*STATISTICA, anno LXIX, n. 4, 2009*

# A COMPARISON OF ADJUSTED BAYES ESTIMATORS OF AN ENSEMBLE OF SMALL AREA PARAMETERS

Enrico Fabrizi

## 1. INTRODUCTION

In recent years sample surveys have been characterized by a growing demand for estimates of population descriptive quantities for domains (or 'areas') obtained classifying the target population according to geography or other criteria. As the sample portion pertaining to domains is often too small to allow for reliable estimation using standard design-based estimators, small area estimation methods have become a relevant research topic (see Rao, 2003 or Lahiri and Jiang, 2006 for a general introduction). Empirical and hierarchical Bayes methods are an important chapter of small area estimation theory and are also widely applied in practice (see Rao, 2003 Chapters 9 and 10 and the references therein). The basic idea behind these methods is to treat domain descriptive quantities of interest (e.g. means, totals, proportions) as random and to estimate them using some summary of their posterior distribution, typically the posterior mean, often referred to as 'Bayes estimator' (Ghosh, 1992).

Bayes estimators may be very effective in improving the precision (sampling Mean Square Error) of 'direct' design-unbiased (or design-consistent) estimators, but this improvement is often achieved at the cost of shrinking the estimates toward a synthetic estimator which is obtained pooling together data from all areas under study. For this reason, Bayes estimators may be proven to be poor for estimating the actual distribution function of a population ('ensemble') of small area parameters (Louis, 1984, Heady and Ralphs, 2004). The interest in the distribution function may be crucial when small area estimates are used in substantive applications such as the analysis of regional disparities in the distribution of economic indicators of poverty and income inequality (Fabrizi *et al.* 2005). A proper representation of the distribution of the ensemble of the parameters is, for these purposes, as important as to dispose of reliable estimates for each sub-national region.

In this paper we discuss the popular Fay-Herriot model (Fay and Herriot, 1979) and a set of adjusted estimators associated to it. With adjusted estimators we mean estimators of the small area parameters that enjoy acceptable properties with respect to the estimation of empirical distribution function (EDF) or other nonlinear functionals of the population ('ensemble') of small area parameters.

The main goal of the paper is to review adjusted estimators within the framework of hierarchical Bayesian modeling and to compare their frequentist properties by means of a Monte Carlo exercise. In particular we focus on: *i*) the distance between the estimated and the true distribution function; *ii*) efficiency as measured by mean square error; *iii*) robustness with respect to the failure of the assumption of normality of the random effects. We emphasize frequentist properties since these are usually relevant for practitioners and more familiar to final data users. Moreover, we will use the same simulation exercise to evaluate whether posterior mean square errors, a natural measure of uncertainty associated to adjusted Bayes estimators, are also good frequentist measures of variability.

As anticipated, we consider the effects of misspecification concerning the random effects: in particular we focus on wrong distributional assumptions. This type of misspecification is quite likely in practice since random effects are not directly observable and departures from normality are difficult to detect (see Sinharay and Stern, 2003). We explore the properties of the adjusted estimators when the models assume normality but the random effects are actually generated by an alternative distribution.

The paper is organized as follows. In section 2, we shortly discuss the failure of Bayes estimators associated to the univariate Fay-Herriot model as estimators of the variance (and thus of the EDF) of the 'ensemble' of parameters. Among the many adjusted estimators discussed in the literature we focus on constrained Bayes estimators (Ghosh, 1992), constrained linear Bayes estimators (Spjøtvoll and Thomsen, 1987) and a simultaneous estimation method proposed by Zhang (2003). These estimators are reviewed in section 3. The simulation exercise and the tools used in comparisons are introduced in section 4. Although all simulations use populations generated under normality, we will focus on the accuracy of EDF estimation and not just on mean and variance (as in Judkins and Liu, 2000) since with a finite number of areas, the EDF of the population of area parameters may show some slight deviation from normality.

Section 5 focuses on simulation's results; the behaviour of adjusted estimators when the normality of random effects fails is discussed in Section 6. Concluding remarks are provided in Section 7.

## 2. FAILURE OF BAYES ESTIMATORS AS ENSEMBLE ESTIMATORS IN THE FAY-HERRIOT MODEL

The Fay-Herriot model may be described by the following set of assumptions:

$$y_i = \theta_i + e_i \tag{1}$$

$$\theta_i = \mathbf{x}_i^t \beta + v_i \tag{2}$$

$$e_i \overset{ind}{\sim} N(0, \psi_i) \tag{3}$$

$$v_i \overset{ind}{\sim} N(0, \sigma_v^2) \tag{4}$$

where $\{y_i\}$ $1 \le i \le m$ is a collection of 'direct' design-unbiased (or approximately design-unbiased) estimators of a set of small area population parameters $\{\theta_i\}$; $\{\psi_i\}$ is the set of assumed known design-based variances associated to direct estimators and $\mathbf{x}_i$ a $k \times 1$ vector of auxiliary information accurately known for area *i*. Moreover it is assumed that

$$E(e_i v_i) = 0 \,. \tag{5}$$

Small area analyses are somewhat idiosyncratic as they mix randomization and model based probability spaces. More precisely, once denoted $E_D(\,)$, $V_D(\,)$ the expectation and variance with respect to the randomization (design) distribution and $E_M(\,)$, $V_M(\,)$ the moments related to the model or data generating process, assumptions (1) – (4) imply that $E_M(y_i \mid \theta_i) = E_D(y_i) = \theta_i$ and $V_M(y_i \mid \theta_i) = V_D(y_i) = \psi_i$, that is the first two moments of $y_i$ according to the model (conditional on $\theta_i$) and randomization distributions are the same. To be consistent in notation let's also write $E_M(\theta_i) = \mathbf{x}_i^t \beta$, $V_M(\theta_i) = \sigma_v^2$ and $E_M(e_i v_i) = 0$.

