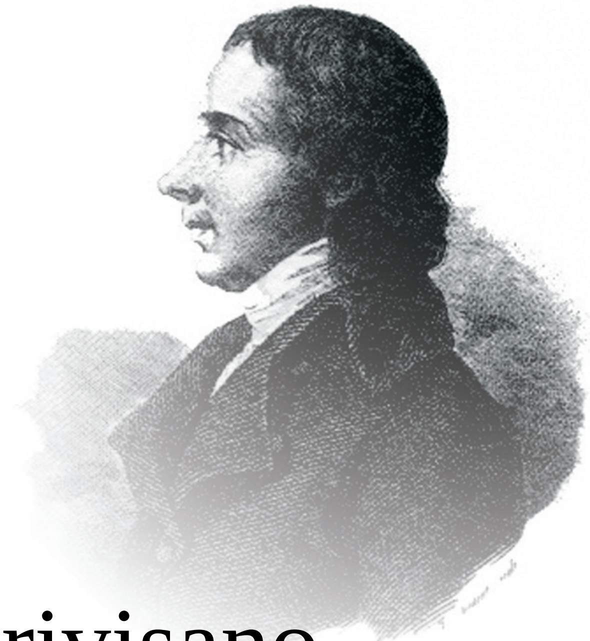
Rapporto n. 11 2007
dmsia unibg.it



E. Fabrizi and C. Trivisano

Robust models for mixed effects in linear mixed models applied to Small Area Estimation

Serie Ricerca

Dipartimento
di Matematica, Statistica, Informatica e Applicazioni
"Lorenzo Mascheroni"

UNIVERSITÀ DEGLI STUDI DI BERGAMO

Robust models for mixed effects in linear mixed models applied to Small Area Estimation

Enrico Fabrizi
DMSIA, Università di Bergamo

Carlo Trivisano
Dipartimento di Scienze Statistiche 'P. Fortunati',
Università di Bologna

Abstract

Hierarchical models are popular in many applied statistics fields including Small Area Estimation. A well known model in this field is the Fay-Herriot model, in which unobservable parameters are assumed Gaussian. In Hierarchical models assumptions about unobservable quantities are difficult to check. Sinharay and Stern (2003) for a special case of the Fay-Herriot model, showed that violations of the assumptions about the random effects are difficult to assess using posterior predictive checks. They conclude that this may represent a form of model robustness . In this paper we consider two extensions of the Fay-Herriot model in which the random effects are supposed to be distributed according to either an Exponential Power (EP) distribution or a skewed EP distribution. The aim is to explore the robustness of the Fay-Herriot model for the estimation of individual area means as well as the Empirical Distribution Function of their 'ensemble'. Based on a simulation experiment our findings are largely consistent with those of Sinharay and Stern as far as the efficient estimation of individual small area parameters is concerned. On the contrary, when the aim is the estimation of the Empirical Distribution Function of the 'ensemble' of small area parameters, results are more sensitive to the failure of distributional assumptions.

Keywords: Fay-Herriot model, Exponential Power distribution, Skewed Exponential Power distribution, 'ensemble' estimators.

1 Introduction

Hierarchical models, in which data $\mathbf{y}$ are modeled conditional on a collection of parameters θ and these parameters are in turn described by a probability distribution involving underlying parameters α are very popular in applied statistics. In recent years, they have been widely applied also in Small Area Estimation.

A well known model in this field, the Normal-Normal Fay-Herriot model (Fay and Herriot, 1979), may be described by the following assumptions:

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

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

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

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

$$\alpha = (\beta, \sigma_v^2) \sim D(\beta, \sigma_v^2) \tag{5}$$

where $1 \leq i \leq m$; $\mathbf{y} = \{y_i\}$ is a collection of 'direct' design unbiased estimates of some set of small area parameters $\boldsymbol{\theta} = \{\theta_i\}$ and whose design-based variances $\{\Psi_i\}$ are assumed to be known. The underlying area-specific parameters are related through a $p-$vector 'auxiliary information' $\{\mathbf{x}_i\}$ and 'random effects' $\{v_i\}$ with a common variance parameter.

In the Fay-Herriot model, normality of e_i $\forall i$ relies on central limit theorem arguments, but normality of v_i $\forall i$ is more difficult to justify. In general, in hierarchical models assumptions about unobservable quantities are difficult to check. Sinharay and Stern (2003), for a special case of the Fay-Herriot model in which $\mathbf{x}_i^T \beta = \mu$ and $\Psi_i = \Psi$ showed that violations of the assumptions about the random effects are difficult to assess using posterior-predictive checks, unless the direct estimates are very informative (that is the sampling variance Ψ is small in face of σ_v^2). They emphasize that this may represent a form of model robustness as 'the posterior distribution of the Normal-Normal model appears to describe adequately the data even when population distribution is not normal'.

By the way, the study of Sinharay and Stern is limited to some special aspect of inference (estimation of σ_v^2). Other aspects may be more sensitive to violations of distributional assumptions. In fact, many small area investigations often have multiple statistical goals: beside efficient estimation of small area parameters $\boldsymbol{\theta}$, we may in fact wish to properly estimate nonlinear functionals of $\{\theta_i\}$ such as their Empirical Distribution Function (EDF). An example in this direction, within the experience of one of the authors (Fabrizi *et al.*, 2005) is represented by the studies in regional economic disparity where poverty parameters should be efficiently estimated on a regional basis without misrepresenting their EDF, which would be equivalent to understate or overstate regional disparities.

In this work, we consider two extensions of the Fay-Herriot model in which the random effects are supposed to be distributed according to either an Exponential Power (EP) Distribution or a Skewed Exponential Power (SEP) Distribution. The EP distribution (Box and Tiao, 1973, ch.3) is a symmetric real valued random variate that generalizes the Normal, allowing for both lighter and heavier tails, while preserving some of its nice properties (presence of location and scale parameters given by the mean and standard deviation). The SEP distribution (Azzalini, 1986) provides a further generalization of the EP distribution encompassing symmetric and skewed, moderately heavy and light tailed distributions.

Also the SEP includes the Normal distribution as a special case.
The use of the EP distribution is often considered in studies of Bayesian robusteness (Berger 1994); in a context very similar to the one we are considering, Datta and Lahiri (1995) suggest the use of an EP prior for all the random effects associated to non-outlying areas.
The main goal of this paper is to explore the performances of the Fay-Herriot model under alternative (and more general) assumptions for the distribution of the random effects. In doing so, we implicitly study the robusteness of the model based on the normality of random effects. A special attention will be devoted to estimation of the EDF of the 'ensemble' of the areas. For this last purpose , we focus on a simultanous estimation method recently proposed by Zhang (2003). In fact, the collection of posterior means, often referred to as 'Bayes estimators' (Ghosh, 1992), has been proven to be a poor estimator of the actual variance of the population of underlying area parameters, let alone its EDF, even when Normality holds (see Louis, 1984 or, more focused on real applications, Heady and Ralphs, 2004). The performances of the Zhang's simultaneous estimation methods have been favorably compared to those of other 'adjusted' estimators discussed in the literature in Fabrizi *et al.* (2007).
We compare the 'Normal-Normal', the 'Normal-EPD' and 'Normal-SEP' Fay-Herriot models by means of a simulation exercise in which data are generated according to a range distributions for the random effects including the Double Exponential Power, the Student's t, the Cauchy, the Lognormal. Not all distributions are special cases of the EP or SEP distributions. We wish in fact to assess whether these more flexible (but still well-behaving) distributions are able to handle cases in which the Normal is known to work poorly.
The paper is organized as follows. In section 2 the EP and SEP distributions are shortly reviewed. In section 3 the robustified Fay-Herriot models based on the EP and SEP distribution for the random effects are introduced and discussed within the framework of Hierarchical Bayes estimation. The Zhang's simultaneous estimation method is reviewed in section 4. Section 5 introduces the simulation exercise, whose results are described in section 6. In section 7 the use of the additional parameters in the EP and SEP distribution as tools for detecting non-normality is discussed. Section 8 provide some concluding discussion.

