*STATISTICS IN TRANSITION new series, Special Issue, August 2020*
*Vol. 21, No. 4, pp. 1–22, DOI 10.21307/stattrans-2020-022*
Received – 31.01.2020; accepted – 30.06.2020

# Small area estimation: its evolution in five decades

Malay Ghosh[1]

## ABSTRACT

The paper is an attempt to trace some of the early developments of small area estimation. The basic papers such as the ones by Fay and Herriott (1979) and Battese, Harter and Fuller (1988) and their follow-ups are discussed in some details. Some of the current topics are also discussed.

**Key words:** template, article, journal.

## 1. Prologue

Small area estimation is witnessing phenomenal growth in recent years. The vastness of the area makes it near impossible to cover each and every emerging topic. The review articles of Ghosh and Rao (1994), Pfeffermann (2002, 2013) and the classic text of Rao (2003) captured the contemporary research of that time very successfully. But the literature continued growing at a very rapid pace. The more recent treatise of Rao and Molina (2015) picked up many of the later developments. But then there came many other challenging issues, particularly with the advent of "big data", which started moving the small area estimation machine faster and faster. It seems real difficult to cope up with this super-fast development.

In this article, I take a very modest view towards the subject. I have tried to trace the early history of the subject up to some of the current research with which I am familiar. It is needless to say that the topics not covered in this article far outnumber those that are covered. Keeping in mind this limitation, I will make a feeble attempt to trace the evolution of small area estimation in the past five decades.

## 2. Introduction

The first and foremost question that one may ask is "what is small area estimation"? Small area estimation is any of several statistical techniques involving estimation of parameters in small 'sub-populations' of interest included in a larger 'survey'. The term 'small area' in this context generally refers to a small geographical area such as a county, census tract or a school district. It can also refer to a 'small domain' cross-classified by

[1]Department of Statistics, University of Florida, Gainesville, FL, USA. E-mail: ghoshm@stat.ufl.edu. ORCID: https://orcid.org/0000-0002-8776-7713.

several demographic characteristics, such as age, sex, ethnicity, etc. I want to emphasize that it is not just the area, but the 'smallness' of the targeted population within an area that constitutes the basis for small area estimation. For example, if a survey is targeted towards a population of interest with prescribed accuracy, the sample size in a particular subpopulation may not be adequate to generate similar accuracy. This is because if a survey is conducted with sample size determined to attain prescribed accuracy in a large area, one may not have the resources available to conduct a second survey to achieve similar accuracy for smaller areas.

A domain (area) specific estimator is 'direct' if it is based only on the domain-specific sample data. A domain is regarded as 'small' if domain-specific sample size is not large enough to produce estimates of desired precision. Domain sample size often increases with population size of the domain, but that need not always be the case. This requires use of 'additional' data, be it either administrative data not used in the original survey, or data from other related areas. The resulting estimates are called 'indirect' estimates that 'borrow strength' for the variable of interest from related areas and/or time periods to increase the 'effective' sample size. This is usually done through the use of models, mostly 'explicit', or at least 'implicit' that links the related areas and/or time periods.

Historically, small area statistics have long been used, albeit without the name "small area" attached to it. For example, such statistics existed in eleventh century England and seventeenth century Canada based on either census or on administrative records. Demographers have long been using a variety of indirect methods for small area estimation of population and other characteristics of interest in postcensal years. I may point out here that the eminent role of administrative records for small area estimation cannot but be underscored even today. A very comprehensive review article in this regard is due to Erciulescu, Franco and Lahiri (2020).

In recent years, the demand for small area statistics has greatly increased worldwide. The need is felt for formulating policies and programs, in the allocation of government funds and in regional planning. For instance, legislative acts by national governments have created a need for small area statistics. A good example is SAIPE (Small Area Income and Poverty Estimation) mandated by the US Legislature. Demand from the private sector has also increased because business decisions, particularly those related to small businesses, rely heavily on local socio-economic conditions. Small area estimation is of particular interest for the transition economics in central and eastern European countries and the former Soviet Union countries. In the 1990's these countries have moved away from centralized decision making. As a result, sample surveys are now used to produce estimates for large areas as well as small areas.

## 3. Examples

Before tracing this early history, let me cite a few examples that illustrate the ever increasing current day importance of small area estimation. One important ongoing small area estimation problem at the U.S. Bureau of the Census is the small area income and poverty estimation (SAIPE) project. This is a result of a Bill passed by the US House of Representatives requiring the Secretary of Commerce to produce and publish at least every two years beginning in 1996, current data related to the incidence of poverty in the United States. Specifically, the legislation states that "to the extent feasible", the secretary shall produce estimates of poverty for states, counties and local jurisdictions of government and school districts. For school districts, estimates are to be made of the number of poor children aged 5-17 years. It also specifies production of state and county estimates of the number of poor persons aged 65 and over.

These small area statistics are used by a broad range of customers including policy makers at the state and local levels as well as the private sector. This includes allocation of Federal and state funds. Earlier the decennial census was the only source of income distribution and poverty data for households, families and persons for such small geographic areas. Use of the recent decennial census data pertaining to the economic situation is unreliable especially as one moves further away from the census year. The first SAIPE estimates were issued in 1995 for states, 1997 for counties and 1999 for school districts. The SAIPE state and county estimates include median household income number of poor people, poor children under age 5 (for states only), poor children aged 5-17, and poor people under age 18. Also starting 1999, estimates of the number of poor school-aged children are provided for the 14,000 school districts in the US (Bell, Basel and Maples, 2016).

Another example is the Federal-State Co-Operative Program (FSCP). It started in 1967. The goal was to provide high-quality consistent series of post-censal county population estimates with comparability from area to area. In addition to the county estimates, several members of FSCP now produce subcounty estimates as well. Also, the US Census Bureau used to provide the Treasury Department with Per Capita Income (PCI) estimates and other statistics for state and local governments receiving funds under the general revenue sharing program. Treasury Department used these statistics to determine allocations to local governments within the different states by dividing the corresponding state allocations. The total allocation by the Treasury Dept. was $675 billion in 2017.

United States Department of Agriculture (USDA) has long been interested in prediction of areas under corn and soybeans. Battese, Harter and Fuller (JASA, 1988) considered the problem of predicting areas under corn and soybeans for 12 counties in North-Central Iowa based on the 1978 June enumerative survey data as well as Landsat Satellite Data. The USDA statistical reporting Service field staff determined the area of corn and soybeans in 37 sample segments of 12 counties in North Central Iowa by interviewing farm operators. In conjunction with LANDSAT readings obtained during August and September 1978, USDA procedures were used to classify the crop cover for all pixels in the 12 counties.

There are many more examples. An important current day example is small area "poverty mapping" initiated by Elbers, Lanjouw and Lanjouw (2003). This was extended as well as substantially refined by Molina and Rao (2010) and many others.

## 4. Synthetic Estimation

An estimator is called 'Synthetic' if a direct estimator for a large area covering a small area is used as an indirect estimator for that area. The terminology was first used by the U.S. National Center for Health Statistics. These estimators are based on a strong underlying assumption is that the small area bears the same characteristic for the large area.

For example, if $y_1, \cdots, y_m$ are the direct estimates of average income for $m$ areas with population sizes $N_1, \cdots, N_m$, we may use the overall estimate $\bar{y}_s = \sum_{j=1}^{m} N_j y_j / N$ for a particular area, say, $i$ ,where $N = \sum_{j=1}^{m} N_j$. The idea is that this synthetic estimator has less mean squared error (MSE) compared to the direct estimator $y_i$ if the bias $\bar{y}_s - y_i$ is not too strong. On the other hand, a heavily biased estimator can affect the MSE as well.

