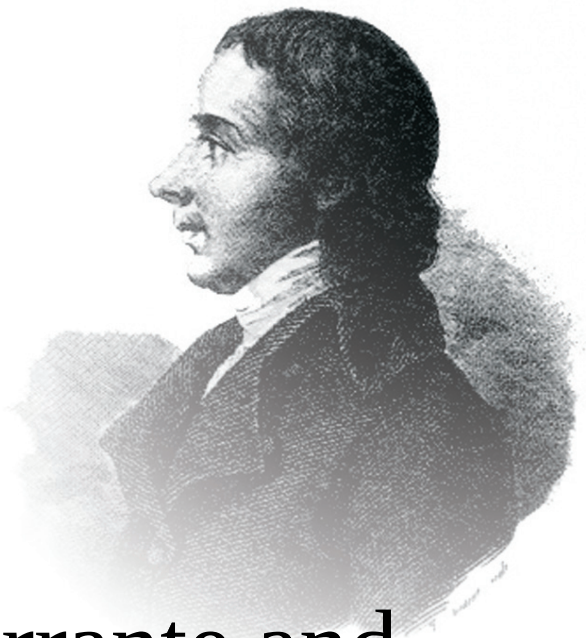
Rapporto n. 1 2007

dmsia unibg.it



E. Fabrizi, M.R. Ferrante and S. Pacei

A Comparison of Adjusted Bayes Estimators of use in Small Area Estimation

Serie Ricerca

Dipartimento
di Matematica, Statistica, Informatica e Applicazioni
“Lorenzo Mascheroni”

UNIVERSITÀ DEGLI STUDI DI BERGAMO

A Comparison of Adjusted Bayes Estimators of use in Small Area Estimation

Enrico Fabrizi
DMSIA, Università di Bergamo

Maria Rosaria Ferrante
Dipartimento di Scienze Statistiche 'P. Fortunati',
Università di Bologna

Silvia Pacei
Dipartimento di Scienze Statistiche 'P. Fortunati',
Università di Bologna

Abstract

Empirical and Hierarchical Bayes methods are often used to improve the precision design-based estimators in Small Area estimation problems. By the way, when posterior means are used to estimate an 'ensemble' of parameters, a poor estimate of the empirical distribution function of the ensemble typically results. Several adjusted estimators have been proposed in the literature in order to obtain better estimates of nonlinear function of an ensemble of parameters. In this paper we discuss a set of adjusted estimators with reference to the univariate and multivariate Fay-Herriot models within the framework of Hierarchical Bayesian modeling. The repeated sampling properties of the considered estimators and the associated measures of uncertainty are evaluated by means of a simulation exercise.

1 Introduction

In recent years sample surveys have been characterized by a growing demand for estimates of population descriptive quantities of domains (or 'areas') obtained cross-classifying the target population according to multiple criteria. As the sample portion pertaining to domains is often too small to produce reliable estimates using standard design-based estimators, Small Area methods have become a relevant research topic (see Ghosh and Rao, 1994, Mukhopadhyay, 1998 or Rao, 2003 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 e 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 means, often referred to as 'Bayes estimators' (Ghosh, 1992).
Bayes estimators may be very effective in improving the precision 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, Bayesian estimators may be proven to be poor estimators of the actual distribution function of a population of Small Area parameters (Louis, 1984, Heady and Ralphs, 2004). The interest in the distribution function may be crucial when Small Areas estimates are used, for instance, in the analysis of regional disparities (Fabrizi *et al.*, 2005). These studies are particularly relevant for applied economists and policy makers in the European Union as the reduction of the territorial disparities in the distribution of income and the promotion of an homogeneous economical development have become a priority for the European Union (European Commission, 2004).
In this paper we discuss the popular Fay-Herriot model (Fay and Herriot, 1979) and a set of alternative adjusted estimators associated to it. With adjusted estimators we mean estimators of the Small area parameters that enjoy also 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. Although we consider Bayesian methods to obtain estimators, we focus on their frequentist properties since these are usually relevant for practitioners. Secondly, 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 measure of variability.
A third goal is that of extend the reviewed adjusted estimators to the multivariate Fay-Herriot model (Rao, 2003, section 5.4) and to apply a parallel comparison exercise to the multivariate estimators. Multivariate models are of practical relevance as they allow to exploit the correlations between various target variables in the population to enhance the estimation of Small Area parameters. An example of multivariate models applied to poverty-related parameters is provided by Ghosh, Nangia and Kim, (1996).
The paper is organized as follows. In section 2, we shortly discuss Bayes estimators associated to the Fay-Herriot model and their failure 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 recently proposed by Zhang (2003). These estimators are reviewed in section 3. In section 4 multivariate extensions are discussed. The simulation exercise and the tools used in comparisons are introduced in section 5. Although all simulations use popula-

tions generated under normality, we focus on the accuracy of EDF 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. Results of the simulations are discussed in section 6, while section 7 offers some concluding remark.

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 $\{y_i\} \quad 1 \leq i \leq m$ 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 $E(e_i v_i) = 0$.
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) and (3) imply that $E_M(y_i|\theta_i) = E_D(y_i) = \theta_i$ and $V_M(y_i|\theta_i) = V_D(y_i) = \psi_i$, that is the first two moments of y_i according to the model (conditional on θ_i) and randomization distribution are the same. To be consistent in notation let's also write $E_M(\theta_i) = \mathbf{x}_i^t\beta$ and $V_M(\theta_i) = \sigma_v^2$.
To keep things simple let's assume that $\psi_i = \psi$ and, for the moment, that $\mathbf{x}_i^t\beta = \mu$. Moreover let's assume β and σ_v^2 are known.
For the considered simplified model, the Bayes estimator (posterior mean) of θ_i is given by $\hat{\theta}_i^B = \gamma y_i + (1-\gamma)\mu$ where $\gamma = \sigma_v^2(\sigma_v^2 + \psi)^{-1}$. It is easy to show that

$$E_M E_D\left[\frac{1}{m-1}\sum_{i=1}^{m}(y_i - \bar{y})^2\right] = \psi + \sigma_v^2 = \frac{\sigma_v^2}{\gamma} \tag{5}$$

As $\gamma \in [0,1]$, we have that direct estimates $\{y_i\}$ are overdispersed with respect to the $\{\theta_i\}$ whose expected variance is σ_v^2. the amount of overdispersion grows with ψ: the more $\{y_i\}$ are imprecise estimate the more they are overdispersed. On the other hand we have that

$$E_M E_D\left[\frac{1}{m-1}\sum_{i=1}^{m}(\hat{\theta}_i^B - \hat{\bar{\theta}}^B)^2\right] = \sigma_v^2\gamma \tag{6}$$

