

A Comparison of model-based methods for Small Area Estimation

N. Best¹, S. Richardson¹, P. Clarke² and V. Gómez-Rubio¹

¹Department of Epidemiology and Public Health, Imperial College London
St. Mary's Campus, Norfolk Place, W2 1PG London - United Kingdom

² Office for National Statistics, United Kingdom

October 20, 2008

Abstract

Government agencies often provide small area estimates that rely on available data and some underlying model that helps to provide estimates in all areas, even in those that were not sampled. Several models have been well-established for the study of data coming from small areas.

In this paper we have made a comparison of some of these methods paying attention to how different types of data sets can be employed efficiently and how to deal with the estimation in areas for which no direct individual data area available. We have considered design-based, regression and EBLUP estimators, which have been fitted using both a likelihood-based and a Bayesian approach. Spatial correlation among areas has also been considered. As in any study of the performance of different models, we have also compared different criteria for model comparison and selection.

1 Introduction

Small Area Estimation methods tackle the problem of providing feasible estimates of the variable of interest in areas where data available is not large enough to provide direct estimates of adequate precision (Rao, 2003; EURAREA Consortium, 2004; Jiang and Lahiri, 2006). Traditionally, two approaches have been followed to provide estimates in small areas: direct estimation and indirect or model-based estimation. Direct estimation relies

on estimates that are computed directly from a survey sample. This implies that for all areas a suitable sample must be available. Depending on the type of study it can be impossible to sample from every area we are interested in.

Rao (2003) also distinguish between area level (Type A) and unit level (Type B) estimators, depending on the level of aggregation of the data. In general, the estimation procedure is similar in both cases, but usually Type B models involve larger matrices which can lead to computational problems. When working with Type A models the aim is to fill the gaps in the areas that have not been sampled but for which some auxiliary covariates have been measured. Similarly, in Type B Models we have some auxiliary information available (i.e., covariates) of all the individuals in all areas. Using the values of the target variable and covariates of the individuals in the sample, the parameters of the model are estimated and the individual values of the target variable are imputed, so that the area estimates are built on them.

Model-based estimation relies on samples from a limited number of areas and a set of auxiliary covariates which are used to obtain estimates in the areas that were not actually sampled. This is also the reason why they are called indirect estimators, because they do not produce an estimate in a region based entirely on the population sampled in that region. The idea behind this techniques is to “borrow” strength in different ways from data available (sample and covariates). Therefore, this comprises a wide range of methods according to how the estimates are built.

Multilevel models (Goldstein, 2003) have played an important role in the development of small area estimators because of its flexibility and its capability to consider different types of effects, specially when these can be grouped into different clusters or layers. Most small area estimators are actually based on an underlying multilevel model.

Hierarchical Bayesian Models (Green *et al.*, 2003) offer a similar approach to the modeling of multilevel models within the Bayesian paradigm. Estimation and inference can be done numerically by means of Markov Chain Monte Carlo techniques (Gilks *et al.*, 1995).

Most data used in Small Area Estimation comes from surveys carried out in some of the areas of interest. This means that in most cases we lack information for many of the areas of study and that additional information will be needed in order to provide small area estimates. This additional data can often be obtained as area level totals or means of the covariates included in the model. Given that different methods can be used to handle these missing data, we will describe different approaches and compare the results obtained. Longford (2005) provides a comprehensive approach to some methods for missing data but in our analysis we will include some models not discussed therein.

This paper is organised as follows. Chapter 2 introduces the standard models used in Small Area Estimation. Different types of models are described to deal with different structures, such as spatial models. Chapter 3 describes analogous Bayesian models for Small Area Estimation. Chapter 4 is about different measures of performance and model selection. Chapter 5 discusses the use of these methods on two real data sets. Finally, Chapter 6 provides a discussion on the contents of this paper.

2 Methods for Small Area Estimation

{sec:SAE}

The information needed to produce small area estimates usually comes from different sources. Let us suppose that we want to produce estimates for k areas of a target variable of interest, which may represent the average or total value of the target variable of the individuals within areas. We will represent the target variable for area i by Y_i (area total) or \bar{Y}_i (area mean). In addition, different auxiliary covariates will be available for each area. As it happens with the target variables, the covariates will express the total (X_i) or average value (\bar{X}_i) of the covariates at the individual level.

A survey can be conducted to obtain a sample of the target variable from individuals within each area (Cochran, 1977; Lehtonen and Pahkinen, 2003). This will be denoted by y_{ij} , $i = 1, \dots, n$ and $j = 1, \dots, n_i$, being n_i the sample size obtained in area i . If auxiliary covariates are also available at the individual level, they will be denoted by x_{ij} . N_i is the number of individuals or units in area i .

The auxiliary covariates are always assumed to be known at the appropriate level (which will depend on the type of estimator used), whereas the target variable may be unknown for some areas if individuals from those areas have not been sampled. Usually, the areas where the sample is going to be taken from are defined in the survey design.

More details on these issues, sampling design and estimation can be found in Cochran (1977) and Särndal *et al.* (1992). The former points out the key issues present in the design of a survey and summarises different ways of conducting the sampling. Different direct and model based estimators are discussed as well.

Throughout the next sections we will describe different estimators of the variable of interest \bar{Y}_i , defined as

$$\bar{Y}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij} \quad (1) \quad \{\text{eq:total}\}$$

where N_i is the number of individuals in area i and y_{ij} the values of the

variable of interest for each individual in area i . Note that the area level total can be estimated by multiplying equation (1) by the population size N_i .

To clarify the terms used, a unit within an area can be either an individual, a household or another type of “entity” depending on the study and target variable. For example, the mean income is usually studied per household, so that within each area we will consider households as units.

2.1 Direct estimators and Sample-based methods

A convenient direct estimator is the Horvitz-Thompson estimator (Cochran, 1977), which can be defined as

$$\widehat{Y}_{D,i} = \frac{1}{N_i} \sum_{j \in s_i} w_{ij} y_{ij} \quad (2)$$

Here s_i represents the set of individuals that have been sampled in area i . Note that w_{ij} are weights that depend on how the sampling has been designed. Intuitively, w_{ij} can be thought of the amount of units in the area “represented” by each observation y_{ij} . The weights can be taken as the inverse of the probability of being sampled ($w_{ij} = 1/\pi_{ij}$), which only depends on the survey design (Särndal *et al.*, 1992). This estimator is also known as the π -estimator and it is an unbiased estimator of the mean in area i , given in equation (1).

Under random sampling without replacement, the variance of this direct estimator (known as *design variance*), is given by

$$\text{Var}[\widehat{Y}_{D,i}] = \frac{1}{n_i} \left(1 - \frac{n_i}{N_i}\right) \sigma_i^2$$

where σ_i^2 is the unit level variance in area i , which is often estimated using the empirical variance of the units y_{ij} . As a result, this variance can be used to provide approximate 95% confidence intervals using a Normal distribution.