One of the early use of synthetic estimation appears in Hansen, Hurwitz and Madow (1953, pp 483-486). They applied synthetic regression estimation in the context of radio listening. The objective was to estimate the median number of radio stations heard during the day in each of more than 500 counties in the US. The direct estimate $y_i$ of the true (unknown) median $M_i$ was obtained from a radio listening survey based on personal interviews for 85 county areas. The selection was made by first stratifying the population county areas into 85 strata based on geographical region and available radio service type. Then one county was selected from each stratum with probability proportional to the estimated number of families in the counties. A subsample of area segments was selected from each of the sampled county areas and families within the selected area segments were interviewed.

In addition to the direct estimates, an estimate $x_i$ of $M_i$, obtained from a mail survey was used as a single covariate in the linear regression of $y_i$ on $x_i$. The mail survey was first conducted by sampling 1,000 families from each county area and mailing questionnaires. The $x_i$ were biased due to nonresponse (about 20% response rate) and incomplete coverage, but were anticipated to have high correlation with the $M_i$. Indeed, it turned out that $\text{Corr}(y_i, x_i) = .70$. For nonsampled counties, regression synthetic estimates were $\hat{M}_i = .52 + .74 x_i$.

Another example of Synthetic Estimation is due to Gonzalez and Hoza (JASA, 1978, pp 7-15). Their objective was to develop intercensal estimates of various population characteristics for small areas. They discussed synthetic estimates of unemployment where the larger area is a geographic division and the small area is a county.

Specifically, let $p_{ij}$ denote the proportion of labor force in county $i$ that corresponds to cell $j$ $(j = 1, \cdots, G)$. Let $u_j$ denote the corresponding unemployment rate for cell $j$

based on the geographic division where county $i$ belongs. Then, the synthetic estimate of the unemployment rate for county $i$ is given by $u_i^* = \sum_{j=1}^{G} p_{ij} u_j$. These authors also suggested synthetic regression estimate for unemployment rates.

While direct estimators suffer from large variances and coefficients of variation for small areas, synthetic estimators suffer from bias, which often can be very severe. This led to the development of composite estimators, which are weighted averages of direct and synthetic estimators. The motivation is to balance the design bias of synthetic estimators and the large variability of direct estimators in a small area.

Let $y_{ij}$ denote the characteristic of interest for the $j$th unit in the $i$th area; $j = 1, \cdots, N_i; i = 1, \cdots, m$. Let $x_{ij}$ denote some auxiliary characteristic for the $j$th unit in the $i$th local area. Note that the population means are $\bar{Y}_i = \sum_{j=1}^{N_i} y_{ij}/N_i$ and $\bar{X}_i = \sum_{j=1}^{N_i} x_{ij}/N_i$. We denote the sampled observations as $y_{ij},\ j = 1, \cdots, n_i$ with corresponding auxiliary variables $x_{ij}$, $j = 1, \cdots, n_i$. Let $\bar{x}_i = \sum_{j=1}^{n_i} x_{ij}/n_i$. $\bar{x}_i$ is obtained from the sample. In addition, one needs to know $\bar{X}_i$, the population average of auxiliary variables.

A Direct Estimator (Ratio Estimator) of $\bar{Y}_i$ is $\bar{y}_i^R = (\bar{y}_i/\bar{x}_i)\bar{X}_i$. The corresponding Ratio Synthetic Estimator of $\bar{Y}_i$ is $(\bar{y}_s/\bar{x}_s)\bar{X}_i$, where $\bar{y}_s = \sum_{i=1}^{m} N_i \bar{y}_i / \sum_{i=1}^{m} N_i$ and $\bar{x}_s = \sum_{i=1}^{m} N_i \bar{x}_i / \sum_{i=1}^{m} N_i$. A Composite Estimator of $\bar{Y}_i$ is

$$(n_i/N_i)\bar{y}_i + (1 - n_i/N_i)(\bar{y}_s/\bar{x}_s)\bar{X}_i',$$

where $\bar{X}_i' = (N_i - n_i)^{-1} \sum_{j=n_i+1}^{N_i} x_{ij}/(N_i - n_i)$. Note $N_i\bar{X}_i = n_i\bar{x}_i + (N_i - n_i)\bar{X}_i'$. All one needs to know is the population average $\bar{X}_i$ in addition to the already known sample average $\bar{x}_i$ to find $\bar{X}_i'$. Several other weights in forming a linear combination of direct and synthetic estimators have also been proposed in the literature.

The Composite Estimator proposed in the previous paragraph can be given a model-based justification as well. Consider the model $y_{ij} \overset{\text{ind}}{\sim} (bx_{ij}, \sigma^2 x_{ij})$. Best linear unbised estimator of $b$ is obtained by minimizing $\sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - bx_{ij})^2/x_{ij}$. The solution is $\hat{b} = \bar{y}_s/\bar{x}_s$. Now estimate $\bar{Y}_i = (\sum_{j=1}^{n_i} y_{ij} + \sum_{j=n_i+1}^{N_i} y_{ij})/N_i$ by $\sum_{j=1}^{n_i} y_{ij}/N_i + \hat{b}\sum_{j=n_i+1}^{N_i} x_{ij}/N_i$. This simplifies to the expression given in the previous paragraph. Holt, Smith and Tomberlin (1979) provided more general model-based estimators of this type.

## 5. Model-Based Small Area Estimation

Small area models link explicitly the sampling model with random area specific effects. The latter accounts for between area variation beyond that is explained by auxiliary variables. We classify small area models into two broad types. First, the "area level" models that relate small area direct estimators to area-specific covariates. Such models are necessary if unit (or element) level data are not available. Second, the "unit level" models that relate the unit values of a study variable to unit-specific covariates. Indirect estimators based on small area models will be called "model-based estimators".

The model-based approach to small area estimation offers several advantages. First, "optimal" estimators can be derived under the assumed model. Second, area specific measures of variability can be associated with each estimator unlike global measures (averaged over small areas) often used with traditional indirect estimators. Third, models can be validated from the sample data. Fourth, one can entertain a variety of models depending on the nature of the response variables and the complexity of data structures. Fifth, the use of models permits optimal prediction for areas with no samples, areas where prediction is of utmost importance.

In spite of the above advantages, there should be a cautionary note regarding potential model failure. We will address this issue to a certain extent in Section 7 when we discuss benchmarking. Another important issue that has emerged in recent years, is design-based evaluation of small area predictors. In particular, design-based mean squared errors (MSE's) is of great appeal to practitioners and users of small area predictors, because of their long-standing familiarity with the latter. Two recent articles addressing this issue are Pfeffermann and Ben-Hur (2018) and Lahiri and Pramanik (2019).

The classic small area model is due to Fay and Herriot (JASA, 1979) with Sampling Model: $y_i = \theta_i + e_i,\ i = 1,\ldots,m$ and Linking Model: $\theta_i = x_i^T b + u_i,\ i = 1,\ldots,m$. The target is estimation of the $\theta_i$, $i = 1,\ldots,m$. It is assumed that $e_i$ are independent $(0, D_i)$, where the $D_i$ are known and the $u_i$ are iid $(0, A)$, where $A$ is unknown. The assumption of known $D_i$ can be put to question because they are, in fact, sample estimates. But the assumption is needed to avoid nonidentifiablity in the absence of microdata. This is evident when one writes $y_i = x_i^T b + u_i + e_i$. In the presence of microdata, it is possible to estimate the $D_i$ as well. An example appears in Ghosh, Myung and Moura (2018).