The $\{\theta_i\}$ are therefore underdispersed, with underdispersion proportional to the amount of shrinkage, that is to the weight given to the synthetic component μ in the Bayes estimator $\hat{\theta}_i^B$. We may also observe that normality of (3) and (4) is not needed to prove (6) as shown in Ghosh (1992); assuming quadratic loss to summarize the posterior distribution and $\sum_{i=1}^m(\hat{\theta}_i^B - \hat{\bar{\theta}}^B)^2 > 0$ are sufficient conditions for the $\hat{\theta}_i^B$ to be underdispersed.
The underdispersion of Bayes estimators is caused by shrinkage toward a common mean μ. If we relax the assumption $\mathbf{x}_i^t\beta = \mu$ we may have either over- or under- dispersion. To see this, let's note first that $\hat{\theta}_i^B = \gamma y_i + (1-\gamma)\mathbf{x}_i^t\beta$ and that (5) continues to be true. By the way, the equivalent of (6) is a little more complicated:

$$E_M E_D\left[\frac{1}{m-1}\sum_{i=1}^m \left(\hat{\theta}_i^B - \hat{\bar{\theta}}^B\right)^2\right] = \gamma\sigma_v^2 + (1-\gamma)^2\beta^t\Sigma_{xx}\beta + 2\gamma(1-\gamma)\Sigma_{x\theta}\beta \quad (7)$$

with $\Sigma_{xx} = (m-1)^{-1}\sum_{i=1}^m(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^t$, $\bar{\mathbf{x}} = \sum_{i=1}^m \mathbf{x}_i$, $\Sigma_{xx} = (m-1)^{-1}\sum_{i=1}^m(\mathbf{x}_i - \bar{\mathbf{x}})(\hat{\theta}_i^B - \hat{\bar{\theta}}^B)$.
Expression in the right hand of (7) may be less or more than σ_v^2. In practical situations it will be only occasionally approximately equal to σ_v^2, so the variance of the population of small area parameters will be under or overstated to some extent (see also Heady and Ralphs, 2004). The case with unequal sampling variances and unknown β and σ_v^2 is more complicated to deal with analytically but leads to the very same conclusions.

3 Univariate Adjusted Bayes Estimators

We consider three different adjusted Bayes estimators associated to the Fay-Herriot model (1) - (4): i) the constrained Bayes approach, ii) the Spjøtvoll and Thomsen method (constrained Linear Bayes approach), iii) the 'ensemble estimation' method proposed by Zhang (2003). These estimators will be reviewed within a Hierarchical Bayes framework; that is we do not assume the hyperparameters β,σ_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 $\hat{\theta}_i^\star$ represent summary of the posterior distribution $p(\theta_i|\mathbf{z})$ different from the posterior mean $\theta_i^{HB} = E(\theta_i|\mathbf{z})$ and are therefore suboptimal with respect to quadratic loss. Uncertainty associated to this alternative posterior summaries may be measured by the posterior Mean Square Error:

$$PMSE\big(\hat{\theta}_i^\star\big) = V(\theta_i|\mathbf{z}) + \big(\hat{\theta}_i^\star - \theta_i^{HB}\big)^2 \quad (8)$$

We denote $E()$ the expectation with respect to the model distribution omitting the deponent M, since randomization moments are not involved anymore.
Of course, $PMSE\big(\hat{\theta}_i^\star\big) > V(\theta_i|\mathbf{z})$. This highlights how 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) for arbitrary distributions. To illustrate these estimators, let's note that the aim is that of obtaining a set of estimators $\{t_i\}$, $1 \leq i \leq m$, optimal under quadratic loss and satisfying the following constraints:

1. $\bar{t}_i = \hat{\bar{\theta}}_{HB}$
2. $(m-1)^{-1}\sum_{i=1}^{m}\big(t_i - \bar{t}\big)^2 = E\Big[\sum_{i=1}^{m}(\theta_i - \bar{\theta})|\mathbf{z}\Big]$

The Constrained Hierarchical Bayes (CHB) estimators are given by:

$$\hat{\theta}_i^{CHB} = \hat{\bar{\theta}}_{HB} + a(\mathbf{z})\big(\hat{\theta}_i^{HB} - \hat{\bar{\theta}}_{HB}\big) \tag{9}$$

where

$$a(\mathbf{z}) = \left[1 + \frac{(\sum_{i=1}^{m} V(\theta_i - \bar{\theta})|\mathbf{z})}{(m-1)^{-1}\sum_{i=1}^{m}\big(\hat{\theta}_i^{HB} - \hat{\bar{\theta}}^{HB}\big)^2}\right]^{\frac{1}{2}} \tag{10}$$

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

3.2 The Constrained Hierarchical Linear Bayes Estimator

Let's suppose, for the moment, that the hyperparameters β, σ_v^2 are known. The Constrained linear Bayes estimator of θ_i is a summary of the posterior distribution of the form $\hat{\theta}_i^L = a_i y_i + b_i$ satisfying the constraints: i) $E(\hat{\theta}_i^L) = \mathbf{x}_i^t\beta$, ii) $E(\hat{\theta}_i^L - \mathbf{x}_i^t\beta)^2 = \sigma_v^2$. Note that when the posterior mean is in linear form we say the distribution enjoys posterior linearity (Goldstein, 1975), a conditions that holds for a variety of distributions. 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

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

where $E_{(\beta,\sigma_v^2|\mathbf{z})}$ is the expectation taken with respect to the posterior distribution of β,σ_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 of Zhang

Given the set $\{\theta_i\}$, $1 \leq i \leq m$ of the area parameters of interest, let $\{\theta_{(i)}\}$ be the associated set ordered set $(\theta_{(1)} \leq \theta_{(2)} \leq ... \leq \theta_{(m)})$. Then $\eta_i = E(\theta_{(i)}|\mathbf{z})$ is the best predictor of $\theta_{(i)}$ under quadratic loss and $\{\eta_i\}$ is the best 'ensemble' estimator of $\{\theta_i\}$ in the same sense. The set of estimators $\{\eta_i\}$ is not area specific in the sense that its elements are not associated to specific areas. To match the $\{\eta_i\}$ with the small areas Zhang (2003) proposes, in the context of an Empirical Bayes estimation approach, to estimate the ranks of $\{\theta_i\}$ using those of the $\{E(\theta_i|\mathbf{z})\}$. By the way, the ranks of the posterior means may be poor estimators of actual ranks, especially if there is much variability in the posterior variances. Following Ghosh and Maiti (1999) and differently from Zhang (2003), we propose $\hat{r}_i = E\big(rank(\theta_i|\mathbf{z})\big)$, $1 \leq i \leq m$, the posterior expectation of ranks, as the estimator needed 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 in detail, we can rank the $\theta_i(s)|\mathbf{z}$ from any draw s of the Markov Chain after convergence. Then we can approximate $\hat{r}_i$ averaging the ranks $rank(\theta_i(s)|\mathbf{z})$ over all draws, obtaining $\hat{r}_i^{MC} = S^{-1}\sum_{s=1}^{S} rank(\theta_i(s)|\mathbf{z})$ 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 Multivariate Adjusted Bayes Estimators