2 The Exponential Power and the Skewed Exponential Power distributions

2.1 The exponential power distribution

The Exponential Power (EP) distribution is a three parameters distribution whose density is given by:

$$f_{EP}(x|\mu,\sigma,\varphi) = \frac{c_1}{\sigma}\exp\left\{-\left|\frac{\sqrt{c_0}}{\sigma}(x-\mu)\right|^{\frac{1}{\varphi}}\right\}\mathbf{I}_{(-\infty,+\infty)}(x) \tag{6}$$

with $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$, $\varphi \in (0,1]$,

$$c_0 = \frac{\Gamma(3\varphi)}{\Gamma(\varphi)} \quad \text{and} \quad c_1 = \frac{\sqrt{c_0}}{2\varphi\Gamma(\varphi)}.$$

This parametrization is essentially the one adopted by Choy and Walker (2003) except for the fact that the shape parameter φ ranges in the unit interval rather than in $(0,2]$. We prefer this parametrization to the more popular one adopted by Box and Tiao (1973) because it implies $E(X) = \mu$ and $Var(X) = \sigma^2$, a property that can be very useful in modeling. The EP distribution is symmetric around μ so all odds non central moments $E(X-\mu)^{2q+1} = 0 \quad \forall q \in \mathbb{N}$. It can be easily shown that μ and σ are location and scale parameter respectively. Moreover φ is a shape (kurtosis) parameter; in fact the excess of kurtosis is

$$\gamma_2 = \frac{\Gamma(\varphi)\Gamma(5\varphi)}{[\Gamma(3\varphi)]^2} - 3$$

The EP distribution has been widely discussed as a robust alternative to the Normal distirbution in the Bayesian literature (see for instance, Box and Tiao (1973, Ch. 3) or Walker and Gutierrez Peña, 1999); in fact the Normal distribution can be obtained as a special case for $\varphi = 0.5$. For $\varphi < 0.5$ the EP distribution is platikurtik. The distribution is not defined for $\varphi = 0$ but for $\varphi \to 0$ a uniform distribution is obtained.
More interestingly, for $\varphi > 0.5$ distributions with tails heavier than the Normal are obtained, up to the Laplace distribution that corresponds to $\varphi = 1$. Note that in this case $\gamma_2 = 6$; the EP distribution cover a wide range of distributions, including heavy tails one, but it always has a regular behaviour and finite absolute moments of all orders. This property makes it suitable as modeling alternative to the Normal except in case of huge deviations from normality, such as those arising from extreme outliers.
Another property of the EP distribution is that for $\varphi \in [0.5, 1]$ it may be represented as a scale mixture of Normal distributions (West, 1987; Andrews and Mallows, 1974): this property is relevant in our context as it makes the EP suitable for modeling heteroskedastic residuals (or random effects).

2.2 The Skewed Exponential Power Distribution

The Skewed Exponential Power (SEP) is a skewed four parameters distribution whose domain is given by $\mathbb{R}$ and whose density may be expressed as:

$$f_{SEP}(x|\mu,\sigma,\varphi,\lambda) = \frac{2\sqrt{c_0}}{\sigma\varphi^{\varphi}}\phi(w) f_{EP}(x|\mu,\sigma,\varphi)\mathbf{I}_{(-\infty,+\infty)}(x) \tag{7}$$

$\mu, \lambda \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$, $\varphi \in (0,1]$. Moreover $w = sign(z)|z|^{\frac{1}{2\varphi}}\lambda(2\varphi)^{\frac{1}{2}}$, $\phi(\quad)$ is the standard Normal density function and

$$z = \frac{x-\mu}{\sigma\varphi^{\varphi}}\sqrt{c_0}$$

The parameters μ, σ, φ are location, scale and shape parameters of an EP distribution but it is worth stressing that they play a different role in the SEP distribution. The fourth parameter λ is a skewness parameter although skewness depends jointly on λ and φ.

The SEP distribution was first introduced by Azzalini (1986) to provide a probabilistic model handling skewness and heavy tails simultaneously. Inferential aspects and application to the analysis of real data of the SEP distribution are discussed in Di Ciccio and Monti (2004).

The SEP distribution is a little difficult to describe. μ and σ are location and scale parameters but they are not elementary functions of the mean and standard deviation. There are not simple formulas for the moments; Di Ciccio and Monti (2004) provide fromulas for a special case ($f_{SEP}(x|0,1,\varphi,\lambda)$) but they involve infinite series expansions. For this special case they also show that the SEP covers a relatively wide ranging values of skewness ($\gamma_1 \in [-2,2]$) and kurtosis ($\gamma_2 \in (-0.4,6)$) thus finding that this subset of SEP distribution family can then be as leptokurtic as the Laplace and as skewed as the one parameter exponential distribution. For the general case we studied the behavior of the SEP in terms of skewness by means of a simulation exercise based on the random number generator described in Di Ciccio and Monti (2004). We found that skewness is not sensitive to the values of μ and σ but depends jointly on λ and φ. In particular we found that, approximately, $\gamma_1 \in (-2\varphi, 2\varphi)$. Moreover, although λ can take any real value, the range of values for which it has an appreciable impact on γ_1 is more limited. For $\varphi = 1$, that is SEP distribution with the heaviest possible tails, the range of the values of λ influential of the skewness is approximately $(-2,2)$ as it is shown Figure 1 where the values of λ are plotted against those of γ_1 for the case $\mu = 0$ and different values of the scale parameter σ. For $\varphi \in (0.5,1)$, that is the subset of distributions with tails heavier than the normal, the range of λ values influencing the distribution skewness is approximately $(-\frac{1}{\varphi}, \frac{1}{\varphi})$, while for $\varphi \to 0$ this range gets larger and larger. But as we noted before the skewness of the distribution depends on φ and as $\varphi \to 0$ the SEP distribution approach symmetry regardless of λ.