A few notations are needed to describe the Fay-Herriot procedure. Let $y = (y_1,\ldots,y_m)^T$; $\theta = (\theta_1,\ldots,\theta_m)^T$: $e = (e_1,\ldots,e_m)^T$; $u = (u_1,\ldots,u_m)^T$; $X^T = (x_1,\ldots,x_m)$; $b = (b_1,\ldots,b_p)^T$. We assume $X$ has rank $p(< m)$. In vector notations, we write $y = \theta + e$ and $\theta = Xb + u$.

For known $A$, the best linear unbiased predictor (BLUP) of $\theta_i$ is $(1 - B_i)y_i + B_i x_i^T \tilde{b}$ where $\tilde{b} = (X^T V^{-1} X)^{-1} X^T V^{-1} y$, $V = \text{Diag}(D_1 + A, \cdots, D_m + A)$ and $B_i = D_i/(A + D_i)$. The BLUP is also the best unbiased predictor under assumed normality of $y$ and $\theta$.

It is possible to give an alternative Bayesian formulation of the Fay-Herriott model. Let $y_i|\theta_i \overset{\text{ind}}{\sim} \text{N}(\theta_i, D_i)$; $\theta_i|b \overset{ind}{\sim} N(x_i^T b, A)$. Then the Bayes estimator of $\theta_i$ is $(1 - B_i)y_i + B_i x_i^T b$, where $B_i = D_i/(A + D_i)$. If instead we put a uniform$(R^p)$ prior for $b$, the Bayes estimator of $\theta_i$ is the same as its BLUP. Thus, there is a duality between the BLUP and the Bayes estimator.

However, in practice, $A$ is unknown. A hierarchical prior joint for both $b$ and $A$ is $\pi(b,A) = 1$. (Morris, 1983, JASA). Otherwise, estimate $A$ to get the resulting empirical Bayes or empirical BLUP. We now describe the latter.

There are several methods for estimation of $A$. Fay and Herriot (1979) suggested solving iteratively the two equations (i) $\tilde{b} = (X^T V^{-1} X)^{-1} X^T V^{-1} y$ and (ii) $\sum_{i=1}^m (y_i - x_i^T \tilde{b})^2 = m - p$. The motivation for (i) comes from the fact that $\tilde{b}$ is the best linear unbiased estimator (BLUE) of $b$ when $A$ is known. The second is a method of moments equation noting that the expectation of the left hand side equals $m - p$.

The Fay-Herriot method does not provide an explicit expression for $A$. Prasad and Rao (1990, JASA) suggested instead a unweighted least squares approach, which provides an exact expression for $A$. Specifically, they proposed the estimator $\hat{b}_L = (X^T X)^{-1} X^T y$. Then $E||y - X\hat{b}_L||^2 = (m-p)A + \sum_{i=1}^m D_i(1-r_i)$, $r_i = x_i^T (X^T X)^{-1} x_i$, $i = 1, \cdots, m$. This leads to $\hat{A}_L = \max\left(0, \frac{||y - X\hat{b}_L||^2 - \sum_{i=1}^m D_i(1-r_i)}{m-p}\right)$ and accordingly $\hat{B}_i^L = D_i/(\hat{A}_L + D_i)$. The corresponding estimator of $\theta$ is $\hat{\theta}_i^{EB} = (1 - \hat{B}_i^L) y_i + \hat{B}_i^L x_i^T \tilde{b}(\hat{A}_L)$, where

$$\tilde{b}(\hat{A}_L) = [X^T V^{-1}(\hat{A}_L) X]^{-1} X^T V^{-1}(\hat{A}_L) y.$$

Prasad and Rao also found an approximation to the mean squared eror (Bayes risk) of their EBLUP or EB estimators. Under the subjective prior $\theta_i \stackrel{\text{ind}}{\sim} \mathrm{N}(x_i^T b, A)$, the Bayes estimator of $\theta_i$ is $\hat{\theta}_i^B = (1 - B_i) y_i + B_i x_i^T b$, $B_i = D_i/(A + D_i)$. Also, write $\tilde{\theta}_i^{EB}(A) = (1 - B_i) y_i + B_i x_i^T \tilde{b}(A)$. Then $E(\hat{\theta}_i^{EB} - \theta_i)^2 = E(\hat{\theta}_i^B - \theta_i)^2 + E(\tilde{\theta}_i^{EB}(A) - \hat{\theta}_i^{B})^2 + E(\hat{\theta}_i^{EB} - \tilde{\theta}_i^{EB}(A))^2$. The cross-product terms vanish due to their method of estimation of $A$, by a result of Kackar and Harville (1984). The first term is the Bayes risk if both $b$ and $A$ were known. The second term is the additional uncertainty due to estimation of $b$ when $A$ is known. The third term accounts for further uncertainty due to estimation of $A$.

One can get exact expressions $E(\theta_i - \hat{\theta}_i^B)^2 = D_i(1 - B_i) = g_{1i}(A)$, say and $E(\hat{\theta}_i^{EB}(A) - \hat{\theta}_i^B)^2 = B_i^2 x_i^T (X^T V^{-1} X)^{-1} x_i = g_{2i}(A)$, say. However, the third term, $E(\hat{\theta}_i^{EB} - \hat{\theta}_i^{EB}(A))^2$ needs an approximation. An approximate expression correct up to $O(m^{-1})$, i.e. the remainder term is of $o(m^{-1})$, as given in Prasad and Rao, is $2B_i^2 (D_i + A)^{-1} A^2 \sum_{i=1}^m (1 - B_i)^2 / m^2 = g_{3i}(A)$, say. Further, an estimator of this MSE correct up to $O(m^{-1})$ is $g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + 2g_{3i}(\hat{A})$. This approximation is justified by noticing $E[g_{1i}(\hat{A})] = g_{1i}(A) - g_{3i}(A) + o(m^{-1})$.

A well-known example where this method has been applied is estimation of median income of four-person families for the 50 states and the District of Columbia in the United States. The U.S. Department of Health and Human Services (HHS) has a direct need for such data at the state level in formulating its energy assistance program for low-income families. The basic source of data is the annual demographic supplement to the March sample of the Current Population Survey (CPS), which provides the median income of four-person families for the preceding year. Direct use of CPS estimates is usually undesirable because of large CV's associated with them. More reliable results are obtained these days by using empirical and hierarchical Bayesian methods.

Here sample estimates of the state medians for the current year (c) as obtained from the Current Population Survey (CPS) were used as dependent variables. Adjusted census median (c) defined as the base year (the recent most decennial census) census median (b) times the ratio of the BEA PCI (per capita income as provided by the Bureau of Economic Analysis of the United States Bureau of the Census) in year (c) to year (b) was used as an independent variable. Following the suggestion of Fay (1987), Datta, Ghosh, Nangia and Natarajan (1996) used the census median from the recent most decennial census as a second independent variable. The resulting estimates were compared against a different regression model employed earlier by the US Census Bureau.