A multivariate extension of the Fay-Herriot model (1) - (4) may be described as follows:

$$\mathbf{y}_i = \boldsymbol{\theta}_i + \mathbf{e}_i \tag{14}$$

$$\boldsymbol{\theta}_i = \mathbf{X}_i\beta + \mathbf{v}_i \tag{15}$$

$$\mathbf{e}_i \overset{ind}{\sim} N(\mathbf{0}, \Psi_i) \tag{16}$$

$$\mathbf{v}_i \overset{ind}{\sim} N(\mathbf{0}, \Sigma_v) \tag{17}$$

$1 \leq i \leq m$. The $p-$vector $\mathbf{y}_i$ contains a set of design-unbiased estimators with mean $\boldsymbol{\theta}_i$ and (design-based) variance-covariance matrix Ψ_i. All $p \times p$ matrices Ψ_i are assumed to be positive definite matrices of known constants. The means

of the area parameters $\boldsymbol{\theta}_i$ are functions of a $p \times k$ matrix of auxiliary variables $\mathbf{X}_i$; their common variance-covariance matrix Σ_v assumed to be positive definite. Moreover it is assumed that $E(\mathbf{e}_i^t \mathbf{v}_i) = 0$. To complete the Bayesian specification of the model a prior for β and Σ_v should be specified.
The Bayes estimators associated to multivariate Fay-Herriot model face the same failure as estimators of the distribution of the population of area parameters illustrated for the univariate case. In the following sections we discuss the extension of methods introduced in section 3 to the multivariate case.
The posterior Mean Square Error introduced in (8) may also be easily generalized to the multivariate case. If $\hat{\boldsymbol{\theta}}_i^{\star}$ is the generic multivariate adjusted estimator its posterior MSE is given by:

$$PMSE(\hat{\boldsymbol{\theta}}_i^{\star}) = V(\boldsymbol{\theta}_i|\mathbf{Z}) + \Big[(\hat{\boldsymbol{\theta}}_i^{\star} - \hat{\boldsymbol{\theta}}_i^{HB})(\hat{\boldsymbol{\theta}}_i^{\star} - \hat{\boldsymbol{\theta}}_i^{HB})|\mathbf{Z}\Big] \tag{18}$$

where $\mathbf{Z} = \{\mathbf{y}_i, \mathbf{x}_i^t, \Psi_i\}_{1 \leq i \leq m}$ denotes the data available in the multivariate problem and $\hat{\boldsymbol{\theta}}_i^{HB} = E(\boldsymbol{\theta}_i|\mathbf{Z})$.

4.1 Multivariate Constrained Hierarchical Bayes estimators

The extension of Constrained Bayes estimators to problems in which a $p-$vector of area parameters is to be estimated is worked out by Ghosh and Maiti (1999). The problem is that of finding a set of estimators $\{\mathbf{t}_i\}$ minimizing $E\big[\sum_{i=1}^m (\boldsymbol{\theta}_i - \mathbf{t}_i)(\boldsymbol{\theta}_i - \mathbf{t}_i)^t|\mathbf{Z}\big]$ under the constraints:

1. $E(\bar{\boldsymbol{\theta}}|\mathbf{Z}) = m^{-1} \sum \mathbf{t}_i$
2. $E\big[\sum_{i=1}^m (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}})^t|\mathbf{Z}\big]$

The multivariate Constrained Hierarchical Bayes estimator is given by:

$$\hat{\boldsymbol{\theta}}_i^{CHB} = \hat{\bar{\boldsymbol{\theta}}}^{CHB} + (\mathbf{H}_1 + \mathbf{H}_2)^{1/2}\mathbf{H}_2^{-1/2}(\hat{\boldsymbol{\theta}}_i^{HB} - \hat{\bar{\boldsymbol{\theta}}}^{HB}) \tag{19}$$

where $\mathbf{H}_1 = \sum_{i=1}^m V(\boldsymbol{\theta}_i|\mathbf{Z}) - mV(\bar{\boldsymbol{\theta}}|\mathbf{Z})$ and $\mathbf{H}_2 = \sum_{i=1}^m (\hat{\boldsymbol{\theta}}_i^{HB} - \mathbf{t}_i)(\hat{\boldsymbol{\theta}}_i^{HB} - \mathbf{t}_i)^t\big]$. The square roots of $\mathbf{H}_1$ and $\mathbf{H}_2$ are well defined since the two matrices are semi-positive definite.

4.2 Multivariate Constrained Hierarchical Linear Bayes estimators

We do not know of a multivariate generalization of the Spjøvoll and Thomsen estimator. By the way, it is straightforward to obtain. In parallel with section 3.2, let's assume temporarily that β and Σ_v are known. We look for an estimator in the form $\hat{\boldsymbol{\theta}}_i^L = \mathbf{A}_i \mathbf{y}_i + \mathbf{b}_i$ where $\mathbf{A}_i$ is a $p \times p$ matrix and $\mathbf{b}_i$ a $p \times 1$ vector of constants, to be determined satisfying the constraints:

1. $E(\hat{\boldsymbol{\theta}}_i^L) = \mathbf{X}_i\beta$

2. $E\big(\hat{\boldsymbol{\theta}}_i^L - \mathbf{X}_i\beta\big)\big(\hat{\boldsymbol{\theta}}_i^L - \mathbf{X}_i\beta\big)^t = \Sigma_v$

It may be easily shown that $\mathbf{b}_i = \big(\mathbf{I}_p - \mathbf{A}_i\big)\mathbf{X}_i\beta$ and $\mathbf{A}_i = \Sigma_v^{1/2}\big(\Sigma_v + \Psi_i\big)^{-1/2}$ since $E\big(\hat{\boldsymbol{\theta}}_i^L - \mathbf{X}_i\beta\big)\big(\hat{\boldsymbol{\theta}}_i^L - \mathbf{X}_i\beta\big)^t = \mathbf{A}_i\big(\Sigma_v + \Psi_i\big)\mathbf{A}_i^t$. As a consequence, the multivariate Constrained Linear Bayes estimator is given by:

$$\hat{\boldsymbol{\theta}}_i^{CLB} = \Gamma_i^{1/2}\mathbf{y}_i + \big(\mathbf{I}_p - \Gamma_i\big)^{1/2}\mathbf{X}_i\beta \tag{20}$$

where $\Gamma_i = \Sigma_v\big(\Sigma_v + \Psi_i\big)^{-1}$ (the estimator is written as function of Γ_i to emphasize the parallel with (11). As (20) is conditional on β, Σ_v the Hierarchical Linear Bayes estimator of $\boldsymbol{\theta}_i$ is given by:

$$\hat{\boldsymbol{\theta}}_i^{CHLB} = E_{(\beta,\Sigma_v|\mathbf{Z})}\big(\hat{\boldsymbol{\theta}}_i^{CLB}\big) \tag{21}$$

where $E_{(\beta,\Sigma_v|\mathbf{Z})}$ is the expectation taken with respect to posterior distribution of β, Σ_v. The estimator $\hat{\boldsymbol{\theta}}_i^{CHLB}$ can be easily be computed on the basis of the output of MCMC algorithms, while an explicit formula will in general be dependent on the chosen prior distributions and generally complicated to obtain.

