from stratified samples. *J. Am. Stat. Assoc.* 82:1153–62
Ghosh M, Lahiri P, Tiwari RC. 1989. Nonparametric empirical Bayes estimation of the distribution and the mean. *Comm. Stat. Theory Methods* 18:121–46
Ghosh M, Myung J, Moura F. 2016. `robustsae`: robust Bayesian small area estimation. *R package*. **https://cran.r-project.org/web/packages/robustsae/index.html**
Ghosh M, Natarajan K, Stroud TWF, Carlin BP. 1998. Generalized linear models for small-area estimation. *J. Am. Stat. Assoc.* 93:273–82
Ghosh M, Rao JNK. 1994. Small area estimation: an appraisal (with discussion). *Stat. Sci.* 9:55–93
Goksel H, Judkins DR, Mosher WD. 1992. Nonresponse adjustments for a telephone follow-up to a national in-person survey. *J. Off. Stat.* 8:417–31
González-Manteiga W, Lombardía MJ, Molina I, Morales D, Santamaría L. 2008. Bootstrap mean squared error of a small-area EBLUP. *J. Stat. Comput. Simul.* 8:443–62
Gu Z. 2008. *Model diagnostics for generalized linear mixed models*. PhD Diss., Univ. Calif., Davis

Hall P, Maiti T. 2006. Nonparametric estimation of mean-squared prediction error in nested-error regression models. *Ann. Stat.* 34:1733–50

Hastie TJ, Tibshirani RJ. 1990. *Generalized Additive Models*. Boca Raton, FL: Chapman & Hall/CRC

Huber PJ, Ronchetti EM. 2009. *Robust Statistics*. Hoboken, NJ: Wiley. 2nd ed.

Jiang J. 2001. Goodness-of-fit tests for mixed model diagnostics. *Ann. Stat.* 29:1137–64

Jiang J. 2003. Empirical best prediction for small area inference based on generalized linear mixed models. *J. Stat. Plann. Inference* 111:117–27

Jiang J. 2007. *Linear and Generalized Linear Mixed Models and Their Applications*. New York: Springer

Jiang J. 2010. *Large Sample Techniques for Statistics*. New York: Springer

Jiang J. 2014. The fence methods. *Adv. Stat.* 2014:830821

Jiang J. 2017. *Asymptotic Analysis of Mixed Effects Models: Theory, Applications, and Open Problems*. Boca Raton, FL: Chapman & Hall/CRC

Jiang J, Lahiri P. 2001. Empirical best prediction for small area inference with binary data. *Ann. Inst. Stat. Math.* 53:217–43

Jiang J, Lahiri P. 2006. Mixed model prediction and small area estimation (with discussion). *TEST* 15:1–96

Jiang J, Lahiri P, Nguyen T. 2018. A unified Monte-Carlo jackknife for small area estimation after model selection. *Ann. Math. Sci. Appl.* 3:405–38

Jiang J, Lahiri P, Wan SM. 2002. A unified jackknife theory for empirical best prediction with M-estimation. *Ann. Stat.* 30:1782–810

Jiang J, Nguyen T. 2009. Comments on: Goodness-of-fit tests in mixed models by G. Claeskens and J.D. Hart. *TEST* 18:248–55

Jiang J, Nguyen T. 2012. Small area estimation via heteroscedastic nested-error regression. *Can. J. Stat.* 40:588–603

Jiang J, Nguyen T, Rao JS. 2009. A simplified adaptive fence procedure. *Stat. Probab. Lett.* 79:625–29

Jiang J, Nguyen T, Rao JS. 2010. Fence method for nonparametric small area estimation. *Surv. Methodol.* 36:3–11

Jiang J, Nguyen T, Rao JS. 2011. Best predictive small area estimation. *J. Am. Stat. Assoc.* 106:732–45

Jiang J, Nguyen T, Rao JS. 2015a. Observed best prediction via nested-error regression with potentially misspecified mean and variance. *Survey Methodol.* 41:37–55