2.2 Indirect estimators and Model-based methods

2.2.1 Synthetic estimator

When the variable that we are measuring is continuous, a linear model can be used to establish a relationship between the target variable and the covariates. Fay and Herriot (1979) propose the following model to combine direct estimation and regression analysis:

{:synthmodel}

$$\bar{Y}_i = \beta \bar{X}_i + \epsilon_i; \epsilon_i \sim N(0, V_i^2) \quad (3)$$

where V_i^2 is the known design variance. The synthetic estimator can be obtained by using the estimate of β from the model (3) and taking

$$\widehat{Y}_{SA,i} = \widehat{\beta} \bar{X}_i \quad (4) \quad \{\text{eq:synthetic}\}$$

as the *synthetic estimate* in area i . As mentioned before, the covariates are available for all small areas. This procedure can be extended to non-normal responses by means of a generalised mixed model in which the response and the linear predictor are linked by an appropriate function.

The estimation of the variance can be done as follows

$$\text{var}[\widehat{Y}_{SA,i}] = E[(\widehat{Y}_{SA,i} - \beta \bar{X}_i - \epsilon_i)^2] = \bar{X}_i^T \text{Var}[\widehat{\beta}] \bar{X}_i + V_i^2$$

The Type B formulation of this estimator involves the samples available for each area and the underlying model is as follows

$$y_{ij} = x_{ij}\beta + e_{ij}$$

where the individual errors e_{ij} are Normally distributed with zero mean and variance σ_i^2 , i.e., we assume that the with-in area variation can be different from area to area.

If the ratios n_i/N_i are very small the finite population correction can be ignored and the Type B synthetic estimator becomes

$$\widehat{Y}_{SB,i} = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij} \widehat{\beta} = \bar{X}_i \widehat{\beta} \quad (5) \quad \{\text{eq:syntheticB}\}$$

However, this relation does not hold for non-Normal or non-linear models.

This type of synthetic estimator has been used extensively by the Office for National Statistics in the United Kingdom to produce different small area reports. Yar *et al.* (2002) and Heady *et al.* (2003) rely on the Type B estimator but, given that no covariates are available at the individual level, they consider $x_{ij} = X_i$.

The variance of this estimator is

$$\text{Var}[\widehat{Y}_i] = \bar{X}_i^T \text{Var}[\widehat{\beta}] \bar{X}_i + \sigma_i^2/N_i$$

Notice also that, whilst the Type A synthetic estimator already provides the estimator for the area, the Type B estimator gives the values of the variable of interest for the individuals that have not been sampled ($\hat{y}_{ij} = \widehat{\beta} x_{ij}, \forall j$ in area i) and that areal estimates must be computed afterwards by averaging over all the individuals in that area.

2.3 General Linear Mixed Models

2.3.1 Area Level Mixed-Effects Models

Mixed-effects Type A models can be fully represented using the following expression:

$$\{\text{eq:GLMMA}\} \quad \hat{Y}_D = \bar{X}\beta + Zu + \varepsilon \quad (6)$$

where \hat{Y}_D is a vector of direct estimators, Z describes the structure of the area random effects u and ε is a random error term. In principle, we will assume that Z is the identity matrix, but a more complex structure can be used when the value in an area depends on several random effects. The random effects u_i are supposed to be independent and Normally distributed with mean 0 and variance σ_u^2 . The sampling error or individual variation ε are considered to be independent with zero mean and known variance $\text{diag}(\hat{\sigma}_i^2)$. We will denote by D and V the variance matrices of u and ε , respectively. Unconditioning on the random effects, the variance of \hat{Y}_D is $G = Z'DZ + V$.

Given that the variance is known, the parameter β can be estimated by means of Generalised Least Squares (GLS), which in this case is equivalent to Maximum Likelihood. The random effects can be estimated in several ways, as described below. See Section 2.3.3 for a discussion on this issue. Hence, the estimator is

$$\{\text{eq:EBLUPA}\} \quad \hat{Y}_{D,i} = \bar{X}_i\hat{\beta} + Z\hat{u} \quad (7)$$

When the direct estimates are missing for some areas but we have area-level covariates, the value of \hat{u} must be set to 0 and a synthetic-type estimator can be computed:

$$\hat{Y}_{D,i} = \bar{X}_i\hat{\beta}$$

If the random effects are given a structure (see, for example Section 2.3.4 below) the previous estimator can be corrected to account for the estimates of the random effects.

2.3.2 Unit Level Mixed-Effects Models

For Type B models we have a similar formulation:

$$\{\text{eq:GLMMB}\} \quad y_{ij} = x_{ij}\beta + \sum_l z_{li}u_l + e_{ij} \quad (8)$$

where e_{ij} is the random variation (or error) of individual or unit j in area i .

These models can also be expressed by splitting the data into the sampled (y_1) and non-sampled (y_2) units:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \beta + \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} u + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \quad (9) \quad \{\text{eq:GLMMB2}\}$$

Note that the length of y_1 and y_2 are $\sum_{i=1}^n n_i$ and $\sum_{i=1}^n (N_i - n_i)$, respectively. All the other matrices have sizes according to that.

The estimation of the model is based on the equation of the observed data

$$y_1 = x_1 \beta + Z_1 u + e_1$$

and different techniques can be used to provide estimates of the parameters of the model. The estimation of the parameter β can be done by Maximum Likelihood (ML) or Penalised Quasi-Likelihood (PQL), while Restricted Maximum Likelihood (REML) can also be used to obtain an unbiased estimate of the variance. McCulloch and Searle (2001) describe in detail these and other strategies to obtain estimates of the different parameters. Furthermore, EURAREA Consortium (2004) provide detailed algorithms and computational tricks for the estimation of the parameters for different types of mixed models.

Once the model has been fit using the observed data (computing estimates for the parameters of the model and the random effects; see Section 2.3.3 below), estimates can be provided for every unit in the population. Then, the small area estimate is obtained by summing over observed and estimated values in an appropriate way:

$$\hat{Y}_i = \frac{n_i}{N_i} \sum_{j=1}^{n_i} y_{ij} + \frac{N_i - n_i}{N_i} \sum_{j=1}^{N_i - n_i} \hat{y}_{ij} = \frac{n_i}{N_i} \sum_{j=1}^{n_i} y_{ij} + \frac{N_i - n_i}{N_i} \sum_{j=1}^{N_i - n_i} x_{ij} \hat{\beta} + \hat{u}_i \quad (10) \quad \{\text{eq:BLUPB}\}$$

where \hat{y}_{ij} is the estimate for unit j in area i .