4.3 A multivariate extension of simultaneous estimation method proposed by Zhang

The extension of Zhang's method to the estimation of multiple area parameters we consider in this paper is rather simple. Let $\boldsymbol{\theta}_h$, $1 \leq i \leq m$ be the set of of estimates for the h-th parameter over all areas.
After the posterior distribution of $\boldsymbol{\theta} = \{\boldsymbol{\theta}_i\}$ have been obtained using the multivariate Fay-Herriot model, the univariate Zhang's method is applied separately for each $\boldsymbol{\theta}_h$, thus obtaining $\eta_{hi} = E(\theta_{hi}|\mathbf{Z})$. The same procedure illustrated in section 3.3 is then used to estimate ranks and match the ensemble estimators to the small areas. As a result the Hierarchical Zhang estimator is given by:

$$\hat{\boldsymbol{\theta}}_i^{ZHB} = \{\eta_{\hat{r}_{hi}}\}_{1\leq h\leq p} \tag{22}$$

with $\hat{r}_{hi}$ approximated by $\hat{r}_{hi}^{MC}$ when computations are carried out using MCMC methods.

5 The simulation experiment

In this section we discuss two different, although parallel, simulation exercises for the comparison of univariate and multivariate ensemble estimators introduced in previous sections. In both exercises 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 visual inspection of chains and standard convergence statistics (Cowles and Carlin, 1996).

5.1 The univariate simulation experiment

The aim of this simulation exercise is to 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 $\hat{\theta}^{ZHB} = \{\hat{\theta}_i^{ZHB}\}$ both in terms of their ability to estimate the EDF of the $\{\theta_i\}$ and their efficiency as measured by the empirical Mean Square Errors over Monte Carlo replications. The simulation is also aimed at assessing whether posterior MSEs have good frequentist properties. The simulation is based on R=1,000 Monte Carlo 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) - (4) setting $\mathbf{x}_i^t\beta = \mu = 0$. With this simplification we may predict that the effect of Bayes estimation will be that of overshrinkage thus making interpretation of results easier.
We consider both the case of moderate and large number of areas setting $m = 30, 100$. Larger values of m are not considered because of computation burden. We set $\sigma_v^2 = 1$ and consider three different configurations of design variances. They are set in the following way: we divide the set of areas in five groups. Variances vary across groups but are constant within them. The considered configurations are illustrated in Table 1. They differ in terms of in-

Table 1: *Different configurations design variances considered in univariate simulation*

	$\psi_{1,\ldots,\frac{m}{5}}$	$\psi_{\frac{m}{5},\ldots,2\frac{m}{5}}$	$\psi_{2\frac{m}{5},\ldots,3\frac{m}{5}}$	$\psi_{3\frac{m}{5},\ldots,4\frac{m}{5}}$	$\psi_{4\frac{m}{5},\ldots,m}$
Population 1	0.1	0.33	1	3	10
Population 2	1	1.33	2	4	10
Population 3	0.1	0.25	0.5	0.75	1

formativeness of direct estimators, that may be measured by $\gamma_i = \sigma_v^2(\sigma_v^2+\psi_i)^{-1}$. Population 1 describes a situation where direct estimators show a wide range of informativeness ($\gamma \in [0.11, 0.91]$); Population 2 a situation in which direct estimators are poorly informative ($\gamma \in [0.11, 0.5]$), while in Population 3 we study the case or rather strongly informative direct estimators ($\gamma \in [0.5, 0.91]$).

We note that considering equal sampling variances, i.e. $\psi_i = \psi$ does not make much sense, first because this is seldom the case in practice and secondly because it may be proved that, for $\mathbf{x}_i^t\beta = \mu$ and μ, σ_v^2 known, $\hat{\theta}_i^{CHB} \longrightarrow \hat{\theta}_i^{CHLB}$ for $m \to \infty$ (Rao, 2003, section 9.6). As a consequence even if do not assume μ and σ_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 μ is unknown. As regards the prior distributions we assume $p(\mu, \sigma_v^2) = p(\mu)p(\sigma_v^2)$ 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 warrant 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 the various estimators, we consider the indicators described below.

i) Overshrinkage correction. Let's define

$$AV(\hat{\theta}^\star) = R^{-1}\sum_{r=1}^{R} v^2(\hat{\theta}_r^\star) \tag{23}$$

where $v^2(\hat{\theta}_r^\star) = \sum_{i=1}^m (m-1)^{-1}\big(\hat{\theta}_{r,i}^\star - \bar{\hat{\theta}}_r^\star\big)^2$, $\star = \{DIR, HB, CHB, CLHB, ZHB\}$. We expect this indicator to be larger than 1 for $\hat{\theta}^{DIR}$, less 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 estimator and that of the $\{\theta_i\}$ is calculated as $D_r(\hat{\theta}^\star, \theta) = \max_x |EDF_{\hat{\theta}^\star}(x) - EDF_\theta(x)|$ where $EDF_{\hat{\theta}^\star}(x) = m^{-1}[\#(\hat{\theta}^\star(x) \leq x)]$, $x \in \mathbb{R}$. The distances calculated at each iteration are then averaged over MC replications. For the ease of comparison we report $\tilde{D}(\hat{\theta}^\star, \theta) = \bar{D}(\hat{\theta}^\star, \theta)/\bar{D}(\hat{\theta}^{HB}, \theta)$, 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. With respect to the Kolmogorov-Smirnov distance is known to be more influenced by the discrepancies in the tails of the distribution. For our purposes the 'empirical' Anderson-Darling distance is given by $A_m^2(\hat{\theta}^\star, \theta) = -\sum_{i=1}^m (2i-1)[\log(EDF_\theta(\hat{\theta}_i^\star)) - \log(1 - EDF_\theta(\hat{\theta}_i^\star))]$. It may happen that $EDF_\theta(\hat{\theta}_i^\star))$ takes value 0 or 1 for some $\hat{\theta}_i^\star$. To avoid $A_m^2(\hat{\theta}^\star, \theta)$ to go to ∞ we use the fact that the $\{\theta_i\}$ are generated from $N(0, \sigma_v^2)$ and replace $EDF_\theta(\hat{\theta}_i^\star)$ with $N_{0,\sigma_v^2}(\hat{\theta}_i^\star)$ in these cases. $A_m^2(\hat{\theta}^\star, \theta)$ is computed for each MC repetition, the average is taken over all R replications, and $\tilde{A}_m^2(\hat{\theta}^\star, \theta) = \bar{A}_m^2(\hat{\theta}^\star, \theta)/\bar{A}_m^2(\hat{\theta}^{HB}, \theta)$ is reported in results.