Assuming $\beta$ and $\sigma_v^2$ as known, the posterior means of $\theta_i$ under model (1) – (5) are given by $\hat{\theta}_i^B = \gamma_i y_i + (1-\gamma_i)\mathbf{x}_i^t \beta$, where $\gamma_i = \sigma_v^2(\sigma_v^2 + \psi_i)^{-1}$ is a shrinkage factor that gives to the direct estimator a weight that is a decreasing function of its sampling variance.

We have that direct estimators $\{y_i\}$ are overdispersed with respect to the underlying $\{\theta_i\}$ and the small area estimators are under-dispersed. The following proposition holds.

*Proposition 1*

Under model (1) – (5) and assuming $\psi_i = \psi$ we have that:

$$1)\ E_M\left[\frac{1}{m-1}\sum_{i=1}^{m}(\theta_i - \overline{\theta})^2\right] = \sigma_v^2 + \beta^t \Sigma_{xx} \beta \tag{6}$$

with $\Sigma_{xx} = (m-1)^{-1}\sum_{i=1}^{m}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^t$, $\overline{\mathbf{x}} = m^{-1}\sum_{i=1}^{m}\mathbf{x}_i$,

$$2)\ E_M E_D\left[\frac{1}{m-1}\sum_{i=1}^{m}(y_i - \overline{y})^2\right] = \frac{\sigma_v^2}{\gamma} + \beta^t \Sigma_{xx} \beta \ge \sigma_v^2 + \beta^t \Sigma_{xx} \beta \tag{7}$$

$$3)\ E_M E_D\left[\frac{1}{m-1}\sum_{i=1}^{m}(\hat{\theta}_i - \hat{\overline{\theta}})^2\right] = \gamma\sigma_v^2 + \beta^t \Sigma_{xx} \beta \le \sigma_v^2 + \beta^t \Sigma_{xx} \beta \tag{8}$$

A proof may be found in the appendix. Note that $\psi_i = \psi$ does not represent a restrictive condition but is helpful to obtain simple formulas. We may observe that the shrinkage factor rules both overdispersion of direct estimates and under-dispersion of posterior means.

## 3. ADJUSTED BAYES ESTIMATORS

We consider three different adjusted Bayes estimators associated to the Fay-Herriot model : *i*) the constrained Bayes, *ii*) the constrained linear Bayes (Spjøtvoll and Thomsen method), *iii*) the estimators based on a simultaneous estimation method proposed by Zhang (2003). These estimators will be reviewed within a hierarchical Bayes framework; that is, we do not assume the hyperparameters $\beta$, $\sigma_v^2$ as known but specify a prior distribution for them. In what follows we denote the data on which the analysis is conditioned as $\mathbf{z} = \{y_i, \psi_i, \mathbf{x}_i\}_{1 \le i \le m}$.

We anticipate that all adjusted estimators $\hat{\theta}_i^*$ are summaries of the posterior distributions $p(\theta_i \mid \mathbf{z})$ different from the posterior mean $\hat{\theta}_i^{HB} = E(\theta_i \mid \mathbf{z})$ and are therefore suboptimal with respect to quadratic loss. Uncertainty associated to these alternative posterior summaries may be measured by the posterior mean square error:

$$PMSE(\hat{\theta}_i^*) = V(\theta_i \mid \mathbf{z}) + (\hat{\theta}_i^* - \hat{\theta}_i^{HB})^2 \tag{9}$$

We denote $E(\ )$ the expectation with respect to the model omitting the subscript $M$, since randomization moments are no longer involved.

Of course, $PMSE(\hat{\theta}_i^*) \ge V(\theta_i \mid \mathbf{z})$; the better representation of the Empirical Distribution Function of the population of Small Area parameters is paid at the price of some loss of efficiency.

### 3.1 *The Constrained hierarchical Bayes estimator*

Constrained Bayes estimators have been introduced and discussed by Louis (1984) under normality and by Ghosh (1992) under less restrictive distributional assumptions. The aim is to obtain a set of estimators $\{t_i\}$, $i = 1,...,m$, optimal under quadratic loss and satisfying the following constraints:

1) $\bar{t} = \hat{\bar{\theta}}^{HB}$

2) $(m-1)^{-1} \sum_{i=1}^{m} (t_i - \bar{t})^2 = E\left[ \sum_{i=1}^{m} (\theta_i - \bar{\theta})^2 \mid \mathbf{z} \right]$

The constrained hierarchical Bayes (CHB) estimators (see Rao, 2003, section 10.13) are given by:

$$\hat{\theta}_i^{CHB} = \hat{\bar{\theta}}^{HB} + a(\mathbf{z})(\hat{\theta}_i^{HB} - \hat{\bar{\theta}}^{HB}) \tag{10}$$

where

$$a(z) = \left[ 1 + \frac{\left(\sum_{i=1}^{m} V(\theta_i - \bar{\theta}) \,|\, \mathbf{z}\right)}{(m-1)^{-1} \sum_{i=1}^{m} (\hat{\theta}_i^{HB} - \hat{\bar{\theta}}^{HB})^2} \right]^{\frac{1}{2}}$$

The posterior mean square error $PMSE(\hat{\theta}_i^{CHB})$ may be used to evaluate the uncertainty associated to this estimator.

3.2 *The constrained hierarchical linear Bayes Estimator*

Let's suppose, for the moment, that the hyperparameters $\beta$, $\sigma_v^2$ are known. The Constrained Linear Bayes estimator of $\theta_i$ is a summary of the posterior distribution in the form $\hat{\theta}_i^L = a_i y_i + b_i$ satisfying the constraints:

1) $E(\hat{\theta}_i^L) = \mathbf{x}_i^t \beta$,

2) $E(\hat{\theta}_i^L - \mathbf{x}_i^t \beta)^2 = \sigma_v^2$.