When the sample size n_i is very small compared to N_i the following estimator is employed:

$$\hat{Y}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij} \hat{\beta} + \hat{u}_i = \bar{X}_i \hat{\beta} + \hat{u}_i \quad (11) \quad \{\text{eq:BLUPB2}\}$$

2.3.3 Best Linear Unbiased Predictor (BLUP)

The way the random effects are estimated differs from that of the fixed effects. A *Best Unbiased Predictor* (McCulloch and Searle, 2001) or BUP is obtained

{subsub:eb1up}

by using the expectation of the random effects given the observed data. It is called *best* in the sense that it minimises the Mean Square Error. When the BUP is linear on the observed data it is called Best Linear Unbiased Predictor (BLUP, Lee and Nelder, 1996). For the model that we have presented before, the estimator is as follows:

$$\{\text{eq:randeff}\} \quad E[u|y] = \hat{u} = DZ_1'\Sigma_{11}^{-1}(y - X_1\beta) \quad (12)$$

When the random effects are estimation in this way, equation (10) becomes a BLUP estimator and is the sum of a direct estimator based on the sample in area i , the fixed part based on the covariates and a estimate of the random effect. Note that if there is no sample available in area i , the BLUP estimator reduces to the synthetic estimator because there is no direct estimate and $\hat{u} = 0$.

When the variance of the random effects is not known and they are estimated from the data, similar estimators can be developed. These are called *Empirical Best Linear Unbiased Predictors* (EBLUP). Rao (2003) discuss the use of BLUP and EBLUP estimators in the context of small area estimation.

For Type A models the estimation of the Mean Square Error of the EBLUP estimator is often computed by approximating it by three components that reflect the uncertainty about the estimator (G_1), the estimation of β (G_2) and the estimation of σ_u^2 (G_3). This is often expressed as follows

$$\{\text{eq:MSE}\} \quad MSE[\hat{Y}_i] \approx G_1 + G_2 + 2G_3 \quad (13)$$

The MSE for Type B models is done analogously. In both cases, see Rao (2003, Chapter 7) for details.

2.3.4 Linear mixed model with spatial correlation

subsub:mixedspatial}

In principle, spatial autocorrelation can be introduced by considering a random effect for each area which is correlated with the random effects of its “neighbours”. Other possible spatial correlations can be used if appropriate.

Salvati (2004) has discussed these models and has computed Spatial EBLUP estimators for SAR and CAR models. These models assume that the area random effects are correlated, with the correlation decaying as the distance between the areas increases. These estimators have been used by Petrucci and Salvati (2004b,a); Pratesi and Salvati (2008) to study the average erosion per acre in the Rathburn lake watershed in Iowa (USA). They report how the Spatial EBLUP produce estimates which have a smaller mean square error than the usual EBLUP without spatial structure.

In the case of a SAR model the random effects are distributed according to a multivariate normal with mean 0 and variance

$$\sigma_u^2[(I - \rho W)(I - \rho W')]^{-1}$$

where ρ is the spatial dependence parameter and W the proximity matrix. W is often a binary matrix so that element $W(i, j)$ is 1 if areas i and j are neighbours and 0 otherwise. In addition, W is taken to be row-standardised to make sure that the values of ρ are between -1 and 1 (Haining, 2003).

Alternatively, for a CAR model the variance of the random effects is

$$\sigma_u^2(I - \rho W)^{-1}$$

Now W is often a binary matrix as before (not necessarily row-standardised). Note that in this case W must be symmetric and $(I - \rho W)$ strictly definite positive.

The EBLUP estimator described in Section 2.3.3 assumes that the random effects are uncorrelated, so that in the areas where no data have been observed, the value is $\hat{u}_i = 0$. The spatial variance-covariance structures induce correlation between the random effects so that it can be exploited to estimate the random effect in areas with no observed data.

Pratesi and Salvati (2008) provide computational details on how to estimate the Spatial EBLUP using a SAR specification for the random effects. In addition, they show how to estimate the MSE of the estimator. The expression of the MSE for the Spatial EBLUP is similar to that shown in equation (13). A new term G_4 is added to reflect the uncertainty about estimating ρ . See Pratesi and Salvati (2008) for full details.

2.3.5 Spatial model fitting with missing data

LeSage and Pace (2004) and Saei and Chambers (2005) show how to deal with missing data when the random effects are spatially correlated and that borrowing information across neighbouring areas improves the small area estimates. In this case, we will use the direct estimates and design variances from the areas with observed data to estimate the parameters of the model. Estimates in the areas with missing data will be provided as in equation (11), where \hat{u}_i is estimated using the fact that \hat{u} is distributed as a Multivariate Normal with mean 0 and variance-covariance matrix $\hat{\sigma}_u^2[(I - \hat{\rho}W)(I - \hat{\rho}W')]^{-1}$ (SAR specification) or $\hat{\sigma}_u^2[(I - \hat{\rho}W)]^{-1}$ (CAR specification).

3 Bayesian approach to Small Area Estimation

{sec:BSAE}

3.1 Spatial models

Ghosh and Rao (1994) describe different hierarchical Bayes approaches to Small Area Estimation. It is based on a previous work by Datta and Ghosh (1991) where they consider Type A and B models as specified in (6) and (8) to estimate the mean value of the variable of interest in each area with *uninformative* priors for β , σ_u^2 and σ_e^2 .

Among other results, they point out that, for Type A models, when a flat prior for β is used and σ_u^2 is known the posterior expectation of the mean value in area i matches the value provided by the BLUP estimator and its variance is equal to the MSE of this BLUP estimator as well.

An analogous model, which also includes a CAR specification for the spatial random effects, is shown in the equation below:

$$\begin{aligned}
 y_i | \beta, u_i, v_i, \sigma_u^2, \sigma_v^2, V_i &\sim N(X_i \beta + u_i + v_i, V_i) \\
 u_i | \sigma_u^2 &\sim N(0, \sigma_u^2) \\
 v_i | v_{-i}, \sigma_v^2 &\sim N\left(\frac{1}{nn_i} \sum_{j \sim i} v_j, \sigma_v^2 / nn_i\right) \\
 f(\beta) &\propto 1 \\
 \sigma_u^2 &\sim Ga^{-1}(a_0, b_0) \\
 \sigma_v^2 &\sim Ga^{-1}(a_0, b_0)
 \end{aligned} \tag{14}$$

Here nn_i is the number of neighbours of area i , V_i is the design variance, which is assumed to be known, and σ_u^2 is not known. The CAR specification used here assumes that W is binary and $\rho = 1$.