iv) Efficiency. Adjusted estimators are sub-optimal by construction. We want to evaluate the average impact of the adjustment on the unconditional frequentist MSE of small area predictors defined as $MSE(\hat{\theta}_i^\star) = V\big(\hat{\theta}_i^\star - \theta_i\big)$ (see Rao, 2003, section 6.2). We estimate these $MSE(\hat{\theta}_i^\star)$ by means of their Monte Carlo approximations $mse_{MC}(\hat{\theta}_i^\star) = R^{-1}\sum_{r=1}^m (\hat{\theta}_{r,i}^\star - \theta_{r,i})^2$. These quantities vary across areas. In particular we focus on the mean of their distribution across

areas: $amse_{MC}(\hat{\theta}^{\star}) = m^{-1}\sum_{i=1}^{m} mse_{MC}(\hat{\theta}_i^{\star})$.
v) *Properties of Posterior MSE.* Frequentist properties of (8) are evaluated using the following measure of relative bias:

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

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

5.2 The multivariate simulation experiment

The simulation exercise for the comparison of multivariate adjusted estimators is similar under many respect to that introduced for the univariate case. The main goal of the simulations is to compare the direct $\hat{\boldsymbol{\theta}}^{DIR} = \{\mathbf{y}_i\}$, the hierarchical Bayes $\hat{\boldsymbol{\theta}}^{HB} = \{\hat{\boldsymbol{\theta}}_i^{HB}\}$ and the various adjusted estimators $\hat{\boldsymbol{\theta}}^{CHB} = \{\hat{\boldsymbol{\theta}}_i^{CHB}\}$, $\hat{\boldsymbol{\theta}}^{CHLB} = \{\hat{\boldsymbol{\theta}}_i^{CHLB}\}$ and $\hat{\boldsymbol{\theta}}^{ZHB} = \{\hat{\boldsymbol{\theta}}_i^{ZHB}\}$ with respect to the estimation of the EDF of the population of Small Area parameters and in terms of efficiency. For simplicity, we will consider a bivariate problem, that is one in which we have two parameters to be estimated for every Small Area. We will consider how well the various competitors estimate the two univariate EDF but we will not consider the joint EDF. Nonetheless, we will pay attention to the ability of the adjusted estimators to reproduce the actual correlation between the components of $\boldsymbol{\theta} = \{\boldsymbol{\theta}_i\}$, $1 \leq i \leq m$. Frequentist properties of $PMSE(\hat{\boldsymbol{\theta}}_i^{\star})$ will also be considered.
Data are generated according to the multivariate Fay-Herriot model (14) - (17) in which we set $p = 2$ and $\mathbf{X}_i\beta = \boldsymbol{\mu}$ on the basis of an argument parallel to that of section 5.1. In particular $\boldsymbol{\mu} = (-1, 1)$.
The experimental factors are, in the case of this second simulation exercise, given by *i*) the number of areas, *ii*) the assumed known sampling variances associated to direct estimates, *iii*) the sampling correlation between direct estimates (that is the off diagonal elements of Ψ_i, iv) The assumed constant correlation between the two components of $\boldsymbol{\theta}$: $\rho = Corr(\theta_{hi}, \theta_{h',i})$.
To avoid an excessively complicated experimental design we make the choices described below. The number of areas m is set to 30 and 100 as in section 5.1. The diagonal elements of the matrixes $(\psi_{hh,i})$ are assigned following a strategy similar to that of section 5.1: we divide, for each component, the set of areas in five groups, each characterized by a different value of $\psi_{hh,i}$. The configurations considered are listed in Table 2. We note that the ratio of sampling variances associated to different components for the same area is constant, an assumption consistent with the fact that in applications we have usually the same sample size to estimate all the parameters of an area. Direct estimates associated to the first component are rather precise, while those associated to the second component are three times larger thus simulating a situation the two study variables show

Table 2: *Different configurations design variances considered in multivariate simulation experiment*

	$\psi_{1,\ldots,\frac{m}{5}}$	$\psi_{\frac{m}{5},\ldots,2\frac{m}{5}}$	$\psi_{2\frac{m}{5},\ldots,3\frac{m}{5}}$	$\psi_{3\frac{m}{5},\ldots,4\frac{m}{5}}$	$\psi_{4\frac{m}{5},\ldots,m}$
θ_{1i}	0.2	0.4	0.6	0.8	1
θ_{2i}	0.6	1.2	1.8	2.4	3

different variability in the population. Off diagonal elements $\psi_{hh',i}$ are set in order to have $\rho_y = Corr(y_{hi}, y_{h'i}) = 0.5$. We choose this rather high correlation level because the case of correlated direct estimators is more interesting for the multivariate estimators; moreover some non-negligible correlation is realistic since the different components of $\boldsymbol{\theta}$ are estimated using the same sample data. We note that, similarly to the univariate case, if we choose equal sampling variances ($\Psi_i = \Psi$), $\hat{\boldsymbol{\theta}}_i^{CHB}$ and $\hat{\boldsymbol{\theta}}_i^{CHLB}$ are approximately identical since it can be shown that in this situation

$$\lim_{m\to\infty} \hat{\boldsymbol{\theta}}_i^{CB} = \hat{\boldsymbol{\theta}}_i^{CLB} \tag{24}$$

where $\hat{\boldsymbol{\theta}}_i^{CB}$ is the estimator (19) for Σ_v and β known and $\hat{\boldsymbol{\theta}}_i^{CLB}$ is defined in (20). A proof of (24) may be found in the Appendix.

As regards the variance-covariance matrix we set the two variances on the main diagonal equal to 1 ($\sigma^2_{v1} = \sigma^2_{v2} = 1$) while three different values are considered for ρ ($\rho = 0, 0.25, 0.75$) corresponding to independence, mild and strong correlation between the components of $\boldsymbol{\theta}$.

When modeling, it is assumed that $\boldsymbol{\mu}$ is unknown, and is given a diffuse Normal prior $\boldsymbol{\mu} \sim N(\mathbf{0}, K\mathbf{I_2})$ where $K = 100$ is 'large' with respect to the scale of the data. Other hyperparameters are assumed a priori independent and given the following priors $p(\sigma_{v1})p(\sigma_{v2}) = Unif(0, L)$, $L = 20$ and $p(\rho) = Unif(-1, 1)$.

For comparison purposes we consider a set of indicators parallel to those introduced in section 5.1: *i*) the average variance across areas defined in (23), calculated separately for $\{\theta^\star_{h1}\}$ and $\{\theta^\star_{h2}\}$, $\star = \{DIR, HB, CHB, CLHB, ZHB\}$; *ii*) the indicators based on the Kolmogorov-Smirnov and the Anderson-Darling distances averaged over MC replications, $\tilde{D}(\hat{\theta}^\star_h, \theta_h)$ and $\tilde{A}^2_m(\hat{\theta}^\star_h, \theta_h)$, $h = 1, 2$; *iii*) $amse_{MC}(\hat{\theta}^\star_h)$ is used to compare the efficiency of estimators; *iv*) $acorr(\hat{\boldsymbol{\theta}}^\star) = R^{-1}\sum_{i=1}^R Corr\big(\hat{\theta}^\star_1(r), \hat{\theta}^\star_2(r)\big)$ is introduced to see whether the set of estimates generated from various estimators reproduce the actual correlations between the components of $\boldsymbol{\theta}$.