Note that when the posterior mean may be expressed in linear form, the distribution enjoys posterior linearity (Goldstein, 1975), a condition that holds for a variety of conjugate models. The constrained linear Bayes estimator is given by:

$$\hat{\theta}_i^{CLB} = \gamma_i^{1/2} y_i + (1 - \gamma_i^{1/2}) \mathbf{x}_i^t \beta \tag{11}$$

with $\gamma_i = \sigma_v^2 (\sigma_v^2 + \psi_i)^{-1}$ (see Spjøtvoll and Thomsen, 1987 and Rao, 2003, section 9.8). This estimator owes its popularity to its similarity to $\hat{\theta}_i^B$: it is still a linear combination of $y_i$ and $\mathbf{x}_i^t \beta$ that, with respect to $\hat{\theta}_i^B$, puts more weight on the 'direct' estimator $y_i$, whereby leading to a set of estimates less shrunken toward the synthetic component.

The estimator (11) may be thought as conditional on ($\beta$, $\sigma_v^2$). We simply propose to define the Constrained Hierarchical Linear Bayes (CHLB) estimator as

$$\hat{\theta}_i^{CHLB} = E_{(\beta, \sigma_v^2 | \mathbf{z})}(\hat{\theta}_i^{CLB}) \tag{12}$$

where $E_{(\beta, \sigma_v^2 | \mathbf{z})}$ is the expectation taken with respect to the posterior distribution of $\beta$, $\sigma_v^2$. An explicit formula for (12) depends on the chosen prior distributions and may be in general difficult to work out. Nonetheless (12) may be easily approximated using the output of Markov Chains Monte Carlo (MCMC) algorithms of common use for the analysis of hierarchical models.

### 3.3 *The simultaneous estimation method proposed by Zhang*

Given the set $\{\theta_i\}$ of the area parameters of interest, let $\{\theta_{(i)}\}$ be the associated ordered set $(\theta_{(1)} \leq \theta_{(2)} \leq ... \leq \theta_{(m)})$. Then $\eta_i = E(\theta_{(i)} \mid \mathbf{z})$ is the best predictor of $\theta_{(i)}$ under quadratic loss and $\{\eta_i\}$ is the best 'ensemble' estimator of $\{\theta_i\}$ under quadratic loss. The set of estimators $\{\eta_i\}$ is not area specific in that its single elements are not associated to specific areas. To match the $\{\eta_i\}$ with the small areas Zhang (2003) proposes, under an empirical Bayes estimation approach, to estimate the ranks of $\{\theta_i\}$ using those of the $\{E(\theta_i \mid \mathbf{z})\}$ set. By the way, the ranks of the posterior means may be poor estimators of actual ranks, especially if there is considerable variability in the posterior variances. Following Ghosh and Maiti (1999) we propose $\hat{r}_i = E(rank(\theta_i \mid \mathbf{z}))$, the posterior expectation of ranks, as the estimator of ranks required in order to match the ensemble estimator $\{\eta_i\}$ with the areas. In the context of hierarchical Bayes modeling, this estimator of ranks may be easily approximated from the output of MCMC algorithms. More specifically, we can rank the $\theta_i(s) \mid \mathbf{z}$ from any draw $s$ of the Markov Chain after convergence. Then we can approximate $\hat{r}_i$ averaging the ranks $rank(\theta_i(s) \mid \mathbf{z})$ over all draws, obtaining $\hat{r}_i^{MC} = S^{-1} \sum_{s=1}^{S} rank(\theta_i(s) \mid \mathbf{z})$ where $S$ is the number of iterations of Markov Chain after convergence used for the estimation of the posterior distribution.

To summarize, the estimator based on Zhang ideas implemented in the context of hierarchical Bayes modeling is given by:

$$\hat{\theta}_i^{ZHB} = \eta_{\hat{r}_i} \tag{13}$$

with $\hat{r}_i$ approximated by $\hat{r}_i^{MC}$ when the posterior distributions are obtained using MCMC algorithms.

## 4. A SIMULATION EXPERIMENT UNDER THE ASSUMPTION OF NORMALITY FOR THE RANDOM EFFECTS

In this section we introduce a simulation experiment whose aim is to compare the effectiveness of the adjusted estimators discussed above in correcting the overshrinkage. We compare also their efficiency and consider whether the $PMSE$ are adequate estimators of their frequentist mean square errors. More specifically, we compare the direct estimators $\hat{\theta}^{DIR} = \{y_i\}$, the posterior means $\hat{\theta}^{HB} = \{\hat{\theta}_i^{HB}\}$ and the various adjusted estimators $\hat{\theta}^{CHB} = \{\hat{\theta}_i^{CHB}\}$, $\hat{\theta}^{CHLB} = \{\hat{\theta}_i^{CHLB}\}$ and $\hat{\theta}^{ZHB} = \{\hat{\theta}_i^{ZHB}\}$.

The simulation is based on R=1,000 Monte Carlo (MC) samples, and all comparisons are referred to the empirical distribution of the various estimators in this replication space.

Data are generated according to the Fay-Herriot model (1) – (5) setting for simplicity $\mathbf{x}_i^t \beta = \mu = 0$.

We consider both the case of moderate and large number of areas setting $m = 30,\ 100$. We set $\sigma_v^2 = 1$ and consider three different configurations of design variances. They are chosen in the following way: we divide the set of areas in five groups. Variances vary across groups but are constant within them. All configurations are illustrated in Table 1. They differ according to informativeness of direct estimators, as measured by $\gamma_i = \sigma_v^2(\sigma_v^2 + \psi_i)^{-1}$. We evaluate the informativeness of direct estimators in comparison to the dispersion of the $\theta_i$ values around the synthetic component $\mu$.