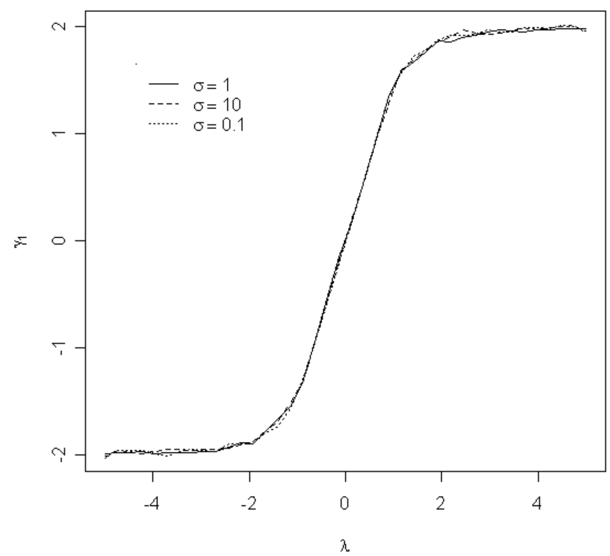


Figure 1: λ plotted against γ_1 for $\mu = 0$, $\varphi = 1$ and various values of σ

3 Robustified Fay-Herriot models

3.1 A Normal-EP two levels model

Let's consider a robustified Fay-Herriot model (**??**) - (**??**) in which the normality of random effects (**??**) is replaced by $v_i \overset{ind}{\sim} EP(0, \sigma, \varphi)$. As μ is the mean of the EP distribution this is equivalent to $\theta_i \overset{ind}{\sim} EP(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma, \varphi)$. Since in Small Area Estimation we are mostly interested in predicting small area means $\boldsymbol{\theta} = \{\theta_i\}_{1 \leq i \leq m}$ the inferential analysis will be focused on

$$p(\boldsymbol{\theta}|\mathbf{y}) = \int_\beta \int_\sigma \int_\varphi p(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma, \varphi|\mathbf{y}) d\beta d\sigma d\varphi \tag{8}$$

It can be shown that the posterior $p(\boldsymbol{\theta}|\mathbf{y})$ cannot be obtained analytically for standard choice of prior distributions. Let's assume a priori independence (i.e. $p(\boldsymbol{\beta}, \sigma, \varphi) = p(\boldsymbol{\beta})p(\sigma)p(\varphi)$ and $p(\boldsymbol{\beta}) \propto \mathbf{I}_{(-\infty,+\infty)}$, $p(\sigma) \propto \mathbf{I}_{(0,+\infty)}$, $p(\varphi) \propto \mathbf{I}_{(0,1]}$. The prior on β is a standard improper flat prior often used in normal regression models and was considered by Fay and Herriot (1979) in the analysis their model. The flat prior on the standard deviation is not Jeffrey's but has recently been recommended by Gelman (2006) as ensuring posterior property in a Normal-Normal model such as the Fay-Herriot's and shown to have a mild impact on posterior distributions. The uniform prior on the shape parameter φ

may justified by simplicity.
This choice of priors simplifies the joint posterior distribution:

$$\begin{aligned} p(\boldsymbol{\theta},\boldsymbol{\beta},\sigma,\varphi|\mathbf{y}) &\propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\beta},\sigma,\varphi) \\ &\propto \left(\frac{1}{\sqrt{2\pi\Psi_i}}\right)^m \exp\Big\{-\frac{1}{2}\sum_{i=1}^m\Big(\frac{y_i-\theta_i}{\Psi_i}\Big)^2\Big\}\times \\ &\times \left(\frac{c_1}{\sigma}\right)^m \exp\Big\{-\sum_{i=1}^m\Big|\frac{\sqrt{c_0}}{\sigma}\big(\theta_i-\mathbf{x}_i^T\boldsymbol{\beta}\big)\Big|^{\frac{1}{\varphi}}\Big\} \end{aligned} \tag{9}$$

The scale parameter σ can be easily integrated out using the notable integral

$$\int_0^{+\infty} x^{-(p+1)}\exp(-bx^{-\alpha})dx = \frac{1}{\alpha}b^{-\frac{p}{\alpha}}\Gamma\Big(\frac{p}{\alpha}\Big)$$

with $p=m-1$, $b=-\sum_{i=1}^m\Big|\sqrt{c_0}\big(\theta_i-\mathbf{x}_i^T\beta\big)\Big|^{\frac{1}{\varphi}}$, $\alpha=\varphi^{-1}$. We then obtain

$$\begin{aligned} p(\boldsymbol{\theta},\boldsymbol{\beta},\varphi|\mathbf{y}) &\propto \left(\frac{1}{\sqrt{2\pi\Psi_i}}\right)^m \exp\Big\{-\frac{1}{2}\sum_{i=1}^m\Big(\frac{y_i-\theta_i}{\Psi_i}\Big)^2\Big\}\times \\ &\times [c_1(\varphi)]^m\varphi\Gamma\big(\varphi(m-1)\big)\Big\{\sum_{i=1}^m\big|\sqrt{c_0}\big(\theta_i-\mathbf{x}_i^T\boldsymbol{\beta}\big)\big|^{\frac{1}{\varphi}}\Big\}^{\varphi(m-1)} \end{aligned} \tag{10}$$

About (**??**) we note that, even if we consider the simplifying special case of $\mathbf{x}_i^T=\mathbf{1}$, φ cannot be integrated out using analytical methods (see Box and Tiao, 1973, p. 160). This rules out the possibility of obtaining closed form expressions for $p(\boldsymbol{\theta}|\mathbf{y})$. Among the many possible approximated methods we advocate the use of Markov Chain Monte Carlo Methods that can be easily and efficiently implemented by widely available softwares and may be, for this reason, appealing to practitioners.
In view of the application of MCMC methods, we slightly modified the discussed choice of prior distributions. The rationale of the modification is that of speeding up computation and avoiding convergence and mixing problems of MCMC algorithms. More in detail we assume:

$$\begin{aligned} p(\boldsymbol{\beta},\sigma,\varphi) &= p(\sigma)p(\varphi)\prod_{i=1}^p p(\beta_j) \\ p(\beta_j) &\sim N(0,L) \qquad (11) \\ p(\sigma) &\sim U(0,K) \qquad (12) \\ p(\varphi) &\sim U(0,1] \qquad (13) \end{aligned}$$

where L and K are positive 'large' constants fixed in order to warrant adequate diffusion of (**??**) and easing computation without representing a genuine constraint in the case of (**??**).

The full conditional distributions involved in the Normal-EP robustified Fay-Herriot can be easily worked out; they are non-standard distributions and require of complex sampling methods.
We consider the approximation of the posterior distributions using the OpenBugs software (Thomas *et al.*, 2006) which is very widely used in applied hierarchical modeling (OpenBugs is the open source version of the BUGS software). In OpenBugs there are several choices for specifying the distribution of random quantities but the EP distribution is not considered. There are two possible solutions to this problem. The first is based on representing the EP distribution as a uniform scale mixture:

$$f_{EP}(x|\mu,\sigma,\varphi) = \int_0^\infty U\Big(x\Big|\mu - \frac{\sigma}{\sqrt{2c_0}}u^\varphi, \mu + \frac{\sigma}{\sqrt{2c_0}}u^\varphi\Big) G\Big(u\Big|1+\varphi, 2^{-\frac{1}{\varphi}}\Big) du$$

where $G(\ |a,b)$ is the density function of a *Gamma* random variable (Choy and Chan, 2003). The second is based on a 'trick' for specifying new prior distributions in BUGS (Spiegelhalter *et al.* 2003). Although first solution seems more appealing in a modeling perspective, it works bad from a MCMC point of view: extremely slow convergence and bad mixing of chains associated to hyperparameters (those indexing the EPD prior for mixed effects). This is likely to depend on the fact that the mixture representation adds a further level to hierarchical model and makes full conditional distributions more difficult to sample. We then advocate the use of second solution that proves to be far better for computation. The BUGS 'trick' (you may find a BUGS code in Appendix A) has been tested in controlled settings to assess its right functioning.