The comparison was based on four criteria recommended by the panel on small area estimates of population and income set up by the US committee on National Statistics. In the following, we use $e_i$ as a generic notation for the $i$th small area estimate, and $e_{i.TR}$ the "truth", i.e. the figure available from the recent most decennial census. The panel recommended the following four criteria for comparison.
Average Relative Absolute Bias $= (51)^{-1}\sum_{i=1}^{51}|e_i - e_{i,TR}|/e_{i,TR}$.
Average Squared Relative Bias $= (51)^{-1}\sum_{i=1}^{51}(e_i - e_{i,TR})^2/e_{i,TR}^2$.
Average Absolute Bias $= (51)^{-1}\sum_{i=1}^{51}|e_i - e_{i,TR}|$.
Average Squared Deviation $= (51)^{-1}\sum_{i=1}^{51}(e_i - e_{i,TR})^2$.

Table 1 compares the Sample Median, the Bureau Estimate and the Empirical BLUP according to the four criteria as mentioned above.

**Table 1.** Average Relative Absolute Bias, Average Squared Relative Bias, Average Absolute Bias and Average Squared Deviation (in 100,000) of the Estimates.

| | Bureau Estimate | Sample Median | EB |
|---|---|---|---|
| Aver. rel. bias | 0.325 | 0.498 | 0.204 |
| Aver. sq. rel bias | 0.002 | 0.003 | 0.001 |
| Aver. abs. bias | 722.8 | 1090.4 | 450.6 |
| Aver. sq. dev. | 8.36 | 16.31 | 3.34 |

There are other options for estimation of $A$. One due to Datta and Lahiri (2000) uses the MLE or the residual MLE (RMLE). With this estimator, $g_{3i}^{DL}$ is approximated by $2D_i^2(A+D_i)^{-3}[\sum_{i=1}^m(A+D_i)^{-2}]^{-1}$, while $g_{1i}$ and $g_{2i}$ remain unchanged. Finally, Datta, Rao and Smith (2005), went back to the original Fay-Herriot method of estimation of $A$, and obtained $g_{3i}^{DRS} = 2D_i^2(A+D_i)^{-3}m[\sum_{i=1}^m(A+D_i)^{-2}]^{-1}$.

The string of inequalities

$$m^{-1}\sum_{i=1}^{m}(A+D_i)^2 \geq [m^{-1}\sum_{i=1}^{m}(A+D_i)]^2 \geq m^2[\sum_{i=1}^{m}(A+D_i)^{-1}]^2$$

leads to $g_{3i}^{PR} \geq g_{3i}^{DRS}$. Another elementary inequality $\sum_{i=1}^{m}(A+D_i)^{-2} \geq m^{-1}[\sum_{i=1}^{m}(A+D_i)^{-1}]^2$ leads to $g_{3i}^{DRS} \geq g_{3i}^{DL}$. All three expressions for $g_{3i}$ equal when $D_1 = \ldots = D_m$. It is also pointed out in Datta, Rao and Smith that while both Prasad-Rao and REML estimators of $A$ lead to the same MSE estimator correct up to $O(m^{-1})$, a further adjustment to this estimator is needed when one uses either the the ML or the Fay-Herriot estimator of $A$. The simulation study undertaken in Datta, Rao and Smith also suggests that the ML, REML and Fay-Herriot methods of estimation of $A$ perform quite similarly in regards to the MSE of the small area estimators, but the Prasad-Rao approach usually leads to a bigger MSE. However, they all perform far superior to the MSE's of the direct estimators.

Over the years, other approaches to MSE estimation have appeared, some quite appealing as well as elegant. The two most prominent ones appear to be the ones due to Jackknife and Bootstrap. Jiang and Lahiri (2001), Jiang, Lahiri and Wan (2002), Chen and Lahiri (2002), Das, Jiang and Rao (2004) all considered Jackknife estimation of the MSE that avoid the detailed Taylor series expansion of the MSE. A detailed discussion paper covering many aspects of related methods appears in Jiang and Lahiri (2006). Pfeffermann and Tiller (2005), Butar and Lahiri (2003) considered bootstrap estimation of the MSE. More recently, Yoshimori and Lahiri (2014) considered adjusted likelihood estimation of $A$. Booth and Hobert (1998) introduced a conditional approach for estimating the MSE. In a different vein, Lahiri and Rao (1995) dispensed with the normality assumption of the random effects, assuming instead its eighth moment in the Fay-Herriot model.

Pfeffermann and Correa (2012) proposed an approach which they showed to perform much better than the "classical" jackknife and bootstrap methods. Pfeffermann and Ben-Hur (2018) used a similar approach for estimating the design-based MSE of model-based predictors.

Small area estimation problems have also been considered for the general exponential family model. Suppose $y_i|\theta_i$ are independent with $f(y_i|\theta_i) = \exp[y_i\theta_i - \psi(\theta_i) + h(y_i)]$, $i = 1, \ldots, m$. An example is the Bernoulli $(p_i)$ where $\theta_i = \text{logit}(p_i) = \log(p_i/(1-p_i))$ and Poisson($\lambda_i$) where $\theta_i = \log(\lambda_i)$. One models the $\theta_i$ as independent $\text{N}(x_i^T b, A)$ and proceeds. Alternately, use beta priors for the $p_i$ and gamma priors for the $\lambda_i$.

The two options are to estimate the prior parameters either using an empirical Bayes approach or alternately using a hierarchical Bayes approach assigning distributions to the prior parameters. The latter was taken by Ghosh et al. (1998) in a general framework. Other work is due to Raghunathan (1993) and Malec et al. (1997). A method for MSE estimation in such contexts appears in Jiang and Lahiri (2001).

Jiang, Nguyen and Rao (2011) evaluated the performance of a BLUP or EBLUP using only the sampling model $y_i \overset{\text{ind}}{\sim} (\theta_i, D_i)$. Recall $B_i = D_i/(A+D_i)$. Then

$$E[\{(1-B_i)y_i + B_i x_i^T b - \theta_i\}^2 | \theta_i] = (1-B_i)^2 D_i + B_i^2(\theta_i - x_i^T b)^2.$$

Noting that $E[(y_i - x_i^T b)^2|\theta_i] = D_i + (\theta_i - x_i^T b)^2$, an unbiased estimator of the above MSE is $(1-B_i)^2 D_i - B_i^2 D_i + B_i^2(y_i - x_i^T b)^2$. When one minimizes the above with respect to $b$ and $A$, then the resulting estimators of of $b$ and $A$ are referred to as observed best predictive estimators. The corresponding estimators of the $\theta_i$ are referred to as the "observed best predictors". These authors suggested Fay-Herriot or Prasad-Rao method for estimation of $b$ and $A$.

## 6. Model Based Small Area Estimation: Unit Specific Models

Unit Specific Models are those where observations are available for the sampled units in the local areas. In addition, unit-specific auxiliary information is available for these sampled units, and possibly for the non-sampled units as well.

To be specific, consider $m$ local areas where the $i$th local area has $N_i$ units with a sample of size $n_i$. We denote the sampled observations by $y_{i1}, \ldots, y_{in_i}$, $i = 1, \ldots, m$. Consider the model

$$y_{ij} = x_{ij}^T b + u_i + e_{ij}, j = 1, \ldots . N_i, i = 1, \ldots, m.$$

The $u_i$'s and $e_{ij}$'s are mutually independent with the $u_i$ iid $(0, \sigma_u^2)$, and the $e_{ij}$ independent $(0, \sigma^2 \psi_{ij})$.