Jiang J, Nguyen T, Rao JS. 2015b. The E-MS algorithm: model selection with incomplete data. *J. Am. Stat. Assoc.* 110:1136–47

Jiang J, Rao JS, Gu Z, Nguyen T. 2008. Fence methods for mixed model selection. *Ann. Stat.* 36:1669–92

Jiang J, Torabi M. 2019. Sumca: simple, unified, Monte-Carlo assisted approach to second-order unbiased MSPE estimation. *J. R. Stat. Soc. B*. In press

Jiang J, Zhang W. 2001. Robust estimation in generalized linear mixed models. *Biometrika* 88:753–765

Koenker P, Bassett G. 1978. Regression quantiles. *Econometrica* 46:33–50

Kott P. 1989. Robust small domain estimation using random effects modelling. *Survey Methodol.* 15:3–12

Kreutzmann AK, Pannier S, Rojas-Perilla N, Schmid T, Templ M, Tzavidis N. 2018. `emdi`: estimating and mapping disaggregated indicators. *R package*. **https://cran.r-project.org/web/packages/emdi/emdi.pdf**

Lahiri P. 2017. *Small area estimation with linked data*. Keynote address, 2017 ISI Satellite Meeting on Small Area Estimation, Paris, France, July 10–12

Lahiri P, Rao JNK. 1995. Robust estimation of mean squared error of small area estimators. *J. Am. Stat. Assoc.* 90:758–66

Lange N, Ryan L. 1989. Assessing normality in random effects models. *Ann. Stat.* 17:624–42

Lehtonen R, Veijanen A. 1998. Logistic generalised regression estimators. *Surv. Methodol.* 24:51–56

Little R, Rubin D. 2014. *Statistical Analysis with Missing Data*. New York: Wiley. 2nd ed.

Lohr SL. 2010. *Sampling: Design and Analysis*. Boston: Brooks/Cole

Lombardía MJ, Sperlich S. 2008. Semiparametric inference in generalized mixed effects models. *J. R. Stat. Soc. B* 70:913–30

Marchetti S, Giusti C, Pratesi M, Salvati N, Giannotti F, et al. 2015. Small area model-based estimators using big data sources. *J. Off. Stat.* 31:263–81