3.2 A Normal-SEP two levels model

Similarly to previous section let's consider an alternative to the standard Normal-Normal Fay Herriot model (**??**) - (**??**) by assuming $v_i \overset{ind}{\sim} SEP(0,\sigma,\varphi,\lambda)$ instead of normality of random effects (**??**). This assumption is equivalent to $\theta_i \overset{ind}{\sim} SEP(\mathbf{x}_i^t\beta,\sigma,\varphi,\lambda)$ since μ is a location parameter in the SEP distribution. But, since it is not the mean of the distribution, we are not assuming zero-mean random effects for the model. By the way, the assumption is consistent with application of the SEP distribution to data analyses reported in Di Ciccio and Monti (2004).
The specification of a prior distribution for $(\beta,\sigma,\varphi,\lambda)$ is complicated by the fact that parameters do not have a clear interpretation. Moreover all the off the diagonal elements of the information matrix (obtained by Di Ciccio and Monti, 2004) are non zero, thus making the specification of Jeffrey priors very difficult. In view of the approximation of the posterior distributions by means of MCMC algorithms we assume a priori independence between the components of $(\mu,\sigma,\lambda,\varphi)$ and specify the following priors for individual parameters: *i)* $\mu \sim N(0,K)$, *ii)* $\sigma \sim Unif(0,H)$, *iii)* $\lambda \sim N(0,4)$, *iv)* $\varphi \sim Unif(0,1]$.
Priors on μ, σ are specified considering the nature of location and scale of these parameters and using arguments similar to those of section 3.1. Constants H, K

will be determined in order to be 'large' with respect to the scale of data being analyzed. The prior on λ is the result of several trials and sensitivity analysis; it tries to strike a good balance between non-informativeness (mild impact on posterior inferences on $\boldsymbol{\theta}$) and computational feasibility. We set $sd(\lambda) = 2$ in order to have a non negligible prior probability density over the whole range of λ values that have a substantial impact on skewness in the case of leptokurtic distributions (see section 2.2). For nearly symmetrical distributions of random effects, the parameter λ is in general difficult to identify (i.e. the posterior is diffuse) unless a very concentrated prior is specified (but in this case, of course, its posterior is prior driven). By the way, since the impact of λ on $p(\boldsymbol{\theta}|\mathbf{y})$ is negligible in these cases, the chosen prior on λ remain uninfluent on the inferences concerning $\boldsymbol{\theta}$. As regards φ the prior is selected using arguments parallel to those of section 3.1.

4 Ensemble estimation

As anticipated in the introduction, among various 'adjusted Bayes estimators' proposed in the literature we focus on a method proposed by Zhang (2003). With respect to the more popular constrained Bayes estimator (Ghosh, 1992), it is not focused on the first two moments of the distribution of the 'ensemble' of estimates, so it can be applied straightforwardly to models characterized by non-normal random effects. Moreover, for the estimation of the Empirical Distribution Function of the 'ensemble' of the small area means, Fabrizi *et al.*(2007) show, by means of a simulation exercise, that the Zhang method compares favorably to the Constrained Bayes estimator even under Normality.

A short description of the Zhang's method follows. Given the set $\{\theta_i\}$, $1 \leq i \leq m$ of the area parameters of interest, let $\{\theta_{(i)}\}$ be the associated ordered set $(\theta_{(1)} \leq \theta_{(2)} \leq ... \leq \theta_{(m)})$. Then $\eta_i = E(\theta_{(i)}|\mathbf{y})$ 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 single 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 posterior means of individual θ_is. 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 $\hat{r}_i = E(rank(\theta_i|\mathbf{y}))$, $1 \leq i \leq m$, 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 $\theta_i(s)|\mathbf{y}$ from any draw s of the Markov chain after convergence. Then we can approximate $\hat{r}_i$ averaging the ranks $rank(\theta_i(s)|\mathbf{y})$ over all draws, obtaining $\hat{r}_i^{MC} = S^{-1}\sum_{s=1}^{S} rank(\theta_i(s)|\mathbf{y})$ where S is the number of iterations of Markov Chain after convergence.

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

$$\theta_i^{ZHB} = \eta_{\hat{r}_i} \tag{14}$$

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

5 The simulation exercise

We consider two separate simulation exercises. In the first one the aim is that of comparing the Normal-Normal and the Normal-EP Fay Herriot models with the random effects v_is generated under different distributions. In particular, we consider symmetric distributions with different fatness of the tails: *a*) $v_i \overset{ind}{\sim} N(0,1)$, *b*) $v_i \overset{ind}{\sim} Dexp(0, 1/\sqrt{2})$ (that is Laplace, Double Exponential), *c*) $v_i \overset{ind}{\sim} t(3)$ that is t distribution with 3 degrees of freedom and *d*) $v_i \overset{ind}{\sim} t(1)$ that is a Cauchy distribution.
In the data generation we also assume, for simplicity that $\mathbf{x}_i = \mathbf{1}$ and $\boldsymbol{\beta} = \mathbf{0}$. When modeling, $\boldsymbol{\beta}$ is assumed unknown and given the flat prior (**??**) with $L = 100$ in both the Normal-Normal and the Normal-EP case.
For comparison purposes, when random effects are generated form a t distribution (that does not belong to the EP family) we consider also a Normal-t Fay Herriot model. In both the *c*) and *d*) cases we treat the degrees of freedom parameter κ parameter as an additional random variable. More in detail, we assume κ to be a discrete random variable for which we specify the discrete uniform prior: $\kappa \sim Unif\{A\}$.
Our choice of A is $A = \{1, 2, 3, 4, 5, 6, 8, 10, 12, 16, 20, 25, 30, 50\}$. We recognize that this choice of A is somewhat arbitrary. By the way, a prior that puts a large part of its weight on low or moderate degrees of freedom allowing at the same time for case close to the Normal ($\kappa = 30$, $\kappa = 50$) is known to work reasonably in general and for this reason is recommended by Spiegelhalter *et al.*(2003). In our cases it allows for a 'fair' comparison between the Normal-EP in which the shape parameter is assumed unknown and the Normal-t Fay-Herriot model.
In the second simulation experiment the focus will be on the SEP-Normal Fay-Herriot model that will be compared the Normal-Normal under *e*) $v_i \overset{ind}{\sim} N(0,1)$, *f*) $v_i \overset{ind}{\sim} Dexp(0, 1/\sqrt{2})$, and three different Lognormal distributions assumed for the random effects. We replace (**??**) with

$$\theta_i = \mathbf{x}_i^T \beta + \lambda + v_i \tag{15}$$

in order to have approximately zero meaned random effects $v_i^\star = \lambda + v_i$. As before, in data generation we set $\mathbf{x}_i^T = 1$ and $\beta = 0$. When modeling β will be assumed unknown. Given the density

$$f_{Lnorm(\phi,\tau)}(v) = \sqrt{\frac{1}{2\pi}} \frac{1}{v\tau} e^{-\frac{1}{2\tau^2}\left(\log v - \phi\right)^2} \mathbf{I}_{[0,+\infty)}(v) \tag{16}$$