Population 1 describes a situation where direct estimators show a wide range of informativeness ($\gamma_i \in [0.11, 0.91]$); Population 2 describes a situation in which direct estimators are poorly informative ($\gamma_i \in [0.11, 0.5]$), while in Population 3 we study the case of rather strongly informative direct estimators ($\gamma_i \in [0.5, 0.91]$).

TABLE 1

*Different configurations of design variances for the simulation*

| | Design variances | | | | |
|---|---|---|---|---|---|
| | $\psi_{1,\dots,\frac{m}{5}}$ | $\psi_{\frac{m}{5}+1,\dots,\frac{2m}{5}}$ | $\psi_{\frac{2m}{5}+1,\dots,\frac{3m}{5}}$ | $\psi_{\frac{3m}{5}+1,\dots,\frac{4m}{5}}$ | $\psi_{\frac{4m}{5}+1,\dots,m}$ |
| Population 1 | 0.1 | 0.333 | 1 | 3 | 10 |
| Population 2 | 1 | 1.333 | 2 | 4 | 10 |
| Population 3 | 0.1 | 0.25 | 0.5 | 0.75 | 1 |

We do not consider equal sampling variances, i.e. $\psi_i = \psi$ since this is seldom the case in practice; moreover it may be proven that, for $\mathbf{x}_i^t \beta = \mu$ and $\mu$, $\sigma_v^2$ known, $\hat{\theta}_i^{CHB} \to \hat{\theta}_i^{CHLB}$ for $m \to \infty$ (Rao, 2003, section 9.6). As a consequence, even if we do not assume $\mu$ and $\sigma_v^2$ as known and work with a finite number of areas, we may expect the two estimators to show close performances.

When modeling, it is assumed that $\mu$ and $\sigma_v^2$ are unknown with prior distributions $p(\mu, \sigma_v^2) = p(\mu) p(\sigma_v^2)$ with $\mu \sim N(0, K)$, $\sigma_v \sim Unif(0, L)$ where $K = 100$ and $L = 20$ are large constants with respect to the scale of the data.

This priors guarantee properness of the posterior distributions, very mild impact on posterior distributions and good behavior (fast convergence and good mixing) of MCMC algorithms (Gelman, 2006).

To compare estimators, we consider the indicators described below.

*i) Overshrinkage correction*. Let's define

$$AV(\hat{\theta}^*) = \mathrm{R}^{-1}\sum_{r=1}^{R} v^2(\hat{\theta}_r^*) \tag{14}$$

where $v^2(\hat{\theta}_r^*) = (m-1)^{-1}\sum_{i=1}^{m}(\hat{\theta}_{i,r}^* - \bar{\hat{\theta}}_r^*)^2$, and $* = \{DIR,\ HB,\ CHB,\ CHLB,\ ZHB\}$. Since we set $\mathbf{x}_i^t\beta = \mu$ and $\sigma_v^2 = 1$, we expect this indicator to be larger than 1 for $\hat{\theta}^{DIR}$, less than 1 for $\hat{\theta}^{HB}$ and close to 1 for the remaining estimators.

*ii) Kolmogorov-Smirnov distance*. For each iteration $r$, the Kolmogorov-Smirnov distance between the EDF of the adjusted estimator $\hat{\theta}_r^* = \{\hat{\theta}_{i,r}\}$ and that of the 'true values' $\{\theta_{i,r}\}$ denoted respectively as $F_*$ and $F_T$ is calculated as $D_r(\hat{\theta}_r^*, \theta_r) = \max_j \left|F_*(u_{j,r}) - F_T(u_{j,r})\right|$ where $j = 1,...,2m$ and $\mathbf{u}_r = (\hat{\theta}_r^*, \theta_r)$ is the $2m$ vector obtained pooling together the 'true values' and those obtained with the adjusted predictor for the *r*-th MC replication. The distances calculated at each iteration are then averaged over MC replications to obtain $\bar{D}(\hat{\theta}^*, \theta) = \mathrm{R}^{-1}\sum_{r=1}^{R} D_r(\hat{\theta}_r^*, \theta_r)$. For the ease of comparison we report

$$\tilde{D}(\hat{\theta}^*, \theta) = \frac{\bar{D}(\hat{\theta}^*, \theta)}{\bar{D}(\hat{\theta}^{HB}, \theta)} \tag{15}$$

whereby assuming the non adjusted hierarchical Bayes estimators as a benchmark.

*iii) Anderson-Darling distance*. Anderson and Darling (1954) introduced a goodness-of-fit statistic that can be used to evaluate the distance of an EDF from a continuous reference distribution. Compared to the Kolmogorov-Smirnov distance, it is known to be influenced to a greater extent by the discrepancies in the tails of the distribution. In our particular context the 'empirical' Anderson-Darling distance is defined as follows

$$A_r^2(\hat{\theta}^*, \theta) = \frac{1}{4}\sum_{j=1}^{2m} \frac{F_*(u_{j,r}) - F_T(u_{j,r})}{F_T(u_{j,r})[1 - F_T(u_{j,r})]} \mathbf{1}_{(0,1)}(F_T(u_{j,r}))$$

The computed distances are then averaged over all R replications, and

$$\tilde{A}^2(\hat{\theta}^*, \theta) = \frac{\bar{A}^2(\hat{\theta}^*, \theta)}{\bar{A}^2(\hat{\theta}^{HB}, \theta)} \tag{16}$$

is reported in results.