A similar Type B model can be written as follows:

$$\begin{aligned}
 y_{ij} | \beta, u_i, v_i, \sigma_u^2, \sigma_e^2 &\sim N(x_{ij} \beta + u_i + v_i, \sigma_e^2) \\
 \log(\sigma_i^2) | \sigma^2 &\sim N(0, \sigma^2) \\
 u_i | \sigma_u^2 &\sim N(0, \sigma_u^2) \\
 v_i | v_{-i}, \sigma_v^2 &\sim N\left(\frac{1}{nn_i} \sum_{j \sim i} v_j, \sigma_v^2 / nn_i\right) \\
 f(\beta) &\propto 1 \\
 \sigma_u^2 &\sim Ga^{-1}(a_0, b_0) \\
 \sigma_v^2 &\sim Ga^{-1}(a_0, b_0) \\
 \sigma^2 &\sim Ga^{-1}(a_0, b_0)
 \end{aligned} \tag{15}$$

Note that now we have included a hierarchical structure on the area level variances. Different effects can be considered by adding more terms to the model or defining a different structure for the random effects. These types of models will be discussed in the sections below.

3.2 Missing data and Bayesian regional models

In this work, Bayesian model fitting is done by simulation using Gibbs Sampling (and the WinBUGS software). The procedure is based on sampling from the full conditionals so that after a suitable burn-in period the simulations can be regarded as samples from the full posterior. In this framework, sampling from the spatial random effects in the area involves the values of these effects in the neighbours and this can be problematic if too many areas are missing, as discussed by Gómez-Rubio *et al.* (2008).

Basically, the full conditional of the spatial random effect in an area with observed data will depend on these data plus the values of the random effects at the neighbours. In the areas with missing data, this full conditional will only depend on the values of the random effect at the neighbours. If too many of them are missing the fitting procedure may not be very efficient.

For this reason, Gómez-Rubio *et al.* (2008) propose the use of models that borrow information at a higher administrative level. For very sparse survey data these models seem to perform better than spatial models that borrow information at a lower administrative level. For example, random effects v_i are replaced by $v_{k(i)}$ in area level model

$$\begin{aligned}\widehat{Y}_{D,i}|\beta, u_i, v_{k(i)}, \hat{\sigma}_i^2 &\sim N(\overline{X}_i\beta + u_i + v_{k(i)}, \hat{\sigma}_i^2) \\ v_{k(i)}|v_{-k(i)}, \sigma_v^2 &\sim N\left(\frac{1}{nn_{k(i)}} \sum_{j\sim k(i)} v_j, \sigma_v^2/nn_{k(i)}\right)\end{aligned}$$

Here $k(i)$ represents the higher administrative level where area i is included and the adjacency structure of the CAR specification is now referred to this higher geographical level.

4 Measures of performance and variation

{sec:modelsec}

4.1 Mean Square Error, intervals and coverage

A common criterion to compare the performance of models when the actual true area level values are known is the Empirical Mean Square Error. If several surveys are available (for example, in a simulation study), the Average Empirical Mean Square Error (AEMSE) can be computed to estimate the overall departure from the actual values:

$$\text{AEMSE} = \sum_{s=1}^m \sum_{i=1}^k (\widehat{Y}_i^{(s)} - \overline{Y}_i)^2$$

where $\widehat{Y}_i^{(s)}$ is the small area estimate for area i using data set s , k is the number of small areas and m the number of different survey samples. The

AEMSE also provides a measure of the variation of the small area estimates around the actual values.

Similarly, credible and confidence intervals can be computed. Credible intervals can be computed for the Bayesian small area estimates by exploiting the output from the MCMC simulations, whilst approximate confidence intervals can be computed using the likelihood-based small area estimates and their estimated variances using a Normal approximation. These intervals can be used in simulation studies to assess how many of the true area values are covered. The coverage will be another criterion that we will employ to assess the performance of the models.

4.2 Selection of mixed-effects models

When different sensible models are available to modelise the same phenomenon we may be interested in choosing the “best” model. We follow the approach given by Spiegelhalter *et al.* (2002) in the sense that there is no true model but some are useful. Traditional criteria based on information theory (see, for example, Akaike, 1973; Burnham and Anderson, 2002) are not designed to deal with random effects and are more focused on the selection of the covariates available in the model. We will not consider the selection of the covariates and we will assume that the same relevant covariates are included in all our models proposed. We will focus on the selection of the best structure of the model by considering measures that account for the complexity of the random effects as well.

Vaida and Blanchard (2005) note the importance of distinguishing between marginal and conditional likelihood in a mixed-effects model. If the interest is only in the fixed part (what they call *population focus*, i.e., we do not care about the random effect and integrate them out) then marginal likelihood (by integrating the random effects out) can be used in conjunction with the AIC to select the best model. Vaida and Blanchard (2005) call this criteria *mAIC*.

For example, in the simple mixed-effects model

$$y_i = X_i\beta + r_i, \quad r_i = Z_i u_i + \varepsilon_i \sim N(0, Z_i D Z_i' + \sigma^2 I_{n_i})$$

this criterion is defined as

$$mAIC = -2 \log g(y|\hat{\beta}) + 2K$$

where $g(y|\hat{\beta})$ is the likelihood of the model integrated over the random effects and evaluated at the MLE of β . K is the number of parameters in the fixed mean and the variance components.

A distinction is made when the random effects are also of interest (what they call *cluster focus*) and in this case they propose using a conditional likelihood to develop a *conditional Akaike Information Criterion* (cAIC):

$$cAIC = -2 \log g(y|\hat{\beta}, \hat{u}) + 2K$$

Here $g(y|\hat{\beta}, \hat{u})$ is the likelihood of the model conditional on the estimated random effects \hat{u} and evaluated at the MLE of β . K is the “effective degrees of freedom” which is related to the trace of the hat matrix and will be corrected depending on whether ML or REML is used and the true random effects variance (particularly, the parameter σ^2) is known. See Vaida and Blanchard (2005) for details.

The cAIC can be used to compare different models which have different random effects structure. For Bayesian models, the Deviance Information Criterion (DIC, Spiegelhalter *et al.*, 2002) can be used. The DIC is defined as

$$DIC = D(\hat{\theta}) + 2p_D$$

where $D(\hat{\theta})$ is the deviance of the model evaluated at the parameter estimates $\hat{\theta}$ (usually the posterior means) and p_D is the effective number of parameters, whose computation for these particular family of mixed-effects models is described in Appendix D.

The cAIC and the DIC provide similar results when a flat prior is used for a mixed-effects model with known variances (D. Spiegelhalter, personal communication). See Appendix D for a proof.

5 Examples

{sec:examples}

5.1 Equivalised Income per Household in Sweden

The LOUISE Population Register in Sweden is a data base that records information at the household level for every municipality in the country. The collection is exhaustive and all households in the country are included in this register. Hence, if we mimic a survey by taking some households at random and we compute some small area estimates, the LOUISE register provides the seldom available true area value to compare to our estimates. This situation is very similar to a simulation study, but with real data.