Table 1: *Different configurations of design variances considered in the simulation experiments*

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

we consider two sets of parameters leading to the following assumptions: $g)\ v_i \overset{ind}{\sim} Lnorm(\log 1.442, \log 1.355)$ and $\lambda = -1.678$; $h)\ v_i \overset{ind}{\sim} Lnorm(\log 0.289, \log 4)$ and $\lambda = -0.578$.
As before, for comparison purposes, when random effects are generated from a Lognormal distribution we consider also a Normal-Lognormal Fay-Herriot model.
The following independent priors are considered: $\exp(\phi) \sim Unif(0, H')$ and $\exp(\tau^2) \sim Unif(1, K')$ where $H' = K' = 100$ are chosen as large constants with respect to the scale of the data. Note that $\exp(\phi)$ is the median of the log-normal distribution, while the coefficients of variation, skewness and kurtosis are all polynomial functions of $\exp(\tau^2)$. We found that for these priors the posterior inferences are very unsensitive to the choice of H', K'. An alternative would be that of specifying priors (e.g. Jeffrey's) on the parameters of $\log(v) = w \sim N(\phi, \tau^2)$; nevertheless we found that these priors have a considerable impact on posterior inferences drawn using MCMC methods in most situations.
The two sets of parameters characterizing the lognormal distributions are set in order to have $Var(v_i) = 1$ and this property holds true also for all distribution $a)$ - $f)$ already discussed for the random effects, with the obvious exception of the Cauchy, whose variance is not finite. The property of $Var(v_i) = 1$ is useful to plug meaningful design variances Ψ_i, $1 \leq i \leq m$ in the data generation. In fact since the main interest is on evaluating different hypotheses on the random effects we assume that

$$e_i \overset{ind}{\sim} N(0, \Psi_i)$$

$1 \leq i \leq m$. We 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 **??**.

They differ in terms of informativeness of direct estimators, that may be measured by $\gamma_i = Var(v_i)[Var(v_i) + \Psi_i]^{-1}$. Population 1 corresponds to a situation in which direct estimators are poorly informative ($\gamma_i \in [0.11, 0.5]$). Population 2 describes a situation where direct estimators show a wide range of informativeness ($\gamma_i \in [0.11, 0.91]$), while in Population 3 we study the case of rather strongly informative direct estimators ($\gamma_i \in [0.5, 0.91]$).
We set the number of areas $m = 100$. We also run the simulations for $m = 30$

(but results are less interesting and not reported) and in some situations, limited by computational burden, in the case $m = 200$. For each experimental setting we generated $1,000$ independent MC samples.
For all models we consider two distinct sets of predictors for the θ_i: the posterior means $\theta_i^{HB} = E(\theta_i|\mathbf{y})$ and the estimator θ_i^{ZHB} defined in (**??**). The comparison between the various models will be based on three main indicators:

1. the average Mean Square Error of the predictors

$$amse_* = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{R}\sum_{r=1}^{R}\left(\theta_{i,r}^* - \theta_{i,r}\right)^2$$

   with $* = HB, ZHB$. Note that R is the number of MC samples generated in the simulation while $\theta_{i,r}^*$, $\theta_{i,r}$ are respectively the HB and ZHB estimators, the actual parameter θ_i and the data in the r^{th} MC generation.

2. The Kolmogorov-Smirnov distance. For each iteration r, the Kolmogorov-Smirnov distance between the actual Empirical Distribution Function of the $\boldsymbol{\theta}_r^{ZHB} = \{\theta_{i,r}^{ZHB}\}$ and that of $\boldsymbol{\theta}_r$ denoted respectively as F_{ZHB} and F_T is calculated as

$$KS_r(\boldsymbol{\theta}_r^{ZHB}, \boldsymbol{\theta}_r) = \max_j |F_{ZHB}(z_{j,r}) - F_T(z_{j,r})|$$

   where $1 \leq j \leq 2m$ and $\mathbf{z}_r = (\boldsymbol{\theta}_r^{ZHB}, \boldsymbol{\theta}_r)$ is the $2m$ vector obtained pooling the true values of the area parameters and the ZHB estimators from the r-th MC replication. The distances calculated at each iteration are then averaged over MC replications:

$$ks = \frac{1}{R}\sum_{r=1}^{R} KS_r(\boldsymbol{\theta}_r^{ZHB}, \boldsymbol{\theta}_r)$$

3. The 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 reference distribution. With respect to the Kolmogorov-Smirnov distance is known to be more influenced by the discrepancies in the tails of the distribution. In the context of our problem, the statistic can be defined as follows:

$$AD_r(\boldsymbol{\theta}_r^{ZHB}, \boldsymbol{\theta}_r) = \frac{1}{4}\sum_{j=1}^{2m}\frac{F_{ZHB}(z_{i,r}) - F_T(z_{i,r})}{F_T(z_{i,r})\big(1 - F_T(z_{i,r})\big)}\mathbf{I}_{(0,1)}(F_T(z_{i,r}))$$

   The distances calculated at each iteration are then averaged over MC replications:

$$ad = \frac{1}{R}\sum_{r=1}^{R} AD_r(\boldsymbol{\theta}_r^{ZHB}, \boldsymbol{\theta}_r)$$

In summary, the first indicator describes how close are point estimators to the true values of the parameters being estimated. We note that it may interpreted as a frequentist measure of variability of the estimators averaged over the set of all areas. The last two indicators measure the performances of the simultaneous estimator $\boldsymbol{\theta}^{ZHB}$ of the EDF of the 'ensemble' of the Small Area means.

6 Simulation results

Results concerning the comparison between the Normal-Normal and the Normal-EP model are reported in Table 2. The first four rows in the table illustrate the situation under distributional assumption a) (normality). The two models perform very closely, with the Normal-EP being a little worse. These close performances are in line with theory expectations and, from a practical point of view reassure us about the mild impact of the prior chosen for the shape parameter φ.