Table 3: *Variance across areas averaged over MC replications ($AV(\hat{\theta}^{\star})$) univariate simulation*

m	Population	θ^{DIR}	θ^{HB}	θ^{CHB}	θ^{CLHB}	θ^{ZHB}
100	1	3.87	0.54	1.11	1.05	1.04
100	2	4.67	0.38	1.14	0.98	0.97
100	3	1.55	0.72	1.04	1.02	1.00

6 Simulations results

6.1 Univariate simulation experiment

The indicators we use to describe simulation results are described in section 5.1. For brevity, we tabulate results for the case $m = 100$ and show those related to the case $m = 30$ only whenever differences between the two cases are remarkable.

Results about the shrinkage correction are displayed in table 3. It is apparent that ordinary direct estimates are overdispersed with overdispersion increasing with the average variance of direct estimates; 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 related values around 1 do not seem to follow any significant pattern. The correction of overshrinkage do not seem to be influenced by the number of areas. Table 4 shows results related to Kolmogorov-Smirnov distance. θ^{HB} appear to be the poorest estimators of the true EDF. All other estimators clearly improve the performances of θ^{ZHB}, and the improvement is larger when $m = 100$ with respect to the case of $m = 30$. Among adjusted estimators, θ^{HB} emerges clearly as best. In fact, given the size of MC errors, all observed differences appearing in table 3 can be taken as 95% significant. Moreover, note that the advantage of θ^{ZHB} over the other adjusted methods is less pronounced for Population 2 (poorly informative direct estimates). θ^{CHB} and θ^{CHLB} perform closely and none of the two seems preferable. Table 5 presents the results related to the Anderson-Darling distance. According to this distance we may note that θ^{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 θ^{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 suggested by Zhang turns out to be best with respect to both measures of distance we have considered and it gives its best when the number of areas is large (as explicitly noted in Zhang, 2003) and the direct estimates are not too imprecise.

The results related to the repeated sampling efficiency, as measured by the

Table 4: *Ratio of Kolmogorov-Smirnov distances between estimated and actual EDF averaged over MC replications over the same quantity calculated for the* $\hat{\theta}^{HB}$ *estimators* $(\tilde{D}(\hat{\theta}^{\star}, \theta))$ *univariate simulation*

m	Population	θ^{DIR}	θ^{HB}	θ^{CHB}	θ^{CLHB}	θ^{ZHB}
30	1	0.92	1	0.80	0.79	0.66
30	2	0.80	1	0.69	0.73	0.64
30	3	0.92	1	0.89	0.89	0.74
100	1	0.96	1	0.66	0.62	0.49
100	2	0.86	1	0.57	0.73	0.50
100	3	0.96	1	0.80	0.80	0.64

Table 5: *Ratio of Anderson-Darling distances between estimated and actual EDF averaged over MC replications over the same quantity calculated for the* $\hat{\theta}^{HB}$ *estimators* $\left(\tilde{A}_m^2(\hat{\theta}^{\star}, \theta)\right)$ *- univariate simulation*

m	Population	θ^{DIR}	θ^{HB}	θ^{CHB}	θ^{CLHB}	θ^{ZHB}
30	1	2.12	1	0.62	0.63	0.48
30	2	1.67	1	0.51	0.55	0.50
30	3	1.03	1	0.74	0.73	0.56
100	1	2.63	1	0.34	0.31	0.22
100	2	2.04	1	0.30	0.30	0.28
100	3	1.20	1	0.50	0.49	0.37

Table 6: $amse_{MC}(\hat{\theta}^{\star})$ - *univariate simulation*

m	Population	θ^{DIR}	θ^{HB}	θ^{CHB}	θ^{CLHB}	θ^{ZHB}
100	1	2.819	0.511	0.601	0.678	0.613
100	2	3.591	0.724	0.907	0.936	0.899
100	3	0.511	0.314	0.341	0.347	0.345

empirical unconditional Mean Square Error are shown in Table 6. We can verify that the adjustment of θ^{HB} has, as expected, a cost in terms of efficiency. The increase in posterior Mean Square Error appear to depend on the precision of direct estimators. When the precision is high (Population 3), the rise is around 10%, but when it is low, $amse_{MC}(\hat{\theta}^{\star})$ are 20% or even 30% (in the case of θ^{CHLB}) higher than in the case of posterior variances. Nonetheless we may note that the improvement with respect to the direct estimators remains substantial. Moreover we may observe that θ^{CHB}, θ^{ZHB} show similar performances, while θ^{CHLB} turns out to be a little less efficient. A more detailed analysis of the distribution of $amse_{MC}(\hat{\theta}_i^{\star})$ across areas (for which we show no tables) highlights the very different behavior of θ^{CHLB} with respect to θ^{CHB} and θ^{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 estimators. From (11) we may note that, with respect to posterior mean, these estimators work giving less weight to the synthetic component and more to the direct one. When y_i is very precise this leads to estimators 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 7 reports an evaluation of frequentist properties of this uncertainty measure as defined in (8). It is apparent that posterior MSEs, which are Bayesian uncertainty measures remain sensible measures of variability also with respect to repeated sampling; in fact they are approximately unbiased 'on average (with respect to set of areas being studied). This property, that was known to hold for the posterior variances as frequentist variability measures of θ^{HB} under careful choice of the priors (Datta, Rao and Smith, 2002), is extensible to the case of posterior mean square errors and the prior chosen in our simulation exercise.

6.2 Multivariate simulation experiment

We will report results concerning the multivariate simulation exercise according to the indicators introduced in Section 5.2, focusing for brevity on the case $m = 100$ and omitting those results that do not add anything relevant to the findings of the univariate simulation exercise.
For instance, the comparison of the various estimators with respect to the cor-

Table 7: $amse_{MC}(\hat{\theta}^{\star})$ - *univariate simulation*

m	Population	θ^{HB}	θ^{CHB}	θ^{CLHB}	θ^{ZHB}
100	1	1.02	1.01	1.02	1.02
100	2	10.96	0.99	0.99	0.98
100	3	11.01	1.02	1.02	1.01