We have considered the case of a fictitious survey carried out in 100 of the 284 municipalities of Sweden, covering 0.1% of the total number of households (with a minimum sample size per area of 5) in the country. The sample sizes

in the areas range from 5 to 545 households. The areas selected, the same in all 20 surveys, have been chosen at random using stratified sampling according to the average of employed people per household and the proportion of head of households with higher education. Furthermore, five covariates have been considered: number of people employed in the household, number of people living in the household, and age, gender and higher education of head of household. This survey design is the same as in Gómez-Rubio *et al.* (2008), but the survey sample data are different.

Table 1 shows the results obtained on the LOUISE data with the methods described in this paper. Direct comparison between the different models can be done by checking the AEMSE. In addition, another criterion based on the AIC has been computed.

Clearly, area level models seems to perform better. This may be due to the fact that they are more robust to the presence of anomalous observations in the data given that these models work with aggregate data. This has also been reported in Gómez-Rubio *et al.* (2008). A possible solution is to use a Student's t instead of a Normal distribution to model the response ****REF****.

Method	AEMSE	Coverage	AIC/cAIC/DIC/	K/ p_D
Direct	28457.6	0.90	—	—
Area Synthetic	4901.9	0.99	1297.88	7
Unit Synthetic	8941.7	0.92	49565.11	7
Area EBLUP	4315.6	0.91	1244.20	31.32
Area Spatial EBLUP	4980.1	0.69	1134.03	8.52
Unit EBLUP	7289.3		AIC:49507.54	8
Unit Spatial EBLUP				
Bay. area u_i	4324.3	0.71	1247.73	33.16
Bay. area v_i	4568.5	0.68	1266.96	15.21
Bay. area $u_i + v_i$	4287.7	0.74	1246.52	34.21
Bay. area regional	4183.6	0.79	1247.40	30.29
Bay. unit u_i	5989.3	0.45	47464.08	132.20
Bay. unit v_i	4556.3	0.84	47458.16	124.54
Bay. unit $u_i + v_i$	4665.9	0.82	47457.54	124.97
Bay. unit regional	4835.6	0.67	47458.92	125.82

{tab:LOUISE} Table 1: Small area estimates for the equalised income per household in Sweden.

Model selection based on the different Information Criteria should be handled with care, specially because we are considering different inferential

approaches. However, cAIC and DIC provide similar values for similar models. In particular, the Area level EBLUP and the Bayesian model with spatial random effects only have very similar values.

5.2 Income per Household in England and Wales

Our second case study is based on the Family Resources Survey (FRS) carried out in 2001 and 2002 by the British Department of Work and Pensions. The FRS recorded a wealth of information from many households in England and Wales. Primary sampling unit were postcode sectors, and the information has been made available to us linked to other different geographies, such as the Middle-layer Super-Output Areas (MSOAs) and Local Authority Districts (LADs). The total sample size is 22,859, which represents a sample size of around the 0.15% of the total number of households. Dhanecha *et al.* (2002) describe the FRS and **** ASK PHILIP ABOUT THIS REPORT **** provide an analysis using standard ONS methodology on likelihood-based models.

The target variable is the total weekly income per household and it is among the variables available in the survey. Another 21 covariates have been included in the model based on the analysis done in **** ASK PHILIP ABOUT THIS ****, which include different socio-economic variables. These covariates are not available at the household level but at the MSOA and LAD levels, the former being our main administrative level of interest. Furthermore, the same covariates for every LAD in England and Wales are available so that small area estimates can be provided.

Hence, in area i we can describe the data for the household j as

$$(y_{ij}, \mathbf{x}_{m(ij)}, \mathbf{x}_{l(ij)})$$

where $m(ij)$ is the index for the MSOA and $l(ij)$ for the LAD. Note that LADs are not nested within MSOAs. Furthermore, another table is available with the same area level covariates for all LADs, so that estimates can be computed for all LADs. However, we found that only a few LADs did not contain any household in the survey.

Another particular advantage of this data set is that given that every household has the LAD covariates attached, we can work with any transformation of the income. To reduce the skewness of the response, the logarithm of the income is often used. Note that this is a non-linear transformation and if we had household level covariates it would not be able to use the fitted model to provide small area estimates ****REF ON ECOL. BIAS****. Hence, the data set that we have used to fit this model is

$$(\log(y_{ij}), \mathbf{x}_{1(ij)})$$

which means that we ignore the MSOA level covariates. The model will provide an estimate of the log-average income per household.

5.2.1 Results

We have fit the Bayesian models described in Section 3 and we have computed the small area estimates of the average income per household in the areas. Note that now it is not possible to assess the goodness of the estimates by means of the AEMSE because we do not know the actual area level true income per household. However, we can attempt to select the best model with the cAIC and DIC.

Given that we lack direct estimates of the average income at the LAD level we are not able to fit area level models. Hence, we have restricted the analysis to unit level models.

Table 2 shows the results of the models computed for this data set. If we consider the unit level Bayesian models, the model with the lowest DIC is the one with non-spatial random effects only.

Model	AIC/cAIC/DIC	K/ p_D
Synthetic (unit)		
EBLUP (unit)		
SEBLUP (unit)		
Bayesian unit u_i	51494.900	363.760
Bayesian unit v_i	51502.100	353.597
Bayesian unit $u_i + v_i$	51502.200	377.413

{tab:FRSDIC} Table 2: Results of different estimators for the Income using data from the Family Resources Survey. Note that the AEMSE and the coverage rate cannot be computed because we do not know the true area level means.

5.2.2 Classification of areas

In addition to the small area estimates of the income we are interested in providing a classification of the areas according to their income. In particular, we are interested in those areas with very low income, which may be targeted by particular policies to increase their wealth.

The criterion for non-Bayesian models is mainly based on plotting the small area estimates and their associated intervals. Areas can be ranked

according to the small area estimates, and checking whether the intervals overlap may be used to separate areas of low and high income. Other methods for ranking areas in a likelihood-based framework include, for example, quantile regression (Chambers and Tzavidis, 2006).

For the Bayesian models a similar ranking can be provided based on the posterior estimates. Furthermore, we have followed Gómez-Rubio *et al.* (2008) and computed different Bayesian criteria to rank the areas. First of all, we can rank the areas according via the posterior rank means. Secondly, we have considered posterior probabilities based on assessing whether the small areas estimates are the most deprived area. Other interesting posterior probabilities include the probability of being among the 10% and 20% most deprived areas.

Some of these criteria are displayed in Figure 1. Areas have been sorted according to the posterior means of the ranks, shown in Figure 1(b). First of all, Figure 1(a) shows that it is difficult to separate low income areas from the rest based on the non-overlapping credible intervals. However, areas of very high income can be separated from the rest. Using posterior probabilities, as shown in Figures 1(c) and 1(d), seems a better approach, and areas can be ranked according to them. However, it is known that these probabilities depend on the actual survey data used in the analysis and we are not able to provide any measure of uncertainty about them. Finally, it is worth mentioning that all classification criteria tend to provide a similar ranking of the areas.