When the random effects are Laplace or $t(3)$ distributed the Normal-Normal model is very close to the Normal-EP and Normal-t models in terms of $amse_{HB}$, that is as far as the estimation of individual θ_i is concerned. The situation gets worse when the random effects are sampled from a Cauchy distribution. Moreover we note that in this case when the direct estimators are precise (population 3) the difference between the Normal-Normal and the other models dwindles. This latter fact may be attributed to the collapse of posterior means on direct estimates caused by the 'overestimation' of σ_v^2, and to the fact that in population 3 direct estimators begin to be more acceptable estimators of the underlying parameters. These findings confirm the results of Sinharay and Stern (2003) that the standard Fay-Herriot model is robust with respect to departures from normality at least as far as point predictors optimal under quadratic loss are involved and unless this departures are huge. Note that under Cauchy distributed random effects the Normal-EP model offers only partial protection against the collapse of small area predictors on the direct estimates as it may expected from the discussion of section 2.1. In this cases alternative estimators such as the 'limited translation' estimator of Efron and Morris (1971) or the one proposed by Datta and Lahiri (1995) may be advisable.
As regards the comparison between $amse_{HB}$ and $amse_{ZHB}$ we note first that in all situations $amse_{HB} < amse_{ZHB}$; this is consistent with the fact that $\boldsymbol{\theta}^{ZHB}$ is a set of sub-optimal predictors of individual θ_i in terms of quadratic loss. As for the effect of distributional assumptions, the behaviour of $amse_{ZHB}$ is similar to that of $amse_{HB}$.
Let's turn to the analysis of the results about the the performances of the simultaneous estimators $\boldsymbol{\theta}^{ZHB}$ as measured by the averaged Kolmogorov-Smirnov and Anderson-Darling distances. If the random effects are Laplace distributed the difference in the performance of the Normal-Normal and Normal-EP is rather small but grows with the precision of direct estimators. Moreover the better performance of the Normal-EP model emerges more clearly for the Anderson-Darling distance. This is consistent with the nature of this statistics that is more

Table 2: *Comparison between the Normal-Normal and Normal-EP Fay-Herriot models*

DGP	Model	Population	$amse_{HB}$	$amse_{ZHB}$	ks	ad
N(0,1)	N-N	1	0.724	0.907	0.123	1.560
N(0,1)	N-EP	1	0.724	0.904	0.123	1.567
N(0,1)	N-N	2	0.512	0.614	0.089	0.614
N(0,1)	N-EP	2	0.514	0.617	0.091	0.654
N(0,1)	N-N	3	0.314	0.343	0.073	0.365
N(0,1)	N-EP	3	0.316	0.344	0.076	0.396
Laplace	N-N	1	0.715	0.899	0.137	1.808
Laplace	N-EP	1	0.711	0.889	0.135	1.757
Laplace	N-N	2	0.511	0.634	0.111	0.931
Laplace	N-EP	2	0.504	0.613	0.099	0.745
Laplace	N-N	3	0.315	0.359	0.098	0.667
Laplace	N-EP	3	0.306	0.341	0.086	0.481
t3	N-N	1	1.366	1.728	0.126	1.704
t3	N-EP	1	1.279	1.574	0.108	1.103
t3	N-t	1	1.231	1.432	0.098	0.757
t3	N-N	2	0.989	1.201	0.103	0.957
t3	N-EP	2	0.918	1.069	0.087	0.544
t3	N-t	2	0.896	1.046	0.084	0.440
t3	N-N	3	0.415	0.475	0.084	0.515
t3	N-EP	3	0.382	0.422	0.073	0.311
t3	N-t	3	0.376	0.409	0.072	0.279
Cauchy	N-N	1	3.439	4.725	0.165	2.817
Cauchy	N-EP	1	2.613	3.489	0.138	1.668
Cauchy	N-t	1	2.022	2.521	0.094	0.510
Cauchy	N-N	2	2.701	3.553	0.131	1.574
Cauchy	N-EP	2	1.998	2.552	0.111	0.939
Cauchy	N-t	2	1.536	2.079	0.079	0.313
Cauchy	N-N	3	0.521	0.618	0.092	0.499
Cauchy	N-EP	3	0.477	0.552	0.081	0.345
Cauchy	N-t	3	0.433	0.480	0.066	0.186

focused on the tails of the distribution with respect to the Kolmogorov-Smirnov. A similar pattern characterizes the results also when the random effects are $t(3)$ distributed, with performances of the Normal-EP model improving with the precision of direct estimators. Moreover the Normal-EP model is quite close to Normal-t model despite the t is not a special case of the EP family. The situation when the random effects are Cauchy distributed is consistent with this scenario, the Normal-EP model being clearly better than the Normal-Normal but rather far to the one based on the right distributional assumption for the random effects.

The results pertaining to the second simulation experiment, in which the Normal-Normal and the Normal-SEP models are compared are displayed in Table 3. First, we may note that, when the random effects are normal, the results obtained under the Normal-SEP are very close to those obtain under the Normal-Normal Fay-Herriot model. This means that, although the former has two additional parameters, they have only a minor impact on the quality of the estimates of small area means. Moreover it implies also that the prior chosen for the parameters of the SEP distribution have mild impact on the posteriors of θs, at least under normality. Similar comments apply when we consider random effects $EP(0, 1, 1)$ (Laplace) distributed and we include also the Normal-EP model in the comparisons.

When the random effects are log-normally distributed we should distinguish what happens in case g (moderate skewness) from case h (large skewness). In the former, the Normal-SEP model performs very closely to the Normal-Lognormal model; in fact although the Lognormal is not within the family of the SEP distributions, the two additional parameters allows a more flexible modeling of the random effects distribution. This advantage is lost when the direct estimates are imprecise (Population 1) and thus provide little information on the distribution of the underlying parameters.

In case h the skewness is higher than that a distribution within the SEP family can reach, and the Fay-Herriot model based on the lognormal assumption for the random effects perform clearly better than the one assuming a SEP. By the way the latter still offers a large improvement with respect to the standard Normal-Normal model.

7 On the use of the posterior distribution of shape and skewness parameters to discriminate between alternative models

The Normal-EP and the Normal-SEP models generalize the ordinary Normal-Normal Fay-Herriot model allowing respectively for 1 and 2 extra parameters in the modeling of the distribution of random effects.

The shape parameter φ in the Normal-EP model has a clear interpretation: it takes value 0.5 under Normality, more if the distribution of random effects is leptokurtic, less if it is platikurtic. The question is whether, in practice, the

Table 3: *Comparison of the Normal-Normal and Normal-SEP Fay-Herriot models*

DGP	Model	Population	amse-HB	amse-ZHB	ks	ad
N(0,1)	N-N	1	0.724	0.897	0.123	1.560
N(0,1)	N-SEP	1	0.727	0.907	0.128	1.683
N(0,1)	N-N	2	0.512	0.614	0.089	0.614
N(0,1)	N-SEP	2	0.515	0.619	0.092	0.669
N(0,1)	N-N	3	0.314	0.343	0.073	0.365
N(0,1)	N-SEP	3	0.316	0.346	0.078	0.426
Laplace	N-N	1	0.715	0.899	0.137	1.808
Laplace	N-EP	1	0.711	0.889	0.135	1.757
Laplace	N-SEP	1	0.711	0.889	0.143	1.801
Laplace	N-N	2	0.511	0.634	0.111	0.931
Laplace	N-EP	2	0.504	0.613	0.099	0.745
Laplace	N-SEP	2	0.504	0.613	0.106	0.832
Laplace	N-N	3	0.315	0.359	0.098	0.667
Laplace	N-EP	3	0.306	0.341	0.086	0.481
Laplace	N-SEP	3	0.309	0.343	0.094	0.525
Lnorm(g)	N-N	1	0.713	0.938	0.175	3.702
Lnorm(g)	N-SEP	1	0.713	0.938	0.158	3.209
Lnorm(g)	N-LOGN	1	0.665	0.812	0.161	3.455
Lnorm(g)	N-N	2	0.508	0.651	0.132	2.027
Lnorm(g)	N-SEP	2	0.476	0.572	0.102	1.011
Lnorm(g)	N-LOGN	2	0.477	0.577	0.112	1.166
Lnorm(g)	N-N	3	0.311	0.377	0.117	1.537
Lnorm(g)	N-SEP	3	0.273	0.308	0.089	0.675
Lnorm(g)	N-LOGN	3	0.273	0.301	0.103	0.842
Lnorm(h)	N-N	1	0.613	0.908	0.309	12.004
Lnorm(h)	N-SEP	1	0.485	0.721	0.304	11.101
Lnorm(h)	N-LOGN	1	0.426	0.570	0.282	10.586
Lnorm(h)	N-N	2	0.459	0.666	0.256	8.153
Lnorm(h)	N-SEP	2	0.335	0.407	0.224	6.142
Lnorm(h)	N-LOGN	2	0.315	0.370	0.285	6.578
Lnorm(h)	N-N	3	0.274	0.392	0.278	7.552
Lnorm(h)	N-SEP	3	0.182	0.239	0.216	5.933
Lnorm(h)	N-LOGN	3	0.165	0.191	0.177	3.855