*iv) Frequentist efficiency*. We evaluate the average impact of the adjustment on the unconditional frequentist MSE of small area predictors defined as $MSE(\hat{\theta}_i^*) = E(\hat{\theta}_i^* - \theta_i)^2$ (see Rao, 2003, section 6.2). We estimate the $MSE(\hat{\theta}_i^*)$ by means of its Monte Carlo approximation $mse_{MC}(\hat{\theta}_i^*) = \mathrm{R}^{-1}\sum_{r=1}^{R}(\hat{\theta}_{r,i}^* - \theta_{r,i})^2$. These quantities are area-specific. We just focus on the mean of their distribution across areas:

$$amse_{MC}(\hat{\theta}^*) = m^{-1}\sum_{i=1}^{m} mse_{MC}(\hat{\theta}_i^*) \tag{17}$$

*v) Frequentist evaluation of PMSE*. Frequentist properties of (9) are evaluated using the following measure of relative bias:

$$apmse(\hat{\theta}^*) = \frac{1}{m}\sum_{i=1}^{m}\frac{\mathrm{R}^{-1}\sum_{r=1}^{R} pmse_{MC}(\hat{\theta}_{r,i}^*)}{mse_{MC}(\hat{\theta}_i^*)} \tag{18}$$

where $pmse_{MC}(\hat{\theta}_{r,i}^*)$ is the $pmse_{MC}(\hat{\theta}_i^*)$ calculated using data from the *r*-th draw of the Monte Carlo exercise. Moreover note that $pmse_{MC}(\hat{\theta}_i^{HB})$ error reduces to posterior variance.

Codes are written in R (R Development Core Team, 2006). For MCMC calculations we used the Brugs package (Thomas and O'Hara, 2006) which recursively calls the MCMC dedicated software OpenBUGS (Thomas *et al.* 2006). As for technical details concerning MCMC calculations, we generate samples of size 20,000 for all chains deleting a conservative 'burn in' sample of size 5,000. In fact, the relatively simple normal models employed in the simulations have all shown very fast convergence rates. Convergence has been checked by means of standard convergence statistics (Cowles and Carlin, 1996).

## 5. SIMULATION RESULTS

For brevity, we report results for the case $m = 100$ and discuss the case $m = 30$ only when differences are remarkable or the comparison highlights important points. Results about the shrinkage correction are displayed in table 2. It is apparent that direct estimates are overdispersed, with overdispersion increasing with the average variance of direct estimators; hierarchical Bayes estimators are underdispersed in all situations, and more seriously so when the direct estimates convey little information (Population 2). More important, all adjusted estimators approximately eliminate the overshrinkage. Fluctuations of values around 1 do not seem to follow any significant pattern. The correction of overshrinkage does not seem to be influenced by the number of areas.

TABLE 2

*Overshrinkage correction as measured by the indicator* $AV(\hat{\theta}^*)$, $m = 100$

| | $\hat{\theta}^{DIR}$ | $\hat{\theta}^{HB}$ | $\hat{\theta}^{CHB}$ | $\hat{\theta}^{CLHB}$ | $\hat{\theta}^{ZHB}$ |
|---|---|---|---|---|---|
| Population 1 | 3.86 | 0.51 | 1.02 | 1.01 | 1.00 |
| Population 2 | 4.65 | 0.33 | 1.03 | 1.00 | 0.97 |
| Population 3 | 2.35 | 0.69 | 0.95 | 1.00 | 0.99 |

Table 3 shows results based on the Kolmogorov-Smirnov distance. All adjusted estimators clearly improve the performances of $\hat{\theta}^{HB}$ and the improvement is larger when $m = 100$ with respect to the case of $m = 30$. Among adjusted estimators, $\hat{\theta}^{ZHB}$ emerges clearly as the best. Note that, given the size of Monte Carlo (MC) errors, all observed differences appearing in table 3 can be taken as 95% significant. Moreover, note that the advantage of $\hat{\theta}^{ZHB}$ over the other adjusted methods is less pronounced for Population 2 (poorly informative direct estimates). This makes sense since the estimation of individual $\theta_{(i)}$ requires more information than the global adjustment on which $\hat{\theta}^{CHB}$ and $\hat{\theta}^{CLHB}$ are based. $\hat{\theta}^{CHB}$ and $\hat{\theta}^{CLHB}$ perform closely and none of the two seems preferable.

TABLE 3

*Ratio of the Kolmogorov-Smirnov distances between estimated and actual EDF averaged over MC replications divided by the same quantity calculated for the* $\hat{\theta}^{HB}$ *estimators, i.e.* $\tilde{D}(\hat{\theta}^*, \theta)$, $m = 100$

| Population | $\hat{\theta}^{DIR}$ | $\hat{\theta}^{HB}$ | $\hat{\theta}^{CHB}$ | $\hat{\theta}^{CLHB}$ | $\hat{\theta}^{ZHB}$ |
|---|---|---|---|---|---|
| Population 1 | 0.97 | 1 | 0.66 | 0.61 | 0.49 |
| Population 2 | 0.89 | 1 | 0.57 | 0.57 | 0.50 |
| Population 3 | 0.97 | 1 | 0.81 | 0.81 | 0.64 |

Table 4 presents the results related to the Anderson-Darling distance. We note that $\hat{\theta}^{HB}$ is in an intermediate position between the direct estimators (that are the worst performers over all settings) and the adjusted estimators that are better. Among adjusted estimators $\hat{\theta}^{ZHB}$ is clearly better than the other two except for Population 2, characterized by poorly informative direct estimates where, by the way, it still performs a little better. We may then conclude that the estimation method proposed by Zhang turns out to be the best with respect to both considered measures of distance and it gives its best when the number of areas is large and the direct estimates are not too imprecise.