Table 8: *Ratio of Kolmogorov-Smirnov distances between estimated and actual EDF averaged over MC replications over the same quantity calculated for the $\hat{\theta}_h^{HB}$ estimators: $\tilde{D}(\hat{\theta}_h^{\star}, \theta_h)$ - multivariate simulation*

m	θ_{hi}	ρ	θ_h^{DIR}	θ_h^{HB}	θ_h^{CHB}	θ_h^{CLHB}	θ_h^{ZHB}
30	θ_{1i}	0	0.90	1	0.85	0.87	0.75
30	θ_{2i}	0	0.84	1	0.56	0.75	0.69
100	θ_{1i}	0	0.97	1	0.89	0.82	0.68
100	θ_{2i}	0	0.88	1	0.52	0.62	0.56
30	θ_{1i}	0.25	0.90	1	0.84	0.86	0.74
30	θ_{2i}	0.25	0.79	1	0.53	0.73	0.68
100	θ_{1i}	0.25	0.96	1	0.87	0.81	0.68
100	θ_{2i}	0.25	0.86	1	0.51	0.60	0.55
30	θ_{1i}	0.75	0.85	1	0.79	0.84	0.71
30	θ_{2i}	0.75	0.81	1	0.54	0.75	0.69
100	θ_{1i}	0.75	0.90	1	0.83	0.77	0.63
100	θ_{2i}	0.75	0.94	1	0.57	0.66	0.58

rection of overshrinkage yields results that parallel perfectly those illustrated for the univariate experiment: direct estimates $\hat{\boldsymbol{\theta}}^{DIR}$ are overdispersed, posterior means $\hat{\boldsymbol{\theta}}^{HB}$ underdispersed, adjusted estimators substantially show a substantially correct sample variance. Moreover the correction appear to be independent of the number of areas and the level of correlation between the two components of $\hat{\boldsymbol{\theta}}$.
From table 8 (which reports results about the Kolmogorov-Smirnov distances) we may note that adjusted estimators improve direct estimators and posterior means. When direct estimates are more precise ($h = 1$, $\theta_{hi} = \theta_{1i}$), $\hat{\theta}_h^{ZHB}$ ($h = 1, 2$) are clearly better than other adjusted methods, both when $m = 100$ and that is more interesting, when $m = 30$. On the contrary when direct estimates are less precise ($h = 2$, $\theta_{hi} = \theta_{2i}$) the picture is different with $\hat{\theta}_h^{CHB}$ clearly better than $\hat{\theta}_h^{ZHB}$ when $m = 30$ and still somewhat better when $m = 100$. The $\hat{\theta}_h^{CHLB}$ ensemble estimators never appear to be the best choice.

If we turn to the Anderson-Darling distance (table 9), the performance of di-

Table 9: *Ratio of Anderson-Darling distances between estimated and actual EDF averaged over MC replications over the same quantity calculated for the* $\hat{\theta}_h^{HB}$ *estimators:* $\tilde{A}_m^2(\hat{\theta}_h^\star, \theta_h)$ *- multivariate simulation*

m	θ_{hi}	ρ	θ_h^{DIR}	θ_h^{HB}	θ_h^{CHB}	θ_h^{CLHB}	θ_h^{ZHB}
30	θ_{1i}	0	1.24	1	0.80	0.80	0.66
30	θ_{2i}	0	9.55	1	0.68	0.71	0.62
100	θ_{1i}	0	1.56	1	0.59	0.60	0.49
100	θ_{2i}	0	10.33	1	0.38	0.38	0.34
30	θ_{1i}	0.25	1.18	1	0.77	0.76	0.63
30	θ_{2i}	0.25	8.62	1	0.62	0.65	0.59
100	θ_{1i}	0.25	1.42	1	0.55	0.56	0.46
100	θ_{2i}	0.25	9.29	1	0.36	0.36	0.33
30	θ_{1i}	0.75	1.05	1	0.73	0.72	0.58
30	θ_{2i}	0.75	9.17	1	0.63	0.66	0.60
100	θ_{1i}	0.75	1.25	1	0.51	0.51	0.41
100	θ_{2i}	0.75	12.91	1	0.41	0.42	0.37

rect estimators $\hat{\theta}_h^{DIR}$ deteriorates and also that the adjusted estimators improve them more sensibly than in the case of the Kolmogorov-Smirnov. Focusing on the adjusted estimators, we observe that $\hat{\theta}_h^{ZHB}$ are best in all situations. By the way, their advantage over the competitors is larger when the number of areas is 100 and when direct estimates are more precise. Moreover, we may also note that different levels of correlation between the components of $\boldsymbol{\theta}$ have a minor impact on comparisons for both distances.

As regards sampling efficiency of estimators as estimated by $amse_{MC}(\hat{\theta}_h^\star)$, a picture very similar to that of the univariate case can be obtained. θ_h^{HB} are by far more efficient than direct estimators: their $amse_{MC}$ is about 1/2 int case of θ_1 and 1/3 in the case of θ_2, regardless of ρ. Adjusted estimators show $amse_{MC}$ about 10% higher than θ_h^{HB} and this increase is not sensitive to the number of areas and the correlation level ρ. All adjusted estimators perform closely in terms of $amse_{MC}$, even though considerations about the distribution of $amse_{MC}(\theta_h^{CHLB})$, similar to those already illustrated in Section 6.1 could be made. Moreover the approximate unbiasedeness of the posterior Mean Square Error (18) as estimator of the frequentist MSE that held in the univariate case continues to hold in the multivariate one. Tables concerning these results are not shown to save space.

Table 10 illustrates whether the considered multivariate adjusted estimators are able to reproduce the true correlation between the components of $\boldsymbol{\theta}$. We may observe that θ_h^{CHB} and θ_h^{CHLB} achieve this result, even if with a some approximation when $m = 30$. This may in part ascribed to the MC error (about 0.02 for this case) inflated by the fact that the sample correlation coefficient computed on 30 items is more unstable than that computed on 100. On the

Table 10: *Actual correlation ρ between the components of and acorr($\hat{\boldsymbol{\theta}}^{\star}$ - multivariate simulation*

m	ρ	θ_h^{DIR}	θ_h^{HB}	θ_h^{CHB}	θ_h^{CLHB}	θ_h^{ZHB}
30	0	0.24	-0.25	-0.03	-0.03	-0.25
100	0	0.25	-0.24	-0.05	-0.02	-0.25
30	0.25	0.36	0.05	0.17	0.18	0.05
100	0.25	0.37	0.13	0.22	0.23	0.13
30	0.75	0.59	0.69	0.59	0.62	0.70
100	0.75	0.60	0.83	0.73	0.73	0.83

contrary θ_h^{ZHB} behaves under this respect just like the ensemble of posterior means, thus failing to recover the actual correlation between the components of in the case of linear independence or low correlation. Nonetheless it should be noted that when ρ is high (a situation in which recovering the actual correlation is really important) the posterior means and all the adjusted estimators show an empirical correlation between the components not far from its true value.

7 Concluding remarks

In this paper we discuss and compare three different methods for adjusting a set of small area estimates in order to make them better estimators of the EDF of the 'ensemble' of Small Area parameters. Two of these methods (CHB and CLHB) are well known in the literature, while the third (ZHB) is more recent. In all cases we consider also their extensions to the multivariate case, as multivariate models are widely employed in Small Area estimation and may provide further gains in efficiency when the vector of parameters being estimated involves correlated variables.
The evaluation of the performances of the three estimators is mainly based on a simulation exercise covering a range of situations hopefully relevant for small area practitioners. All simulations are carried out assuming normality of residuals and random effects. A first conclusion from this analysis is that all the methods considered meet the goal of substantially correcting the 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 the gain in precision with respect to direct estimators remain substantial.
In the univariate case, the Zhang seems as better than the other two. The comparisons based on the bivariate simulation exercise provide a little more blurred picture. All adjusted estimators still appear very effective, but there is no a so clear edge of one method over the others as in the univariate case. The Zhang method in this case seems better in terms of the distance measures considered but fails to recover the actual correlation between the components of the vectors

of small area means being estimated for the cases of moderate or null correlation. This was in part to be expected, given the fact that the generalization of the Zhang's method is more difficult and deserves further investigation and refinements.
Eventually we also studied the frequentist behavior of measure of uncertainty associated to adjusted Hierarchical Bayes estimators, notably finding that, at least for the prior distributions chosen in the simulation exercise, the posterior MSEs have good frequentist properties.