The above nested error regression model was considered by Battese, Harter and Fuller (BHF, 1988), where $y_{ij}$ is the area devoted to corn or soybean for the $j$th segment in the $i$th county; $x_{ij} = (1, x_{ij1}, x_{ij2})^T$, where $x_{ij1}$ denotes the no. of pixels classified as corn for the $j$th segment in the $i$th county and $x_{ij2}$ denotes the no. of pixels classified as soybean for the $j$th segment in the $i$th county; $b = (b_0, b_1, b_2)^T$ is the vector of regression coefficients. BHF took $\psi_{ij} = 1$. The primary goal of BHF was to estimate the $\bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}$, the population average of area under corn or soybean for the 12 areas in North Central Iowa, $N_i$ denoting the population size in area $i$.

A second example appears in Ghosh and Rao (1994). Here $y_{ij}$ denotes wages and salaries paid by the $j$th business firm in the $i$th census division in Canada and $x_{ij} = (1, x_{ij})^T$, where $x_{ij}$ is the gross business income of the $j$th business firm in the $i$th census division. In this application, $\psi_{ij} = x_{ij}$ was found more appropriate than the usual model involving homoscedasticity.

I consider in some detail the BHF model. Their ultimate goal was to estimate the population means $\bar{Y}_i = (N_i)^{-1} \sum_{j=1}^{N_i} y_{ij}$, In matrix notation, we write $y_i = (y_{i1}, \ldots, y_{in_i})^T$,

$X_i = (x_{i1}, \ldots, x_{in_i})^T$, $e_i = (e_{i1}, \ldots, e_{in_i})^T$, $i = 1, \ldots, m$. Thus, the model is rewritten as

$$y_i = X_i b + u_i 1_{n_i} + e_i, i = 1, \ldots, m.$$

Clearly, $E(y_i) = X_i b$ and $V_i = V(y_i) = \sigma_e^2 I_{n_i} + \sigma_u^2 J_{n_i}$, where $J_{n_i}$ denote the matrix with all elements equal to 1. Write $\bar{x}_i = \sum_{j=1}^{n_i} x_{ij}/n_i$ and $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$. The target is estimation of $\bar{X}_i^T b + u_i 1_{n_i}$, where $\bar{X}_i = N_i^{-1} \sum_{j=1}^{N_i} x_{ij}$, $i = 1, \ldots, m$.

For known $\sigma_u^2$ and $\sigma_e^2$, the BLUP of $\bar{x}_i^T b + u_i 1_{n_i}$ is $(1 - B_i) y_i + B_i \bar{x}_i^T \tilde{b}$, where $B_i = (\sigma_e^2/n_i)/(\sigma_e^2/n_i + \sigma_u^2)$ and $\tilde{b} = (\sum_{i=1}^m X_i^T V_i^{-1} X_i)^{-1} (\sum_{i=1}^m X_i^T V_i^{-1} y_i)$. Hence, the BLUP of $\bar{X}_i^T b + u_i 1_{n_i}$ is $[(1 - B_i)[\bar{y}_i + (\bar{X}_i - \bar{x}_i)^T \tilde{b}] + B_i \bar{X}_i^T \tilde{b}$.

BHF used method of moment estimation to get unbiased estimators of unknown $\sigma_u^2$ and $\sigma_e^2$. The EBLUP of $\bar{X}_i^T b + u_i$ is now found by substituting these estimates of $\sigma_u^2$ and $\sigma_e^2$ in the BLUP formula. Estimation of $\sigma_e^2$ is based on the moment identity

$$E[\sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i - (x_{ij} - \bar{x}_i)^T \tilde{b}]^2 = (n - m - p_1),$$

where $p_1$ is the number of non-zero $x$ deviations. The second moment identity is given by

$$E[\sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - x_{ij})^T \hat{b})^2] = (n - p)\sigma_e^2 + \sigma_u^2 [m - \sum_{i=1}^m n_i^2 \bar{x}_i^T (X^T X)^{-1} \bar{x}_i],$$

where $\hat{b} = (X^T X)^{-1} X^T y$, $y = (y_1^T, \ldots, y_m^T)^T$. If this results in a negative estimator of $\sigma_u^2$, they set the estimator equal to zero.

Of course, the method of moments estimators can be replaced by maximum likelihood, REML or other estimators as discussed in the previous section. Alternately, one can adopt a hierarchical Bayesian approach as taken in Datta and Ghosh (1991). First, it may be noted that if the variance components $\sigma_e^2$ and $\sigma_u^2$ were known, a uniform prior on $b$ leads to a HB estimator of $\bar{X}_i^T b + u_i$, which equals its BLUP. Another interesting observation is that the BLUP of $\bar{X}_i^T b + u_i$ depends only on the variance ratio $\sigma_u^2/\sigma_e^2 = \lambda$, say. Rather than assigning priors separately for $\sigma_u^2$ and $\sigma_e^2$, it suffices to assign a prior to $\lambda$. This is what was proposed in Datta and Ghosh (1991), who assigned a Gamma prior to $\lambda$. The Bayesian approach of Datta and Ghosh (1991) did also acccommodate the possibility of multiple random effects.

## 7. Benchmarking

The model-based small area estimates, when aggregated, may not equal the corresponding estimated for the larger area. On the other hand, the direct estimate for a larger area, for example, a national level estimate, is quite reliable. Moreover, matching the latter may be a good idea, for instance to maintain consistency in publication, and very often for protection against model failure. The latter may not always be achieved, for example in time series models, as pointed out by Wang, Fuller and Qu (2008).

Specifically, suppose $\theta_i$ is the $i$th area mean and $\theta_T = \sum_{i=1}^m w_i\theta_i$ is the overall mean, where $w_j$ may be the known proportion of units in the $j$th area. The direct estimate for $\theta_T$ is $\sum_{i=1}^m w_i\hat{\theta}_i$. Also, let $\tilde{\theta}_i$ denote an estimator of $\theta_i$ based on a certain model. Then $\sum_{i=1}^m w_i\tilde{\theta}_i$ is typically not equal to $\sum_{i=1}^m w_i\hat{\theta}_i$

In order to address this, people have suggested (i) ratio adjusted estimators

$$\hat{\theta}_i^{RA} = \hat{\theta}_i^G(\sum_{j=1}^m w_j\hat{\theta}_j)/(\sum_{j=1}^m w_j\hat{\theta}_j^G)$$

and (ii) difference adjusted estimator $\hat{\theta}_i^{DA} = \hat{\theta}_i^G + \sum_{j=1}^m w_j\hat{\theta}_j - \sum_{j=1}^m w_j\hat{\theta}_j^G$, where $\hat{\theta}_j^G$ is some generic model-based estimator of $\theta_j$.

One criticism against such adjustments is that a common adjustment is used for all small areas regardless of their precision. Wang, Fuller and Qu (2008) proposed instead minimizing $\sum_{j=1}^m \phi_j E(e_j - \theta_j)^2$ for some specified weights $\phi_j (> 0)$ subject to the constraint $\sum_{j=1}^m w_j e_j = \hat{\theta}_T$. The resulting estimator of $\theta_i$ is

$$\hat{\theta}_i^{WFQ} = \hat{\theta}_i^{BLUP} + \lambda_i(\sum_{j=1}^m w_j\hat{\theta}_j - \sum_{j=1}^m w_j\hat{\theta}_j^{BLUP}),$$

where $\lambda_i = w_i\phi_i^{-1}/(\sum_{j=1}^m w_j^2\phi_j^{-1})$.