TABLE 4

*Ratio of the Anderson-Darling distances between estimated and actual EDF averaged over MC replications divided by the same quantity calculated for the* $\hat{\theta}^{HB}$ *estimators, i.e.* $\tilde{A}^2(\hat{\theta}^*, \theta)$, $m = 100$

| Population | $\hat{\theta}^{DIR}$ | $\hat{\theta}^{HB}$ | $\hat{\theta}^{CHB}$ | $\hat{\theta}^{CLHB}$ | $\hat{\theta}^{ZHB}$ |
|---|---|---|---|---|---|
| Population 1 | 2.63 | 1 | 0.34 | 0.31 | 0.22 |
| Population 2 | 2.04 | 1 | 0.30 | 0.30 | 0.28 |
| Population 3 | 1.20 | 1 | 0.49 | 0.49 | 0.37 |

The results related to the repeated sampling efficiency, as measured by the empirical unconditional Mean Square Error are shown in Table 5, where the adjustment of $\hat{\theta}^{HB}$ has, as expected, a cost in terms of efficiency. The increase in Mean Square Errors depends on the precision of direct estimators. When the precision is high (Population 3), the rise is around 10%, but when it is low, $amse_{MC}(\hat{\theta}^{*})$ are 20% or even 30% (in the case of $\hat{\theta}^{CLHB}$) higher than in the case of $\hat{\theta}^{HB}$. Nonetheless we may note that the improvement with respect to the direct estimators remains substantial. Moreover $\hat{\theta}^{CHB}$ and $\hat{\theta}^{ZHB}$ show similar performances, while $\hat{\theta}^{CLHB}$ turns out to be a little less efficient.

TABLE 5

*Efficiency of the considered estimators as measured by* $amse_{MC}(\hat{\theta}^{*})$, $m = 100$

| | $\hat{\theta}^{DIR}$ | $\hat{\theta}^{HB}$ | $\hat{\theta}^{CHB}$ | $\hat{\theta}^{CLHB}$ | $\hat{\theta}^{ZHB}$ |
|---|---|---|---|---|---|
| Population 1 | 2.82 | 0.51 | 0.60 | 0.68 | 0.61 |
| Population 2 | 3.59 | 0.72 | 0.91 | 0.94 | 0.90 |
| Population 3 | 0.51 | 0.31 | 0.34 | 0.35 | 0.34 |

A more detailed analysis of the distribution of $amse_{MC}(\hat{\theta}^{*})$ across areas (for which we do not show tables) highlights the very different behavior of $\hat{\theta}^{CLHB}$ with respect to $\hat{\theta}^{CHB}$ and $\hat{\theta}^{ZHB}$: it performs clearly better when direct estimates are more precise than the average and far worse in the case of areas characterized by the most imprecise direct estimates. This behavior depends on the nature of the estimator. From (11) we may note that, with respect to posterior mean, $\hat{\theta}^{CHB}$ gives less weight to the synthetic component and more to the direct one. When $y_i$ is very precise, this leads to more efficient estimators than other adjusted methods; unfortunately $y_i$ receives more weight even when it is unreliable, thus producing a large loss in efficiency with respect to and the other adjusted methods.

TABLE 6

*Average ratio of the estimated* PMSE *to the* MSE *:* $apmse(\hat{\theta}^{*})$, $m = 100$

| | $\hat{\theta}^{HB}$ | $\hat{\theta}^{CHB}$ | $\hat{\theta}^{CLHB}$ | $\hat{\theta}^{ZHB}$ |
|---|---|---|---|---|
| Population 1 | 1.01 | 1.00 | 1.00 | 1.00 |
| Population 2 | 0.98 | 0.99 | 0.98 | 0.98 |
| Population 3 | 0.99 | 1.00 | 1.00 | 1.00 |

Table 6 reports an evaluation of frequentist properties of the Posterior Mean Square Error defined in (9). It is apparent that this Bayesian uncertainty measure, represents a sensible measure of variability also with respect to repeated sampling; in fact in all cases it is approximately unbiased; although it should be noted that this property holds 'on average' with respect to the set of areas being studied. The property, that was known to hold for the posterior variances as frequentist variability measures of $\hat{\theta}^{HB}$ under careful choice of the priors (Ganesh and Lahiri, 2008), is then extensible to the case of posterior mean square errors, at least for the prior distributions chosen in our simulation exercise.

## 6. PROPERTIES OF THE ADJUSTED ESTIMATORS UNDER FAILURES OF THE NORMALITY OF RANDOM EFFECTS

In this section we compare the performance of adjusted estimators already considered when the assumed normality of random effects does not hold, that is, we evaluate whether they are 'robust' with respect to this departure from the assumptions under which they are obtained.

We generate population data according to the same simulation experiment discussed in Section 4, changing the distribution assumed for the random effects. We consider: 1) $v_i \sim Laplace(0,1/\sqrt{2})$ and 2) $v_i \sim Exp(1)$. In both cases $V(v_i)=1$, so the interpretation of Population 1, Population 2, Population 3 in terms of informativeness of the direct estimators remains unchanged. Case 1) represents a mild deviation from normality while case 2) represents a more serious departure from this assumption.

We note that the performances of hierarchical Bayes estimators of the area means $\theta_i$ remain approximately unchanged when random effects are generated by a Laplace distribution but a normal model is assumed. This robustness with respect to moderate failures of the assumptions on the distribution of the random effects is noted in Sinharay and Stern (2003) and Fabrizi and Trivisano (2009) . When random effects are generated from an Exponential distribution, the deterioration of the performances of $\theta^{HB}$ is substantial, even though not catastrophic. Under the assumption on the direct estimators' variances of 'Population 1' the Kolmogorov-Smirnov distance grows by a 30%, the Anderson-Darling by a 60%, while the Mean Square Error averaged over the set of all the areas only by 10%. In Table 7 results about the Kolmogorov-Smirnov and the Anderson-Darling distance are reported for $m=100$. Results related to other indicators and $m=30$ are not reported since they are substantially similar. From Table 7 we may note that all adjusted predictors provide smaller improvements with respect to $\theta^{HB}$ than in the case of normality; this means that the estimation of the Empirical Distribution Function of the area averages or some of its features is more sensitive to wrong distributional assumptions than the estimation of individual area averages.