Acknowledgments

The work of Enrico Fabrizi was partially supported by the grants 60FABR06 and 60BIFF04, University of Bergamo.

A Proof of (24)

To prove that

$$\lim_{m\to\infty} \hat{\boldsymbol{\theta}}_i^{CB} = \hat{\boldsymbol{\theta}}_i^{CLB}$$

when $\Psi_i = \Psi$, note first that $\hat{\boldsymbol{\theta}}_i^{CB} = \hat{\bar{\boldsymbol{\theta}}}^B + (\mathbf{H}_1 + \mathbf{H}_2)^{1/2}\mathbf{H}_2^{-1/2}(\hat{\boldsymbol{\theta}}_i^B - \hat{\bar{\boldsymbol{\theta}}})$ with $\hat{\boldsymbol{\theta}}_i^B = \Gamma\mathbf{y}_i + (\mathbf{I}_p - \Gamma)\boldsymbol{\mu}$, $\Gamma = \Sigma_v(\Psi + \Sigma_v)^{-1}$, $\hat{\bar{\boldsymbol{\theta}}}^B = m^{-1}\sum_{i=1}^m \hat{\boldsymbol{\theta}}_i^B = \Gamma\bar{\mathbf{y}} + (\mathbf{I}_p - \Gamma)\boldsymbol{\mu}$. It can be shown (Ghosh and Maiti, 1999) that $\mathbf{H}_1 = (m-1)(\Psi^{-1} + \Sigma_v^{-1})^{-1}$ and $\mathbf{H}_2 = \Gamma\big[\sum_{i=1}^m(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^t\big]\Gamma^t$. From standard matrix algebra we have that

$$(\Psi^{-1} + \Sigma_v^{-1})^{-1} = \Sigma_v - \Sigma_v(\Psi + \Sigma_v)^{-1}\Sigma_v = (\mathbf{I}_p - \Gamma)\Sigma_v$$

As a consequence $\mathbf{H}_1 = (m-1)(\mathbf{I}_p - \Gamma)\Sigma_v$.
Moreover $\lim_{m\to\infty}(m-1)^{-1}\sum_{i=1}^m(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^t = (\Psi + \Sigma_v)$ so, for large m we may write that, approximately $\mathbf{H}_2 = (m-1)\Gamma(\Sigma_v + \Psi)\Gamma^t$ and as a consequence $(\mathbf{H}_1 + \mathbf{H}_2)^{1/2}\mathbf{H}_2^{-1/2} = \Gamma^{-1/2}$.
This implies that $\hat{\boldsymbol{\theta}}_i^{CB} = \Gamma^{1/2}\mathbf{y}_i + (\mathbf{I}_p - \Gamma^{1/2})\boldsymbol{\mu} = \hat{\boldsymbol{\theta}}_i^{CLB}$ and as $\lim_{m\to\infty}\bar{\mathbf{y}} = \boldsymbol{\mu}$, $\lim_{m\to\infty}\hat{\boldsymbol{\theta}}_i^{CB} = \hat{\boldsymbol{\theta}}_i^{CLB}$.

References

[1] Anderson T.W. and Darling D.A. (1954) A test of goodness of fit, *Journal of the American Statistical Association*, 49, 765-769.

[2] Cowles M.K. and Carlin B.P. (1996) Markov Chain Monte Carlo convergence diagnostics: a comparative review, *Journal of the American Statistical Association*, 91, 833-904.

[3] Datta G.S., Rao J.N.K. and Smith D.D. (2002) On measures of uncertainty of small area estimators in the Fay-Herriot model, *Technical Report*, University of Georgia, Athens.

[4] European Commission (2004) A new partnership for cohesion: convergence, competitiveness, cooperation. *Third report on economic and social cohesion*. COM/2004/107, Luxembourg, Office for Official Publications of the European Commission

[5] Fabrizi E., Ferrante M.R. and Pacei S. (2005) Estimation of poverty indicators at sub-national level using multivariate small area models, *Statistics in Transition*, 7, 587-608.

[6] Fay R. and Herriot R.A. (1979) Estimates of income for small places: an application of James-Stein procedures to Census data, *Journal of the American Statistical Association*, 74, 269-277.

[7] Gelman A. (2006) Prior distribution for variance parameters in hierarchical models, *Bayesian Analysis*, 1, 515-533.

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

[9] Ghosh M. and Maiti T. (1999) Adjusted Bayes estimators with applications to Small Area Estimation, *Sankhya*, Series B, 61, 71-90.

[10] Ghosh M., Nangia N. and Kim D. (1996) Estimation of median income of four-person familes: a Bayesian time series approach, *Journal of the American Statistical Association*, 91,1423-1431.

[11] Ghosh M. and Rao J.N.K. (1994) Small Area Estimation: an Appraisal, *Statistical Science*, 9, 55-93.

[12] Goldstein H. (1975) A Note on some Bayesian Nonparametric Estimates, *Annals of Statistics*, 3, 736-740.

[13] Heady P. and Ralphs M. (2004) Some findings of the EURAREA project and their implications for statistical policy, *Statistics in Transition*, 6, 641-653.

[14] Judkins D.R. and Liu J. (2000) Correcting the Bias in the Range of a Statistic across Small Areas, *Journal of Official Statistics*, 16, 1-13.

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

[16] Mukhopadhyay P. (1998) Small Area Estimation in Survey Sampling, Narosa Publishing House, New Delhi.

[17] Rao J.N.K. (2003) Small Area Estimation, Wiley Series on Survey Methodology, John Wiley and Sons, New York.

[18] 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 `http://www.R-project.org`.

[19] Spjøtvoll E. and Thomsen I. (1987) Application of some empirical Bayes methods to Small Area Estimation, *Bullettin of the International Statistical Institute, vol. 2* pp. 435-449.

[20] Thomas A. and O'Hara B. (2006) The BRugs package, Software and Documentation, *downloadable at* `http://cran.r-project.org/org/packages/BRugs`.

[21] Thomas A., O Hara B., Ligges U., Sturz S. (2006) Making BUGS Open, ***R** News*, 6, pp. 12-17.

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

Redazione

Dipartimento di Matematica, Statistica, Informatica ed Applicazioni
Università degli Studi di Bergamo
Via dei Caniana, 2
24127 Bergamo
Tel. 0039-035-2052536
Fax 0039-035-2052549

La Redazione ottempera agli obblighi previsti dall'art. 1 del D.L.L. 31.8.1945, n. 660 e successive modifiche

Stampato nel 2007
presso la Cooperativa
Studium Bergomense a r.l.
di Bergamo