Table 4: *Posterior summaries of $p(\varphi|\mathbf{y})$ averaged over MC replications. The argument $\varphi|\mathbf{y}$ is omitted to save space.*

m	DGP	Pop.	$E_{MC}(E)$	$E_{MC}(F_{0.025})$	$E_{MC}(F_{0.975})$	$E_{MC}(\mathbf{I}_{F_{0.025}>0.5})$
100	Laplace	1	0.522	0.077	0.959	0.025
100	Laplace	2	0.695	0.321	0.976	0.245
100	Laplace	3	0.730	0.389	0.978	0.345
200	Laplace	1	0.576	0.122	0.975	0.085
200	Laplace	2	0.774	0.481	0.981	0.555
200	Laplace	3	0.810	0.554	0.984	0.715
100	t3	1	0.734	0.333	0.987	0.310
100	t3	2	0.821	0.553	0.989	0.695
100	t3	3	0.863	0.648	0.992	0.845
100	Cauchy	1	0.978	0.927	0.999	0.995
100	Cauchy	2	0.981	0.935	0.999	1
100	Cauchy	3	0.985	0.946	0.999	1

posterior distribution $p(\varphi|\mathbf{y})$ provides enough evidence of non-normality in the distribution of the random effects when it is the case.
Some results on this point , based on the simulation exercise introduced in section 5, are displayed in Table 4, where the average value over MC replications of some summary statistics of $p(\varphi|\mathbf{y})$, i.e. $E_{MC}\{E(\varphi|\mathbf{y}_r)\}$, $E_{MC}\{F_{0.025}(\varphi|\mathbf{y}_r)\}$, $E_{MC}\{F_{0.975}(\varphi|\mathbf{y}_r)\}$ are reported along with the proportion of $E_{MC}\{\mathbf{I}_{F_{0.025}(\varphi|\mathbf{y}_r)>0.5}\}$. This latter statistics simply tells us the frequency of posterior 0.95 probability intervals not including 0.5, thus providing evidence of non-normality in the random effects.
Random effects generated according to $EP(0,1,1)$, $t(3)$ and $t(1)$ are considered.

It is easy to note that, when deviation from Normality is huge (the Cauchy case) $p(\varphi|\mathbf{y})$ clearly identifies non-normality. This is case also when $v_i \overset{ind}{\sim} t(3)$ but only when the variance of the direct estimators is low.
For Laplace generated random effects we have considered both the cases of $m = 100$ and $m = 200$. In the first, the ability of $p(\varphi|\mathbf{y})$ to signalize non-normality ranges from virtually null to moderate depending on the precision of direct estimators. The situation improves in the second case. Results for values of m smaller than 100, not reported, show that the ability of $p(\varphi|\mathbf{y})$ to signalize non-normality is reduced; results for larger m are not produced due to computational burden. In conclusion it seems that the use of the posterior distribution $p(\varphi|\mathbf{y})$ to draw conclusions about non-normality of random effects even in cases of 'mild' deviations from Normality requires that the number of areas in the problem is large, where how large is enough depends critically on the precision of direct estimators.
Similar findings can be obtained about the skewness parameter λ in the Normal-

Table 5: *Posterior summaries of $p(\lambda|\mathbf{y})$ averaged over MC replications. The argument $\lambda|\mathbf{y}$ is omitted to save space.*

m	DGP	Pop.	$E_{MC}(E)$	$E_{MC}(F_{0.025})$	$E_{MC}(F_{0.975})$	$E_{MC}(\mathbf{I}_{F_{0.025}>0.5})$
100	Lognorm(g)	1	0.861	-2.574	4.321	0.165
100	Lognorm(g)	2	1.801	-0.767	4.748	0.565
100	Lognorm(g)	3	2.088	-0.109	4.874	0.725
100	Lognorm(h)	1	1.016	-2.179	4.391	0.285
100	Lognorm(h)	2	2.230	-0.025	5.020	0.740
100	Lognorm(h)	3	2.459	0.458	5.119	0.835

SEP Fay-Herriot model. The posterior $p(\lambda|\mathbf{y})$ detects the skewness in the distribution of the random effects with an effectiveness that grows with the precision of direct estimators and the number of areas involved in the estimation. Results are reported in Table 5.

With reference to the Normal-SEP model we note that, as far as the detection of excess in kurtosis is concerned, the behavior of the kurtosis parameter φ is similar to that observed in the Normal-EP model.
Sinharay and Stern (2003) base their discussion of the robusteness of the Normal-Normal model on posterior predictive checks (Rubin, 1984). Posterior predictive checks are based on a discrepancy measure for which we compare the observed value to the one obtained under the predictive distribution based on the fitted model. As pointed out by many authors, implicit in the choice of this discrepancy measure is the idea of a wider model of which the one in question represents a restriction. We do not consider posterior predictive checks in this paper as our focus is on evaluating how the posterior distribution of parameters of wider models can be used to detect departures from normality.

8 Concluding remarks

This paper was originally motivated by the intention to revise critically the opinion, common among small area practitioners, that failures of the normality assumption for random effects in Fay-Herriot type models is very difficult to detect unless deviations from normality are huge; an opinion well supported by the findings of Sinharay and Stern (2003). We considered two different distributions for the random effects, both encompassing normality as a special case and compare the models based on this assumptions with the standard Normal-Normal Fay-Herriot model in different situations.
Our findings are largely consistent with those of Sinharary and Stern (2003) as far as the efficient estimation of individual small area parameters is concerned. We also considered the problem of estimating the Empirical Distribution Function of the 'ensemble' of small area parameters, a problem that is more likely to

be sensitive to the failure of distributional assumptions. As expected, we find the alternative models to be better than the Normal-Normal one for this problem when normality fails. Moreover we found that the Normal-EP and Normal-SEP may represent viable alternative to the standard Fay-Herriot model in general applications even when normality holds, provided that the number of areas involved in estimation is not too small.