Matloff NS. 1981. Use of regression functions for improved estimation of means. *Biometrika* 68:685–89
Mendez G. 2008. *Tree-based methods to model dependent data*. PhD Thesis, Ariz. State Univ.
Mendez G, Lohr S. 2011. Estimating residual variance in random forest regression. *Comput. Stat. Data Anal.* 55:2937–50
Meza J, Lahiri P. 2005. A note on the C_p statistic under the nested error regression model. *Surv. Methodol.* 31:105–9
Molenberghs G, Kenward MG. 2007. *Missing Data in Clinical Studies*. New York: Wiley
Molina I, Marhuenda Y. 2018. `sae`: small area estimation. *R package*. **https://cran.r-project.org/web/packages/sae/sae.pdf**
Molina I, Rao JNK. 2010. Small area estimation of poverty indicators. *Can. J. Stat.* 38:369–85
Molina I, Rao JNK, Datta GS. 2015. Small area estimation under a Fay–Herriot model with preliminary testing for the presence of area random effects. *Survey Methodol.* 41:1–19
Morris CN, Christiansen CL. 1995. Hierarchical models for ranking and for identifying extremes with applications. In *Bayesian Statistics*, Vol. 5, ed. JO Bernando, JO Berger, AP Dawid, AFM Smith, pp. 278–95. Oxford, UK: Oxford Univ. Press
Newey WK. 1985. Generalized method of moments specification testing. *J. Econom.* 29:229–56
Opsomer JD, Breidt FJ, Claeskens G, Kauermann G, Ranalli MG. 2008. Nonparametric small area estimation using penalized spline regression. *J. R. Stat. Soc. B* 70:265–86
Pan Z, Lin DY. 2005. Goodness-of-fit methods for generalized linear mixed models. *Biometrics* 61:1000–9
Pfeffermann D. 2013. New important developments in small area estimation. *Stat. Sci.* 28:40–68
Pfeffermann D, Barnard CH. 1991. Some new estimators for small-area means with application to the assessment of farmland values. *J. Bus. Econ. Stat.* 9:73–84
Plass J, Augustin T, Cattaneo M, Schollmeyer G. 2015. Statistical modelling under epistemic data imprecision: some results on estimating multinomial distributions and logistic regression for coarse categorical data. In *Proceedings of the 9th International Symposium on Imprecise Probability: Theories and Applications*, ed. T Augustin, S Doria, E Miranda, E Quaeghebeur, pp. 247–56. **https://pdfs.semanticscholar.org/64d7/a6c79502a898ec370774792500c83779139d.pdf**
Plass J, Omar A, Augustin T. 2017. Towards a cautious modelling of missing data in small area estimation. *Proc. Mach. Learn. Res.* 62:253–64
Polettini S. 2017. A generalised semiparametric Bayesian Fay–Herriot model for small area estimation shrinking both means and variances. *Bayesian Anal.* 12:729–52
Prasad NGN, Rao JNK. 1990. The estimation of mean squared errors of small area estimators. *J. Am. Stat. Assoc.* 85:163–71
Prasad NGN, Rao JNK. 1999. On robust small area estimation using a simple random effects model. *Surv. Methodol.* 25:67–72
Quenouille M. 1949. Approximation tests of correlation in time series. *J. R. Stat. Soc. B* 11:18–84
Rao JNK, Molina I. 2015. *Small Area Estimation*. New York: Wiley. 2nd ed.
Rao JNK, Sinha SK, Dumitrescu L. 2014. Robust small area estimation under semi-parametric mixed models. *Can. J. Stat.* 42:126–41
Särndal CE. 1984. Design-consistent versus model-dependent estimation for small domains. *J. Am. Stat. Assoc.* 79:624–31
Schouten B, de Nooij G. 2005. *Nonresponse adjustment using classification trees*. Disc. Pap. 05001, Stat. Neth., Voorburg/Heerlen
Sinha SK, Rao JNK. 2009. Robust small area estimation. *Can. J. Stat.* 37:381–99
Sugasawa S, Kubokawa T. 2017. Heteroscedastic nested error regression models with variance functions. *Stat. Sin.* 27:1101–23
Sun H, Nguyen T, Luan Y, Jiang J. 2018. Classified mixed logistic model prediction. *J. Multivar. Anal.* 168:63–74
Tang M. 2010. *Goodness-of-fit tests for generalized linear mixed models*. PhD Diss., Univ. Md., Coll. Park
Torabi M. 2012. Likelihood inference in generalized linear mixed models with two components of dispersion using data cloning. *Comput. Stat. Data Anal.* 56:4259–65

Toth D, Eltinge J. 2011. Building consistent regression trees from complex sample data. *J. Am. Stat. Assoc.* 106:1626–36

Vaida F, Blanchard S. 2005. Conditional Akaike information for mixed-effects models. *Biometrika* 92:351–70

Wand M. 2003. Smoothing and mixed models. *Comput. Stat.* 18:223–49

Wang J, Fuller WA, Qu Y. 2008. Small area estimation under restriction. *Surv. Methodol.* 34:29–36

Yan G, Sedransk J. 2007. Bayesian diagnostic techniques for detecting hierarchical structure. *Bayesian Anal.* 2:735–60

Yan G, Sedransk J. 2010. A note on Bayesian residuals as a hierarchical model diagnostic technique. *Stat. Pap.* 51:1

You Y, Rao JNK. 2002. A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *Can. J. Stat.* 30:431–39

You Y, Rao JNK. 2003. Pseudo-hierarchical Bayes small area estimation combining unit-level models and survey weights. *J. Stat. Plann. Inference* 111:197–208

Annual Review of Statistics and Its Application

Volume 7, 2020

Contents

Errata

An online log of corrections to *Annual Review of Statistics and Its Application* articles may be found at http://www.annualreviews.org/errata/statistics