Datta, Ghosh, Steorts and Maples (2011) took instead a general Bayesian approach and minimized $\sum_{j=1}^m \phi_j[E(e_j - \theta_j)^2|data]$ subject to $\sum_{j=1}^m w_j e_j = \hat{\theta}_T$ and obtained the estimator $\hat{\theta}_i^{AB} = \hat{\theta}_i^B + \lambda_i(\sum_{j=1}^m w_j\hat{\theta}_j - \sum_{j=1}^m w_j\hat{\theta}_j^B)$, with the same $\lambda_i$. This development is similar in spirit to those of Louis (1984) and Ghosh (1992) who proposed constrained Bayes and empirical Bayes estimators to prevent overshrinking. The approach of Datta, Ghosh, Steorts and Maples extends readily to multiple benchmarking constraints. In a frequentist context. Bell, Datta and Ghosh (2013) extended the work of Wang, Fuller and Qu (2008) to multiple benchmarking constraints.

There are situations also when one needs two-stage benchmarking. A current example is the cash rent estimates of the Natural Agricultural Statistics Service (NASS), where one needs the dual control of matching the aggregate of county level cash rent estimates to the corresponding agricultural district (comprising of several counties) level estimates, and the aggregate of the agricultural district level estimates to the final state level estimate. Berg, Cecere and Ghosh (2014) adopted an approach of Ghosh and Steorts (2013) to address the NASS problem.

Second order unbiased MSE estimators are not typically available for ratio adjusted benchmarked estimators. In contrast, second order unbiased MSE estimators are available for difference adjusted benchmarked estimators, namely, $\hat{\theta}_i^{DB} = \hat{\theta}_i^{EB} + (\sum_{j=1}^m w_j \hat{\theta}_j - \sum_{j=1}^m w_j \hat{\theta}_j^{EB})$. Steorts and Ghosh (2013) have shown that MSE$(\hat{\theta}_i^{DB})$ = MSE$(\hat{\theta}_i^{EB}) + g_4(A) + o(m^{-1})$, where MSE$(\hat{\theta}_i^{EB})$ is the same as the one given in Prasad and Rao (1990), and

$$g_4(A) = \sum_{i=1}^m w_i^2 B_i^2 (D_i + A) - \sum_{i=1}^m \sum_{j=1}^m w_i w_j B_i B_j x_i^T (X^T V^{-1} x_j).$$

We may recall that $B_i = D_i/(A+D_i)$, $X^T = (x_1, \ldots, x_m)$ and $V = \text{Diag}(A+D_1, \ldots, A+D_m)$ in the Fay-Herriot model. A second order unbiased estimator of the benchmarked EB estimator is thus $g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + 2g_{3i}(\hat{A}) + g_{4i}(\hat{A})$.

There are two available approaches for self benchmarking that do not require any adjustment to the EBLUP estimators. The first, proposed in You and Rao (2002) for the Fay-Herriot model replaces the estimator $\hat{b}$ in the EBLUP by an estimator which depends both on $\hat{b}$ as well as the weights $w_i$. This changes the MSE calculation. Recall the Prasad-Rao MSE of the EBLUP given by MSE$(\hat{\theta}_i^{EB}) = g_{1i} + g_{2i} + g_{3i}$, where $g_{1i} = D_i(1-B_i)$, $g_{2i} = B_i^2 x_i^T (X^T V^{-1} X)^{-1} x_i$ and $g_{3i} = 2D_i^2 (A+D_i)^{-3} m^{-2} \{\sum_{j=1}^m (A+D_j)^2\}$. For the Benchmarked EBLUP, $g_{2i}$ changes.

The second approach is by Wang, Fuller and Qu (2008) and it uses an augmented model with new covariates $(x_i, w_i, D_i)$. This second approach was extended by Bell, Datta and Ghosh (2013) to accommodate multiple benchmarking constraints.

## 8. Fixed versus Random Area Effects

A different but equally pertinent issue has recently surfaced in the small area literature. This concerns the need for random effects in all areas, or whether even fixed effects models would be adequate for certain areas. Datta, Hall and Mandal (DHM, 2011) were the first to address this problem. They suggested essentially a preliminary test-based approach, testing the null hypothesis that the common random effect variance was zero. Then they used a fixed or a random effects model for small area estimation based on acceptance or rejection of the null hypothesis. This amounted to use of synthetic or regression estimates of all small area means upon acceptance of the null hypothesis, and composite estimates which are weighted averages of direct and regression estimators otherwise. Further research in this area is due to Molina, Rao and Datta (2015).

The DHM procedure works well when the number of small areas is moderately large, but not necessarily when the number of small areas is very large. In such situations, the null hypothesis of no random effects is very likely to be rejected. This is primarily due to a few large residuals causing significant departure of direct estimates from the regression estimates. To rectify this, Datta and Mandal (2015) proposed a Bayesian approach with "spike and slab" priors. Their approach amounts to taking $\delta_i u_i$ instead of $u_i$ for random effects where the $\delta_i$ and the $u_i$ are independent with $\delta_i$ iid Bernoulli($\gamma$) and $u_i$ iid N$(0, \sigma_u^2)$.

In contrast to the spike and slab priors of Datta and Mandal (2015), Tang, Ghosh, Ha and Sedransk (2018) considered a different class of priors that meets the same objective. as the spike and slab priors, but uses instead absolutely continuous priors. These priors allow different variance components for different small areas, in contrast to the priors of Datta and Mandal, who considered prior variances to be either zero or else common across all small areas. This seems to be particularly useful when the number of small areas is very large, for example, when one considers more than 3000 counties of the US, where one expects a wide variation in the county effects. The proposed class of priors, is usually referred to as "global-local shrinkage priors" (Carvalho, Polson and Scott (2010); Polson and Scott (2010)).

The global-local priors, essentially scale mixtures of normals, are intended to capture potential "sparsity", which means lack of significant contribution by many of the random effects, by assigning large probabilities to random effects close to zero, but also identifying random effects which differ significantly from zero. This is achieved by employing two levels of parameters to express prior variances of random effects. The first, the "local shrinkage parameters", acts at individual levels, while the other, the "global shrinkage parameter" is common for all random effects. This is in contrast to Fay and Herriot (1979) who considered only one global parameter. These priors also differ from those of Datta and Mandal (2015), where the variance of random effects is either zero or common across all small areas.

Symbolically, the random effects $u_i$ have independent N$(0, \lambda_i^2 A)$ priors. While the global parameter $A$ tries to cause an overall shrinking effect, the local shrinkage parameters $\lambda_i^2$ are useful in controlling the degree of shrinkage at the local level. If the mixing density corresponding to local shrinkage parameters is appropriately heavy-tailed, the large random effects are almost left unshrunk. The class of "global-local" shrinkage priors includes the three parameter beta normal (TPBN) priors (Armagon, Clyde and Dunson, 2011) and Generalized Double Pareto priors (Armagon, Dunson and Lee, 2012). TPBN includes the now famous horseshoe (HS) priors (Scott and Berger, 2010) and the normal-exponential-gamma priors (Griffin and Brown, 2005).

As an example, consider estimation of 5-year (2007–2011) county-level overall poverty ratios in the US. There are 3,141 counties in the data set. The covariates are foodstamp participation rates. The map given in Figure 1 gives the poverty ratios for all the counties of US. Some salient findings from these calculations are given below.

(i) Estimated poverty ratios are between 3.3% (Borden County, TX) and 47.9% (Shannon County, SD). The median is 14.7%.