The Normal-EP model is a special case of the Normal-SEP that encompasses the case of skewed residuals and is shown to perform almost equally well under normality and other distributions in the EP family. By the way, we treated the Normal-EP model in detail because of its nice properties: the parameters are easily interpreted (location and scale parameter coincide with expected value and variance) and when $\varphi > 0.5$, the EP distribution may be interpreted as a scale mixture of normals, which makes this model appealing as an alternative in presence of heteroskedastic random effects. Both these properties do not hold for the Normal-SEP model. We feel that especially the lack of direct interpretation for the location and scale parameters may represent an obstacle for practitioners. In general the SEP distribution is more complex, its properties less well known and the task of eliciting priors for its parameters difficult.

In conclusion we think that both the Normal-EP and Normal-SEP models deserve further investigation. In this paper we just considered an Hierarchical Bayes approach based on MCMC, which has many advantages but do not yield point estimators of closed and interpretable form. The analysis of these models adopting an Empirical Bayes or Frequentist approach may represent an interesting topic of future research.

A. BUGS code for the Normal-EPD Fay-Herriot model

```
model; {
for ( i in 1:m){
y[i]~dnorm(theta[i],precs[i])
theta[i]<-mu+v[i] }
# trick for specifying EP priors
for (i in 1:m){
zero[i]<-0
v[i]~dunif(-10000,10000)
phi[i]<- -(log(c1)-log(sigmav)-pow(abs(sqrt(c0)*(v[i]-0)/sigmav),1/psi))
zero[i]~dpois(phi[i])}
c0<-exp(loggam(3*psi))/exp(loggam(psi))
c1<-sqrt(c0)/(2*psi*exp(loggam(psi)))
# end of trick
psi~dunif(0,1)
sigmav~dunif(0,K)
sig2v<-pow(sigma,2)
mu~dnorm(0,1/L)
}
```

B. BUGS code for the Normal-SEPD Fay-Herriot model

```
model; {
for (i in 1:m){ y[i] dnorm(theta[i],precs[i])
theta[i]<-v[i] }
# trick for specifying SEP priors for (i in 1:m){ z[i]<-(v[i]-mu)/sigma
sign.z[i]<-z[i]/abs(z[i])
tau[i]<-sign.z[i]*pow(abs(z[i]),alpha/2)*lambda*sqrt(2/alpha)
zero[i]<-0
v[i]~dunif(-10000,10000)
phi[i]<-(-1)*(-(1/alpha-1)*log(alpha)-loggam(1/alpha)-
-log(sigma)+log(phi(tau[i]))-pow(abs(z[i]),alpha)/alpha)
zero[i]~dpois(phi[i]) }
# end of trick sigma.n~dunif(0,K)
mu~dnorm(0,1/L)
psi~dunif(0,1)
lambda ~ dnorm(0,.25)
alpha<-1/psi
sigma<-sigma.n*pow(psi,psi)/sqrt(c0)
c0<-exp(loggam(3*psi))/exp(loggam(psi))
c1<-sqrt(c0)/(2*psi*exp(loggam(psi)))
}
```


Acknowledgements

The authors would like to thank prof. Partha Lahiri, University of Maryland, for the informal discussion that led to this research. The work of Enrico Fabrizi was partially supported by the grants 60FABR06 and 60BIF06, University of Bergamo. The work of Carlo Trivisano was partially supported by the grand n. 2004137478.001, PRIN 2004.

References

Anderson, T.W., Darling, D.A., 1954. A test of goodness of fit. J. Amer. Statist. Assoc. 49, 765-769.

Azzalini, A., 1986. Further Results on a Class of Distributions which Includes the Normal One. Statistica 46, 171-178.

Andrews, D.F., Mallows, C.L., 1974. Scale Mixtures of Normal Distributions. J. Roy. Statist. Soc. B 36, 99-102.

Berger, J.O., 1994. An overview of robust Bayesian analysis (with discussion). Test 3, 5-126.

Box, G.E.P., Tiao, G.C., 1973. Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading MS.

Choy, S.T.B., Walker, S.G., 2003. The Extended Exponential Power Distribution and Bayesian Robustness. Stat. Probabil. Lett. 63, 227-232.

Choy, S.T.B., Chan, C.M., 2003. Scale Mixture Distributions in Insurance Applications. ASTIN Bullettin 33, 93-104.
Datta, G.S., Lahiri, P., 1995. Robust Hierarchical Bayes Estimation of Small Area Characteristics in the Presence of Covariates and Outliers. J. Multivariate Anal. 54, 310-328.
Datta, G.S., Rao, J.N.K., Smith, D.D., 2002. On measures of uncertainty of small area estimators in the Fay-Herriot model. Technical Report, University of Georgia, Athens.
Di Ciccio, T.J., Monti, A.C., 2004. Inferential Aspects of the Skew Exponential Power Distribution. J. Amer. Statist. Assoc. 99, 439-450.
Efron, B., Morris, C., 1971. Limiting the risk of Bayes estimators - Part I. The Bayes case. J. Amer. Statist. Assoc. 66, 807-815.
Fabrizi, E., Ferrante, M.R., Pacei, S., 2005. Estimation of poverty indicators at sub-national level using multivariate small area models. Statistics in Transition 7, 587-608.
Fabrizi, E., Ferrante, M.R., Pacei, S., 2007. A comparison of adjusted Bayes estimators of use in Small Area Estimation. Techical Report, DMSIA, University of Bergamo, 2007 n. 1.
Fay, R.A., Herriot, R.E., 1979. Estimation of income for small places: an application of James-Stein procedures to Census Data. J. Amer. Statist. Assoc., 78, 879-884.
Gelman, A., 2006. Prior Distributions for Variance Parameters in Hierarchical Models. Bayesian Analysis 1, 515-533.
Ghosh, M., 1992. Constrained Bayes estimation with applications, J. Amer. Statist. Assoc., 87, 533-540.
Ghosh, M., Maiti, T., 1999. Adjusted Bayes estimators with applications to Small Area Estimation. Sankhya B 61, 71-90.
Heady, P., Ralphs, M., 2004. Some findings of the EURAREA project and their implications for statistical policy. Statistics in Transition 6, 641-653.
Louis, T.A., 1984. Estimating a Population of Parameters Values Using Bayes and Empirical Bayes Methods. J. Amer. Statist. Assoc. 79, 393-398.
Rubin, D.B., 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician. The Annals of Statistics 12, 1151-1172.
Sinharay, S., Stern, H. S., 2003. Posterior Predictive Model Checking in Hierarchical Models. J. Stat. Plan. Infer. 111, 209-221.
Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003. WinBUGS User Manual Version 1.4 , downlodable at `http://www.mrc-bsu.cam.ac.uk/bugs`.
Thomas, A., O Hara, B., Ligges, U., Sturz, S., 2006. Making BUGS Open. **R** News 6, 12-17.
Walker, S.G., Gutierrez-Peña, E., 1999. Robustifying Bayesian Procedures. In: Bernardo, J.M., Berger, J.O., David, A.P., Smith A.M.F. (Eds.), Bayesian Statistics 6, 685-710.
West, M., 1987. On Scale Mixtures of Normal Distributions. Biometrika 74, 646-648.
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