Although it may be difficult to identify the areas with the lowest income, if we consider the probabilities of, for example, being the most deprived area, we can see that many areas have zero probability. Hence, these areas can be excluded as being the most deprived.

6 Discussion

{sec:discussion}

In this paper we have compared some of the most relevant methods used in Small Area Estimation using likelihood-based and Bayesian inference.

7 Acknowledgements

We would like to thank Statistics Sweden for providing the LOUISE Register data set and ONS for access to the Family Resources Survey data. This research has been funded by the Economic and Social Research Council through its National Centre for Research Methods, contract/grant number

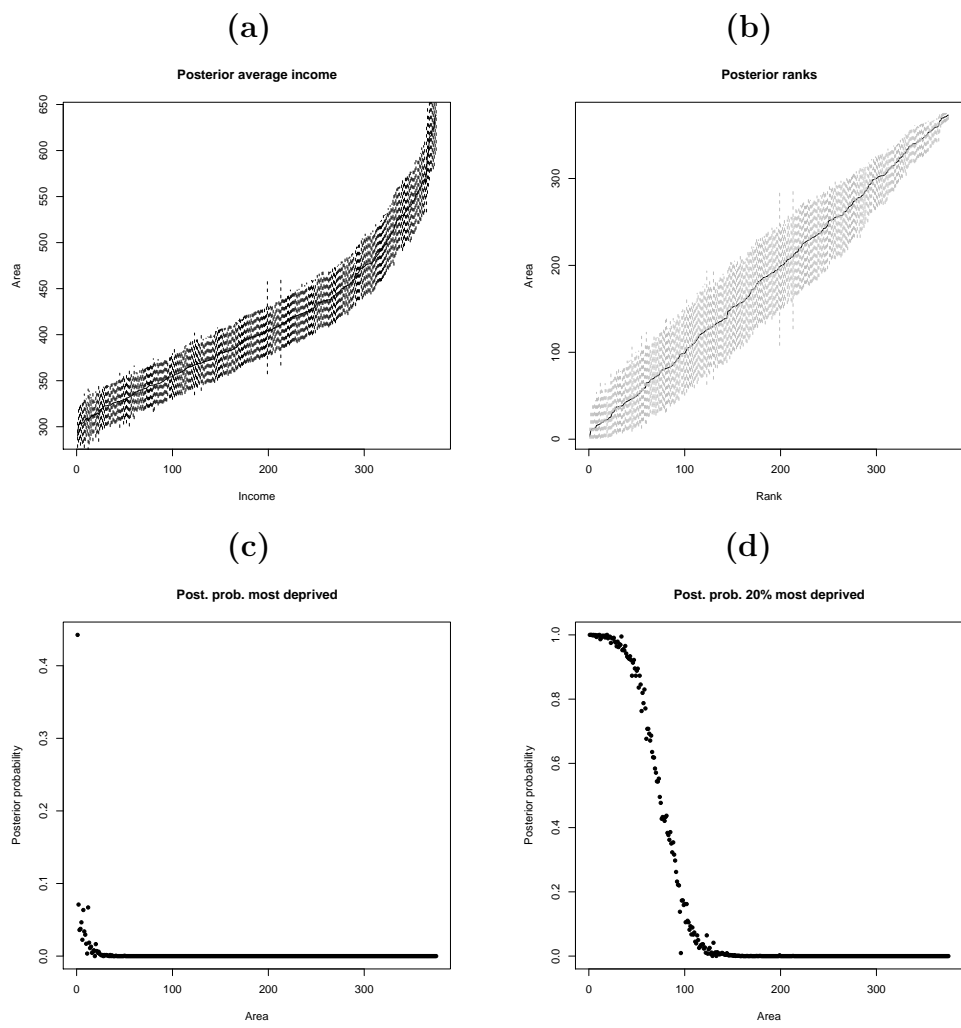


Figure 1: Classification criteria based on Bayesian Hierarchical Models. **Labels should read:**(a) x-label: Area, y-label: Weekly Income (pounds); (b) x-label: Posterior rank, y-label: Posterior distribution of rank; (c) x-label: Posterior rank, y-label: Posterior probability; (d) x-label: Posterior rank, y-label: Posterior probability

{fig:FRS}

R000239598.

References

- Akaike H (1973). “Information theory as an extension of the maximum likelihood principle.” In BN Petrov, F Csaki (eds.), “Second International Symposium on Information Theory,” pp. 267–281. Akademiai Kiado, Budapest.
- Burnham KP, Anderson DR (2002). *Model Selection and Multimodel Inference*. Springer-Verlag New York, Inc., 2nd edition.
- Chambers R, Tzavidis N (2006). “M-quantile Models for small area estimation.” *Biometrika*, (93).
- Cochran WG (1977). *Sampling Techniques*. John Wiley & Sons, Inc., New York, 3rd edition.
- Datta GS, Ghosh M (1991). “Bayesian Prediction in Linear Models: Applications to Small Area Estimation.” *The Annals of Statistics*, **19**(4), 1748–1770.
- Dhanecha N, Ellerd-Elliott S, Herring I, Horsfall E, Matejic P, Shome J, Snow J (2002). “Family Resources Survey 2001-2002.” *Technical report*, Department for Work and Pensions.
- EURAREA Consortium (2004). “Project Reference Volume.” *Technical report*, EURAREA Consortium.
- Fay RE, Herriot RA (1979). “Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data.” *Journal of the American Statistical Association*, **74**(366), 269–277.
- Ghosh M, Rao JNK (1994). “Small Area Estimation: An Appraisal.” *Statistical Science*, **9**(1), 55–76.
- Gilks WR, Richardson S, Spiegelhalter DJ (eds.) (1995). *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC, Boca Raton, Florida.
- Goldstein H (2003). *Multilevel Statistical Models, 3rd ed.* Arnold Publishers, London.
- Gómez-Rubio V, Best N, Richardson S, Clarke P (2008). “Bayesian Statistics for Small Area Estimation.” *Submitted for publication*.
- Green PJ, Hjort NL, Richardson S (eds.) (2003). *Highly Structured Stochastic Systems*. Oxford University Press Inc., New York.