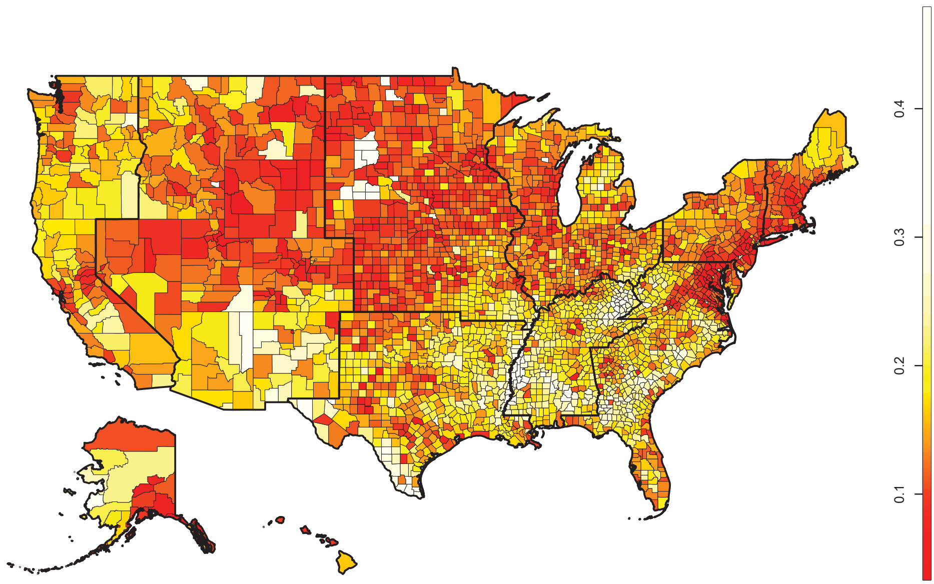


Figure 1: Map of posterior means of $\theta$'s.

(ii) In Mississippi, Georgia, Alabama and New Mexico, 55%+ counties have poverty rates $>$ the third quartile (18.9%).
(iii) In New Hampshire, Connecticut, Rhode Island, Wyoming, Hawaii and New Jersey, 70%+ counties have poverty rates $<$ the first quartile (11.1%).
(iv) Examples of counties with high poverty ratios are Shannon, SD; Holmes, MS; East Carroll, LA; Owsley, KY; Sioux, IA.
(v) Examples of counties with large random effects are Madison, ID; Whitman, WA; Harrisonburg, VA; Clarke, GA; Brazos, TX.

Dr. Pfeffermann suggested splitting the counties, whenever possible, into a few smaller groups, and then use the same global-local priors for estimating the random effects separately for the different groups. From a pragmatic point of view, this may sometimes be necessary for faster implementation. It seems though that the MCMC implementation even for such a large number of counties was quite easy since all the conditionals were standard disributions, and samples could be generated easily from these distributions at each iteration.

## 9. Variable Transformation

Often the normality assumption can be justified only after transformation of the original data. Then one performs the analysis based on the transformed data, but transform back properly to the original scale to arrive at the final predictors. One common example is transformation of skewed positive data, for example, income data where log transformation gets a closer normal approximation. Slud and Maiti (2006) and Ghosh and Kubokawa (2015) took this approach, providing final results for the back-transformed original data.

For example, consider a multiplicative model $y_i = \phi_i \eta_i$ with $z_i = \log(y_i)$, $\theta_i = \log(\phi_i)$ and $e_i = \log(\eta_i)$. Consider the Fay-Herriott (1979) model (i) $z_i|\theta_i \overset{\text{ind}}{\sim} \mathrm{N}(\theta_i, D_i)$ and (ii) $\theta_i \overset{\text{ind}}{\sim} \mathrm{N}(x_i^T\beta, A)$. $\theta_i$ has the $\mathrm{N}(\hat{\theta}_i^B, D_i(1-B_i))$ posterior with $\hat{\theta}_i^B = (1-B_i)z_i + B_i x_i^T\beta$, $B_i = D_i/(A+D_i)$. Now $E(\phi_i|z_i) = E[\exp(\theta_i)|z_i] = \exp[\hat{\theta}_i^B + (1/2)D_i(1-B_i)]$.

Another interesting example is the variance stabilizing transformation. For example, suppose $y_i \overset{\text{ind}}{\sim} \mathrm{Bin}(n_i, p_i)$. The arcsine transformation is given by $p_i = \sin^{-1}(2p_i - 1)$. The back transformation is $p_i = (1/2)[1+\sin(\theta_i)]$.

A third example is the Poisson model for count data. There $y_i \overset{\text{ind}}{\sim} \mathrm{Poisson}(\lambda_i)$. Then one models $z_i = y_i^{1/2}$ as independent $\mathrm{N}(\theta_i, 1/4)$ where where $\theta_i = \lambda_i^{1/2}$. An added advantage in the last two examples is that the assumption of known sampling variance, which is really untrue, can be avoided.

## 10. Final Remarks

As acknowledged earlier, the present article leaves out a large number of useful current day topics in small area estimation. I list below a few such topics which are not covered at all here. But there are are many more. People interested in one or more of the topics listed below and beyond should consult the book of Rao and Molina (2015) for their detailed coverage of small area estimation and an excellent set of references for these topics.

- Design consistency of small area estimators.
- Time series models.
- Spatial and space-time models.
- Variable Selection.
- Measurement errors in the covariates.
- Poverty counts for small areas.
- Empirical Bayes confidence intervals.
- Robust small area estimation.
- Misspecification of linking models.

- Informative sampling.
- Constrained small area estimation.
- Record Linkage.
- Disease Mapping.
- Etc, Etc., Etc.

## Acknowledgements

I am indebted to Danny Pfeffermann for his line by line reading of the manuscript and making many helpful suggestions, which improved an earlier version of the paper. Partha Lahiri read the original and the revised versions of this paper very carefully, and caught many typos. A comment by J.N.K. Rao was helpful. The present article is based on the Morris Hansen Lecture delivered by Malay Ghosh before the Washington Statistical Society on October 30, 2019. The author gratefully acknowledges the Hansen Lecture Committee for their selection.

## REFERENCES

ARMAGAN, A., CLYDE, M., and DUNSON, D. B., (2013). Generalized double pareto shrinkage. *Statistica Sinica*, 23, pp. 119–143.

ARMAGAN, A., DUNSON, D. B., LEE, J., and BAJWA, W. U., (2013). Posterior consistency in linear models under shrinkage priors. *Biometrika*, 100, pp. 1011–1018.

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, pp. 28–36.

BELL, W. R., DATTA, G, S., and GHOSH, M., (2013). Benchmarking small area estimators. *Biometrika*, 100, pp. 189–202.

BELL, W. R., BASEL, W. W., and MAPLES, J. J., (2016). An overview of U.S. Census Bureau's Small Area Income and Poverty Estimation Program. In *Analysis of Poverty Data by Small Area Estimation*. Ed. M. Pratesi. Wiley, UK, pp. 349–378.

BERG, E., CECERE, W., and GHOSH, M., (2014). Small area estimation of county level farmland cash rental rates. *Journal of Survey Statistics and Methodology*, 2, pp. 1–37. Bivariate hierarchical Bayesian model for estimating cropland cash rental rates at the county level. *Survey Methodology*, in press.

BOOTH, J. G., HOBERT, J., (1998). Standard errors of prediction in generalized linear mixed models. *Journal of the American Statistical Association*, 93, pp. 262–272.