The performance of the adjusted predictors are now closer than under normality of random effects. $\hat{\theta}^{ZHB}$ remains slightly better in terms of Kolmogorov-Smirnov distance, while $\hat{\theta}^{CHB}$ is a little better if we consider the Anderson-Darling distance and $amse_{MC}(\hat{\theta}^*)$. Since $\hat{\theta}^{ZHB}$ turned out to be better under normality, we may conclude that its performances deteriorate more when this assumption fails with respect to other adjusted estimators. This makes sense, since $\hat{\theta}^{ZHB}$ makes an heavier use of the assumption of normality in the estimation of $\{\theta_{(i)}\}$ than the other predictors, whose corrections are based on the estimation of moments. Note, however, that the performances of $\hat{\theta}^{ZHB}$ remain always comparable to those of other adjusted predictors even under big departures from normality of random effects.

TABLE 7

*Comparison of adjusted estimators under failures of the distributional assumption on random effects,* $m = 100$

| Actual distribution of the $v_i$ | Population | $\hat{\theta}^{DIR}$ | $\hat{\theta}^{HB}$ | $\hat{\theta}^{CHB}$ | $\hat{\theta}^{CLHB}$ | $\hat{\theta}^{ZHB}$ |
|---|---|---|---|---|---|---|
| | | | $\tilde{D}(\hat{\theta}^*,\theta)$ | | | |
| Laplace | Population 1 | 1.241 | 1 | 0.799 | 0.844 | 0.763 |
| | Population 2 | 1.093 | 1 | 0.696 | 0.719 | 0.667 |
| | Population 3 | 1.254 | 1 | 1.023 | 1.022 | 0.956 |
| Exponential | Population 1 | 1.108 | 1 | 0.793 | 0.766 | 0.761 |
| | Population 2 | 1.038 | 1 | 0.720 | 0.713 | 0.703 |
| | Population 3 | 1.146 | 1 | 0.921 | 0.915 | 0.951 |
| | | | $\tilde{A}^2(\hat{\theta}^*,\theta)$ | | | |
| Laplace | Population 1 | 3.558 | 1 | 0.494 | 0.582 | 0.531 |
| | Population 2 | 2.580 | 1 | 0.414 | 0.449 | 0.436 |
| | Population 3 | 2.115 | 1 | 0.903 | 0.920 | 0.923 |
| Exponential | Population 1 | 1.490 | 1 | 0.469 | 0.431 | 0.434 |
| | Population 2 | 1.412 | 1 | 0.387 | 0.391 | 0.406 |
| | Population 3 | 0.793 | 1 | 0.653 | 0.638 | 0.661 |

## 7. CONCLUSIONS

In this paper we discuss and compare three different methods for adjusting a set of small area estimators in order to improve estimation of the EDF of the 'ensemble' of small area parameters. Two of these predictors (CHB and CLHB) are well known in the literature, while the third (ZHB) is more recent. The evaluation of the performances of the three estimators is based on a simulation exercise covering a range of situations which are potentially relevant for small area practitioners.

A first conclusion from this analysis is that all the methods considered meet the goal of correcting overshrinkage and leading to more realistic estimates of EDF of the 'ensemble' of small area means. On the other hand, adjusted estimators are less efficient than posterior means, but their gain in precision with respect to direct estimators remains substantial.

Zhang's predictor seems as better than the other two when the normality of random effects holds. We also considered failures of this assumption, generating populations with Laplace and Exponentially distributed random effects. In this latter cases Zhang's predictor and the better known CHB predictor show close performances. Thereby Zhang's predictor is more sensitive to the failure of distributional assumptions.

We studied the frequentist behavior of measures of uncertainty associated to adjusted hierarchical Bayes estimators, finding that, at least for the prior distributions chosen in the simulation exercise, the posterior MSEs have good frequentist properties and may be assumed by practictioners as acceptable measures of frequentist variability.

*Dipartimento di Scienze Economiche e Sociali* ENRICO FABRIZI
*Università Cattolica del S. Cuore, sede di Piacenza*

APPENDIX

Proof of (6). Now

$$E_M\left[\frac{1}{m-1}\sum_{i=1}^{m}(\theta_i-\overline{\theta})^2\right]=E\left[\frac{1}{m-1}\sum_{i=1}^{m}\theta_i^2-\frac{m}{m-1}\overline{\theta}^2\right]=$$

$$=\frac{m}{m-1}\sigma_v^2-\frac{1}{m-1}\sigma_v^2+\frac{1}{m-1}\sum_{i=1}^{m}\beta^t(\mathbf{x}_i-\overline{\mathbf{x}})(\mathbf{x}_i-\overline{\mathbf{x}})^t\beta$$

as $E_M(\theta_i^2)=V_M(\theta_i)+[E_M(\theta_i)]^2=\sigma_v^2+\beta^t\mathbf{x}_i\mathbf{x}_i^t\beta$

and $E_M(\overline{\theta}^2)=V_M(\overline{\theta})+[E(\overline{\theta})]^2=m^{-1}\sigma_v^2+\beta^t\overline{\mathbf{x}\mathbf{x}}^t\beta$

To prove (7) note that

$$E_M E_D\left[\frac{1}{m-1}\sum_{i=1}^{m}(y_i-\overline{y})^2\right]=\frac{1}{m-1}E_M E_D\left[\sum_{i=1}^{m}y_i^2-m\overline{y}^2\right].$$