- Haining RP (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge University Press, Cambridge.
- Heady P, Clarke P, others (2003). “Small Area Estimation Project Report.” *Technical report*, Office for National Statistics, United Kingdom.
- Hodges J, Sargent D (2001). “Counting degrees of freedom in hierarchical and other richly parameterized models.” *Biometrika*, **88**, 367–379.
- Jiang J, Lahiri P (2006). “Mixed Model Prediction and Small Area Estimation.” *Test*, **15**(1), 1–96.
- Lee Y, Nelder JA (1996). “Hierarchical Generalised Linear Models.” *Journal Royal Statistical Society B*, pp. 619–678.
- Lehtonen R, Pahkinen E (2003). *Practical Methods for Design and Analysis of Complex Surveys*. John Wiley & Sons Ltd., Chichester.
- LeSage JP, Pace RK (2004). “Models for Spatially Dependent Missing Data.” *Journal of Real Estate Finance and Economics*, **29**(2), 233–254.
- Longford NT (2005). *Missing Data and Small-Area Estimation. Modern Analytical Equipment for the Survey Statistician*. Springer Science+Business Media, Inc., New York.
- McCulloch CE, Searle SR (2001). *Generalized, Linear, and Mixed Models*. John Wiley & Sons, Inc., New York.
- Petrucci A, Pratesi M, Salvati N (2005). “Geographic Information in Small Area Estimation: Small Area Models and Spatially Correlated Random Area Effects.” *Statistics in Transition*, **3**(7), 609–623.
- Petrucci A, Salvati N (2004a). “Small Area Estimation considering spatially correlated errors: the unit level random effects model.” *Technical report*, University of Florence, Department of Statistics.
- Petrucci A, Salvati N (2004b). “Small Area Estimation Using Spatial Information. The Rathbun Lake Watershed Case Study.” *Technical report*, University of Florence, Department of Statistics.
- Petrucci A, Salvati N (2005). “Small Area Estimation for Spatial Correlation in Watershed Erosion Assessment.” *Journal of Agricultural, Biological & Environmental Statistics*, **11**(2), 169–182.

- Pratesi M, Salvati N (2008). “Small area estimation: the EBLUP estimator based on spatially correlated random area effects.” *Statistical Methods and Applications*, **17**, 113–141.
- Rao JNK (2003). *Small Area Estimation*. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Saei A, Chambers R (2005). “Working Paper M05/03: Empirical Best Linear Unbiased Prediction for Out of Sample Areas.” *Technical report*, Southampton Statistical Sciences Research Institute, University of Southampton.
- Salvati N (2004). “Small Area Estimation by Spatial Models: the Spatial Empirical Best Linear Unbiased Prediction (Spatial EBLUP).” *Technical report*, University of Florence, Department of Statistics.
- Särndal CE, Swensson B, Wretman J (1992). *Model Assisted Survey Sampling*. Springer-Verlag Inc., New York.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002). “Bayesian measures of model complexity and fit (with discussion).” *J. R. Statist. Soc. B*, **64**, 583–639.
- Vaida F, Blanchard S (2005). “Conditional Akaike information for mixed-effects models.” *Biometrika*, **92**(2), 351–370.
- Yar M, Hennell S, Clarke P, Meltzer H, Gatward R (2002). “Childhood Mental Disorder in England: Ward Estimates.” *Technical report*, Office for National Statistics, United Kingdom.

A cAIC for the Type A and B Models under heterocedasticity

Vaida and Blanchard (2005) describe the conditional Akaike Information Criterion (cAIC, henceforth) that can be used to compare different mixed-effects models of the form:

$$y = X\beta + Zu + \varepsilon \tag{16}$$

where y , X and Z are the response, covariates and structure of the random effects, respectively. This general form can be used to represent area and unit level models, as described later.

Based on the AIC, Vaida and Blanchard (2005) developed the cAIC, which can be considered an extension to the case of mixed-effect models. The form of the cAIC is

$$cAIC = -2 \log(f(y|\hat{\beta}, \hat{u})) + 2\eta$$

where the first term is the conditional deviance of the model and η is bias-correction penalty term that measures the effective number of parameters. This is taken as the trace of the hat matrix H_1 that maps y into their respective estimated values:

$$\{\text{eq:fitted}\} \quad \hat{y} = X\hat{\beta} + Z\hat{u} = H_1 y \quad (17)$$

Vaida and Blanchard (2005) basically follow the steps described by Hodges and Sargent (2001) to compute the hat matrix.

In this appendix we describe how to extend the computation of the cAIC to models where the variance of y is not homocedastic, that is, it is a general matrix Σ in the context of Small Area Estimation.

A.1 Computation of the penalty term

We will assume that the variance matrix of ε is a known (block) diagonal matrix. We can rescale the matrix by dividing by σ^2 , the maximum value in the (block) diagonal.

We will follow the indications given in Vaida and Blanchard (2005), App. 1 to derive the hat matrix. We express our model as

$$\{\text{eq:YAdelta}\} \quad Y = A\delta + e \quad (18)$$

where

$$Y = \begin{pmatrix} y \\ 0 \end{pmatrix}, \quad \delta = \begin{pmatrix} \beta \\ u \end{pmatrix}, \quad A = \begin{pmatrix} X & Z \\ 0 & -I_r \end{pmatrix}, \quad e = \begin{pmatrix} \varepsilon \\ u \end{pmatrix}$$

Note that the variance of e is now $\text{diag}(\Sigma, G)$, which can also be written as $\sigma^2 \text{diag}(\Sigma_0, G_0)$, with $G_0 = \frac{\sigma^2}{\sigma^2} G$. Given that G is symmetric and positive definite there exists a matrix Δ such as $G_0 = (\Delta^T \Delta)^{-1}$.

We define now $\Gamma = \text{diag}(I, \Delta)$ and we premultiply equation (18) by it to obtain

$$\Gamma Y = Y, \quad w = \Gamma e = \begin{pmatrix} \varepsilon \\ \Delta u \end{pmatrix}, \quad M = \Gamma A = \begin{pmatrix} X & Z \\ 0 & -\Delta \end{pmatrix}$$

We have now that $w \sim N(0, \sigma^2 V)$ where $V = \text{diag}(\Sigma_0, I_d)$. We can obtain a generalised least squares estimator of δ :

$$\begin{aligned}\hat{\delta} &= \begin{pmatrix} \hat{\beta} \\ \hat{u} \end{pmatrix} = (M^T V^{-1} M)^{-1} M^T V^{-1} V^{-1} Y = \\ &= (M^T V^{-1} M)^{-1} \begin{pmatrix} X^T \\ Z^T \end{pmatrix} \Sigma_0^{-1} y = \\ &= \begin{pmatrix} X^T \Sigma_0^{-1} X & X^T \Sigma_0^{-1} Z \\ Z^T \Sigma_0^{-1} X & Z^T \Sigma_0^{-1} Z + G_0^{-1} \end{pmatrix}^{-1} \begin{pmatrix} X^T \\ Z^T \end{pmatrix} \Sigma_0^{-1} y\end{aligned}$$

If we plug $\hat{\delta}$ into equation (17) the hat matrix becomes

$$H_1 = (X \ Z) \hat{\delta} = (X \ Z) \begin{pmatrix} X^T \Sigma_0^{-1} X & X^T \Sigma_0^{-1} Z \\ Z^T \Sigma_0^{-1} X & Z^T \Sigma_0^{-1} Z + G_0^{-1} \end{pmatrix}^{-1} \begin{pmatrix} X^T \\ Z^T \end{pmatrix} \Sigma_0^{-1} \quad (19) \quad \{\text{eq:hatmatrix}\}$$