BUTAR, F. B., LAHIRI, P., (2003). On measures of uncertainty of empirical Bayes small area estimators. *Journal of Statistical Planning and Inference*, 112, pp. 63–76.

CARVALHO, C. M., POLSON, N. G., SCOTT, J. G., (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97, pp. 465–480.

CHEN, S., LAHIRI, P., (2003).A comparison of different MPSE estimators of EBLUP for the Fay-Herriott model. In *Proceedings of the Section on Survey Research Methods*. Washington, D.C. American Statistical Association, pp. 903–911.

DAS, K., JIANG, J., RAO, J. N. K., ((2004). Mean squared error of empirical predictor. *Annals of Statistics*, 32, pp. 818–840.

DATTA, G. S., GHOSH, M., (1991). Bayesian prediction in linear models: applications to small area estimation. *The Annals of Statistics*, 19, pp. 1748–1770.

DATTA, G., GHOSH, M., NANGIA, N., and NATARAJAN, K., (1996). Estimation of median income of four-person families: a Bayesian approach. In *Bayesian Statistics and Econometrics: Essays in Honor of Arnold Zellner*. Eds. D. Berry, K. Chaloner and J. Geweke. North Holland, pp. 129–140.

DATTA, G. S., LAHIRI. P., (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, 10, pp. 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, pp. 183–196.

DATTA, G. S., GHOSH, M., STEORTS, R., and MAPLES, J. J., (2011). Bayesian benchmarking with applications to small area estimation. *TEST*, 20, pp. 574–588.

DATTA, G. S., HALL, P., and MANDAL, A., (2011). Model selection and testing for the presence of small area effects and application to area level data. *Journal of the American Statistical Association*, 106, pp. 362–374.

DATTA, G. S., MANDAL, A., (2015). Small area estimation with uncertain random effects. *Journal of the American Statistical Association*, 110, pp. 1735–1744.

ELBERS, C., LANJOUW, J. O., and LANJOUW, P., (2003). Micro-level estimation of poverty and inequality. *Econometrica*, 71. pp. 355–364.

ERCIULESCU, A. L., FRANCO, C., and LAHIRI, P., (2020). Use of administrative records in small area estimation. To appear in *Administrative Records for Survey Methodology*. Eds. P. Chun and M. Larson. Wiley, New York.

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

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

GHOSH, M., (1992). Constrained Bayes estimation with applications. *Journal of the American Statistical Association*, 87, pp. 533–540.

GHOSH, M., RAO, J. N. K., (1994). Small area estimation: an appraisal. *Statistical Science*, pp. 55–93.

GHOSH, M., NATARAJAN, K., STROUD, T. M. F., and CARLIN, B. P., (1998). Generalized linear models for small area estimation. *Journal of the American Statistical Association*, 93, pp. 273–282.

GHOSH, M., STEORTS, R., (2013). Two-stage Bayesian benchmarking as applied to small area estimation. *TEST*, 22, pp. 670–687.

GHOSH, M., KUBOKAWA, T., and KAWAKUBO, Y., (2015). Benchmarked empirical Bayes methods in multiplicative area-level models with risk evaluation. *Biomerika*, 102, pp. 647–659.

GHOSH, M., MYUNG, J., and MOURA, F. A. S., (2018). Robust Bayesian small area estimation. *Survey Methodology*, 44, pp. 101–115.

GONZALEZ, M. E., HOZA, C., (1978). Small area estimation with application to unemployment and housing estimates. *Journal of the American Statistical Association*, 73, pp. 7–15.

GRIFFIN, J. E., BROWN, P. J., (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5, pp. 171–188.

HANSEN, M. H., HURWITZ, W. N., and MADOW, W. G., (1953). *Sample Survey Methods and Theory*. Wiley, New York.

HOLT, D., SMITH, T. M. F., and TOMBERLIN, T. J., (1979). A model-based approach for small subgroups of a population. *Journal of the American Statistical Association*, 74, pp. 405–410.

JIANG, J., LAHIRI, P., (2001). Empirical best prediction of small area inference with binary data. *Annals of the Institute of Statistical Mathematics*, 53, pp. 217–243.

JIANG, J., LAHIRI, P., and WAN, S-M., (2002). A unified jackknife theory. *The Annals of Statistics*, 30, pp. 1782–1810.

JIANG, J., LAHIRI, P., (2006). Mixed model prediction and small area estimation (with discussion). *TEST*, 15, pp. 1–96.

JIANG, J., NGUYEN, T., and RAO, J. S., (2011). Best predictive small area estimation. *Journal of the American Statistical Association*, 106, pp. 732–745.

KACKAR, R. N., HARVILLE, D. A., (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association*, 79, pp. 853–862.

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

LAHIRI, P., PRAMANIK, S., (2019). Evaluation of synthetic small area estimators using design-based methods. *Austrian Journal of Statistics*, 48, pp. 43–57.

LOUIS, T. A., (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods. *Journal of the American Statistical Association*, 79, pp. 393–398.

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

MOLINA, I., RAO, J. N. K., (2010). Small area estimation of poverty indicators. *Canadian Journal of Statistics*, 38, pp. 369–385.

MOLINA, I., RAO, J. N. K., and DATTA, G. S., (2015). Small area estimation under a Fay-Herriot model with preliminary testing for the presence of random effects. *Survey Methodology*.

PFEFFERMANN, D., TILLER, R. B., (2005). Bootstrap approximation of prediction MSE for state-space models with estimated parameters. *Journal of Time Series Analysis*, 26, pp. 893–916.

MORRIS, C. N., (1983). Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association*, 78, pp. 47–55.

POLSON, N. G., SCOTT, J. G., (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction. *Bayesian Statistics*, 9, pp. 501–538.

PFEFFERMANN, D., (2002). Small area estimation: new developments and direction. *International Statistical Review*, 70, pp. 125–143.

PFEFFERMANN, D., (2013). New important developments in small area estimation. *Statistical Science*, 28, pp. 40–68.

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

RAGHUNATHAN, T. E., (1993). A quasi-empirical Bayes method for small area estimation. *Journal of the American Statistical Association*, 88, pp. 1444–1448.

RAO, J. N. K., (2003). Some new developments in small area estimation. *Journal of the Iranian Statistical Society*, 2, pp. 145–169.

RAO, J. N. K., (2006). Inferential issues in small area estimation: some new developments. *Statistics in Transition*, 7, pp. 523–526.

RAO, J. N. K., Molina, I., (2015). *Small Area Estimation*, 2nd Edition. Wiley, New Jersey.

SCOTT, J. G., BERGER, J. O., (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, 38, pp. 2587–2619.

SLUD, E. V., MAITI, T., (2006). Mean squared error estimation in transformed Fay-Herriot models. *Journal of the Royal Statistical Society, B*, 68, pp. 239–257.

TANG, X., GHOSH, M., Ha, N-S., and SEDRANSK, J., (2018). Modeling random effects using global-local shrinkage priors in small area estimation. *Journal of the American Statistical Association*, 113, pp. 1476–1489.

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

YOSHIMORI, M., LAHIRI, P., (2014). A new adjusted maximum likelihood method for the Fay-Herriott small area model. *Journal of Multivariate Analysis*, 124, pp. 281–294.

YOU, Y., RAO, J. N. K., and HIDIROGLOU, M. A., (2013). On the performance of self-benchmarked small area estimators under the Fay-Herriott area level model. *Survey Methodology*, 39, pp. 217–229.