As $E_D(y_i^2)=V_D(y_i)+[E_D(y_i)]^2=\psi+\theta_i^2$ and

$E_D(\overline{y}^2)=V_D(\overline{y})+[E_D(\overline{y})]^2=m^{-1}\psi+\overline{\theta}^2$ we have that

$$\frac{1}{m-1}E_M E_D\left[\sum_{i=1}^{m}y_i^2-m\overline{y}^2\right]=\frac{1}{m-1}E_M\left[(m-1)\psi+\sum_{i=1}^{m}\theta_i^2-m\overline{\theta}^2\right].$$

With the same argument used in the proof of (6) we obtain

$$\frac{1}{m-1}E_M\left[(m-1)\psi+\sum_{i=1}^{m}\theta_i^2-m\overline{\theta}^2\right]=\psi+\sigma_v^2+\frac{1}{m-1}\sum_{i=1}^{m}\beta^t(\mathbf{x}_i-\overline{\mathbf{x}})(\mathbf{x}_i-\overline{\mathbf{x}})^t\beta$$

Eventually, to prove (8) note that $\overline{\theta}^B=\gamma\overline{y}+(1-\gamma)\overline{\mathbf{x}}^t\beta$ and

$$\frac{1}{m-1}\sum_{i=1}^{m}(\hat{\theta}_i^B - \hat{\bar{\theta}}^B)^2 = \gamma^2\left[\frac{1}{m-1}\sum_{i=1}^{m}(y_i-\bar{y})^2\right] + (1-\gamma)^2\beta^t\Sigma_{xx}\beta$$
$$+\frac{2\gamma(1-\gamma)}{m-1}\sum_{i=1}^{m}(y_i-\bar{y})(\mathbf{x}_i-\bar{\mathbf{x}})^t$$

As a consequence

$$E_M E_D\left[\frac{1}{m-1}\sum_{i=1}^{m}(\hat{\theta}_i^B - \hat{\bar{\theta}}^B)^2\right] =$$
$$= \gamma^2\beta^t\Sigma_{xx}\beta + \gamma\sigma_v^2 + (1-\gamma)^2\beta^t\Sigma_{xx}\beta + 2\gamma(1-\gamma)\beta^t\Sigma_{xx}\beta$$
$$= \gamma\sigma_v^2 + \beta^t\Sigma_{xx}\beta$$

REFERENCES


T.W. ANDERSON AND D.A. DARLING (1954) *A test of goodness of fit*, "Journal of the American Statistical Association", 49, pp. 765-769.

M.K. COWLES, B.P. CARLIN (1996) *Markov Chain Monte Carlo convergence diagnostics: a comparative review*, "Journal of the American Statistical Association", 91, pp. 833-904.

E. FABRIZI, M.R. FERRANTE AND S. PACEI (2005) *Estimation of poverty indicators at sub-national level using multivariate small area models.* "Statistics in Transition", 7, pp. 587-608.

E. FABRIZI AND C. TRIVISANO (2009) *Robust Linear Mixed Models for Small Area Estimation,* "Journal of Statistical Planning and Inference", 140, 433-443.

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

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

A. GELMAN (2006) *Prior distribution for variance parameters in hierarchical models*, "Bayesian Analysis", 1, pp. 515-533.

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

M. GHOSH AND T. MAITI (1999) *Adjusted Bayes estimators with applications to Small Area Estimation*, "Sankhya", Ser. B, 61, pp. 71-90.

H. GOLDSTEIN (1975) *A Note on some Bayesian Nonparametric Estimates*, "Annals of Statistics", 3, pp. 736-740.

P. HEADY AND M. RALPHS (2004) *Some findings of the EURAREA project – and their implications for statistical policy*, "Statistics in Transition", 6, pp. 641-653.

D.R. JUDKINS AND J. LIU (2000) *Correcting the Bias in the Range of a Statistic across Small Areas*, "Journal of Official Statistics", 16, pp. 1-13.

P. LAHIRI, J. JIANG (2006) *Mixed model prediction and small area estimation*, "Test", 15, pp. 1-96.

T.A. LOUIS (1984) *Estimating a Population of Parameters Values Using Bayes and Empirical Bayes Methods*, "Journal of the American Statistical Association", 79, pp. 393-398.

J.N.K. RAO (2003) *Small Area Estimation*, John Wiley and Sons, New York.

R DEVELOPMENT CORE TEAM (2006) *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL downloadable at http://www.R-project.org.

S. SINHARAY, H.S. STERN (2003) *Posterior predictive model checking in hierarchical models*, "Journal of Statistical Planning and Inference", 111, pp. 209-221.

E. SPJØTVOLL, I. THOMSEN (1987) *Application of some empirical Bayes methods to Small Area Estimation*, "Bullettin of the International Statistical Institute", 2, pp. 435-449.

A. THOMAS, B. O'HARA (2006) *The BRugs package, Software and Documentation*, downloadable at http://cran.r-project.org/org/packages/BRugs.

A. THOMAS, B. O HARA, U. LIGGES, S. STURZ (2006) *Making BUGS Open*, "R News", 6, pp. 12-17.

L.C. ZHANG (2003) *Simultaneous estimation of the mean of a binary variable from a large number of small areas*, "Journal of Official Statistics", 19, pp. 253-263.

SUMMARY

*A comparison of adjusted Bayes Estimators of an ensemble of small area parameters*

With "ensemble properties" of small area estimators, we mean their ability to reproduce the Empirical Distribution Function (EDF) characterizing the collection of underlying small area parameters (means, totals). Good "ensemble properties" may be relevant when estimation of non-linear functionals of the EDF of small area parameters (such as their variance) is needed. Small area estimators associated to the popular Fay-Herriot model are considered. "Bayes estimators", i.e. posterior means, do not enjoy of good ensemble properties. In this paper three different adjusted predictors are compared, by means of a simulation exercise, under the assumption of correctly specified model. As the distributional assumptions on the random effects are difficult to assess, the considered predictors are compared also with respect to their robustness to the presence of failures in the distributional assumptions on the random effects.