Rearranging terms in equation (19) we get that

$$\begin{aligned}H_1 &= \begin{pmatrix} X^T \Sigma_0^{-1} X & X^T \Sigma_0^{-1} Z \\ Z^T \Sigma_0^{-1} X & Z^T \Sigma_0^{-1} Z + G_0^{-1} \end{pmatrix}^{-1} \begin{pmatrix} X^T \\ Z^T \end{pmatrix} \Sigma_0^{-1} (X \ Z) = \\ &= \begin{pmatrix} X^T \Sigma_0^{-1} X & X^T \Sigma_0^{-1} Z \\ Z^T \Sigma_0^{-1} X & Z^T \Sigma_0^{-1} Z + G_0^{-1} \end{pmatrix}^{-1} \begin{pmatrix} X^T \Sigma_0^{-1} X & X^T \Sigma_0^{-1} Z \\ Z^T \Sigma_0^{-1} X & Z^T \Sigma_0^{-1} Z \end{pmatrix} \quad (20) \quad \{\text{eq:eta}\}\end{aligned}$$

The previous equation involves a series of terms which appear twice and that only need to be computed once.

The penalty term is computed as $\eta = \text{tr}(H_1)$. Vaida and Blanchard (2005) note that H_1 is not exactly a hat matrix but the upper-left element of the hat matrix that maps $(y, 0)$ into Y .

It is worth noting that this approach can be considered for models with more complex variance-covariance structures and that the cAIC for more general mixed-effects models can be computed in a similar way.

B cAIC for Small Area Estimation

B.1 Area level models

Area level model can be expressed using the notation in equation (16) as follows. $y = (\hat{Y}_{D,1}, \dots, \hat{Y}_{D,k})$ is the vector of direct estimators, $X = [X_1, \dots, X_k]$

the matrix with area level covariates, and $Z = I_k$ is the identity matrix of dimension $k \times k$. Finally, $\Sigma = \text{diag}(\hat{V}_1^2, \dots, \hat{V}_k^2)$ is a diagonal matrix with the design variances. Hence, $\Sigma_0 = \frac{1}{\sigma^2}\Sigma$, where $\sigma^2 = \max_{i=1, \dots, k}(\hat{V}_i^2)$. Regarding G_0 , it is defined as $G_0 = \frac{\sigma_u^2}{\sigma^2}I_k$. Having defined the different terms needed to compute the penalty term, it is computed as in (20).

B.2 Unit level models

In unit level models, we consider the individual units sampled in the survey. The response of the model is given by the unit level values of the response $y = (y_1, \dots, y_k)$ where $y_i = y_{i1}, \dots, y_{in_i}$. X is the matrix with unit level covariates and Z is a block diagonal matrix where each block is a column of 1's of length n_i . Σ is also block diagonal, with each block equal to $\text{diag}(\sigma_i^2, \dots, \sigma_i^2)$ of dimension $n_i \times n_i$. Regarding G_0 , it is defined as $G_0 = \frac{\sigma_u^2}{\sigma^2}I_k$. Having defined the different terms needed to compute the penalty term, it is computed as in (20).

C cAIC for the Spatial EBLUP estimator

The Spatial EBLUP (Petrucci *et al.*, 2005; Petrucci and Salvati, 2005) is based on assuming a Spatial Autoregressive specification on the random effects. This can be expressed as that the area means are:

$$y = X\beta + Zv$$

where v is a set of random effects

$$v = \rho Wv + u$$

Here, W is an adjacency matrix (probably, row standardised) and u are multivariate Normal with zero mean and variance σ_u^2 . Hence, the Fay-Herriot estimator becomes

$$\{\text{eq:SEBLUP}\} \quad y = X\beta + Z(I_k - \rho W)^{-1}u \quad (21)$$

C.1 Area Level SEBLUP

From equation (21) it is possible to derive the computation of the penalty term in an easy way. We will take the same values of the vectors and matrices involved as in the Area Level EBLUP, but matrix G is defined in a different

way. In particular, $G = \sigma_u^2[(I_k - \rho W)(I_k - \rho W^T)]^{-1}$, so that $G_0 = \frac{\sigma_u^2}{\sigma^2}[(I_k - \rho W)(I_k - \rho W^T)]^{-1}$.

C.2 Unit Level SEBLUP

The computation is similarly done as in the case of the Unit Level EBLUP, with the difference that matrix G is defined in the same way as in the Area Level SEBLUP.

D Comparisson of cAIC and DIC

{app:IC}

The cAIC can be written as

$$cAIC = D(\theta) + 2\eta$$

$D(\theta)$ is the conditional deviance of the model evaluated at the parameters θ and values of the random effects, and η is the effective number of parameters of the model. Analogously, the DIC can be written down as

$$DIC = D(\theta) + 2p_D$$

where $D(\theta)$ is as before and p_D is computed as $E[D(\theta)] - D(E[\theta])$.

For the general area level mixed effects model shown in equation (16), Ghosh and Rao (1994) show that the EBLUP estimator fitted using Maximum Likelihood (ML) and the posterior expectation of the small area estimate are equal. Given that the area level variances are known, the value of the deviances in the cAIC and DIC are equivalent.

The value of η is given in equation (20), whilst the value of p_D for this particular model (assuming that σ_u^2 is known) is given in Spiegelhalter *et al.* (2002, Section 4.3, page 593) and it is

$$p_D = tr(L^*L^{-1}) \tag{22} \quad \{\text{eq:pD}\}$$

with

$$L^* = \begin{pmatrix} X^T \Sigma^{-1} X & X^T \Sigma^{-1} Z \\ Z^T \Sigma^{-1} X & Z^T \Sigma^{-1} Z \end{pmatrix}$$

$$L = \begin{pmatrix} X^T \Sigma^{-1} X & X^T \Sigma^{-1} Z \\ Z^T \Sigma^{-1} X & Z^T \Sigma^{-1} Z + G^{-1} \end{pmatrix}$$

Note that (22) is essentially the same expression as for η using the matrices in equation (20), which have been rescaled by dividing by σ^2 . Hence, for

this particular model when the variances of the response Σ and the random effects G are known the cAIC and the DIC are equivalent and they should provide the same values for the same model, regardless it is fit by ML or MCMC.

The case of the unit level models is slightly different because the area level variances $\sigma_i^2, i = 1, \dots, k$ need to be estimated. However, if they are assumed to be known, then the cAIC and the DIC are also equivalent. Similarly, if the ML and Bayesian estimates of these variances are similar, then the cAIC and DIC should be very similar.

VIRGILIO: It may be worth re-checking Ghosh and Rao to see the impact of estimating the variances on the computation of cAIC and DIC or keep the final comments as they are.