

Linear Mixed Model with Correlated Area and Time Effects in Small Area Estimation

Modelli lineari misti con correlazione degli effetti casuali di area e di tempo nella stima per piccole aree

Piero Demetrio Falorsi, Stefano Falorsi

Italian National Institute of Statistics; Via Magenta, 2, Rome, Italy;

e-mail: falorsi@istat.it, stfalors@istat.it.

Aldo Russo

University of Rome, Roma 3; Via Segre, 2, Rome, Italy;

e-mail: aldoruss@poli.uniroma3.it.

Riassunto: Generalmente, le indagini campionarie su larga scala hanno la finalità di produrre stime inerenti non solo all'intera popolazione oggetto di studio ma anche a sottopopolazioni di quest'ultima - usualmente denominate domini di studio o piccole aree- riferite ad aree geografiche, a gruppi socio-demografici oppure ad insiemi di imprese. In questa nota presentiamo una metodologia per l'ottenimento di stime riferite alle sottopopolazioni in oggetto basata sull'utilizzo di predittori definiti nell'alveo della teoria del modello lineare generale misto a componenti di varianza. Il modello di riferimento dei predittori qui proposti presenta l'importante requisito di considerare simultaneamente gli effetti di area e gli effetti di tempo in cui sia gli uni che gli altri sono caratterizzati da una struttura di correlazione.

Keywords: Parameter of interest, Predictors, Variance components, MSE.

1. Introduction

Sample surveys are widely used to provide estimates of population parameters (e.g. totals and means) for the total population under study but also for subpopulations such as geographic areas and socio-demographic groups. *Direct estimates*, generally *design-based*, of a subpopulation parameter are based only on the subpopulation specific sample data. However it is seldom possible to have a large enough overall sample size to have reliable direct estimates for all the subpopulations of interest. In our context, we denote with the term *small area* any subpopulation for which direct estimates of adequate precision cannot be provided. In order to produce reliable small area estimates, it is often necessary to use *indirect estimates* that *borrow strength* by using values of the variables of interest from related areas, thus increasing the *effective sample size*. These values are brought into the estimation method through *models* that provide a link to related subpopulations through the use of auxiliary data related to the variables of interest. These models are classified into two broad groups: (a) area level models that relate the small area means to area specific auxiliary variables; (b) unit level models that relate the unit values of study variables to unit specific auxiliary variables.

The basic unit level may be described as follows

$$y_{di} = \mathbf{x}_{di}' \boldsymbol{\beta} + u_d + e_{di} \quad i = 1, \dots, N_d, \quad d = 1, \dots, D \quad (1)$$

in which we have denoted with: y_{di} the value of the variable of interest y for the population unit i ($i = 1, \dots, N_d$) of *small area* d ($d = 1, \dots, D$); $\mathbf{x}_{di} = (x_{di,1}, \dots, x_{di,P})'$ the unit-specific auxiliary data; $\boldsymbol{\beta}$ the $P \times 1$ vector of regression coefficients; u_d the area-specific effects, assumed to be *iid* random variables satisfying $E_m(u_d) = 0$ and $V_m(u_d) = \sigma_u^2$; e_{di} the errors, assumed to be *iid* random variables satisfying $E_m(e_{di}) = 0$ and $V_m(e_{di}) = \sigma_e^2$.

In the following sections we will handle an extension of model (1) that allows to consider spatial correlated area effects; furthermore we will present: (i) the Best Linear Unbiased Predictors (BLUP); (ii) Empirical BLUP or EBLUP estimators; (iii) the Mean Squared Error (MSE) of EBLUP estimators taking into account of the variability in the estimated variances and covariance components.

2. General Linear Mixed Model

2.1. Notation and Model

Utilizing also the notation above introduced, let's consider the general linear mixed model

$$y_{dtgi} = \beta_0 + \mathbf{x}_{dtgi}' \boldsymbol{\beta}_t + \sum_{j=1}^J u_{j(dt)} + e_{dtgi} \quad (2)$$

for: $i = 1, \dots, N_{dtg}$; $g = 1, \dots, G$; $t = 1, \dots, T$, $d = 1, \dots, D$; in which y_{dtgi} , $\mathbf{x}_{dtgi}'$ and e_{dtgi} are the quantity introduced in section 1 but related to unit i of group g at time t in small area d ; in the following pages this unit will be indicated synthetically as *unit dtgi*. The vector $\boldsymbol{\beta}_t = (\beta_{t1}', \dots, \beta_{tg}', \dots, \beta_{tG}')$, of order Q , contains regression coefficients at time t for each group g where $\boldsymbol{\beta}_{tg} = (\beta_{tg,1}, \dots, \beta_{tg,P})'$ with $Q = P \times G$, β_0 is the intercept and random effects u_j and e_{dtgi} are assumed to be mutually independent and to follow normal distributions with zero means and given variances. The (dt) in $u_{j(dt)}$ indicates that the random effects u_j are related to time and small area; the number of levels of generic random effect u_j is denoted as A_j ($j = 1, \dots, J$) and A is the total number of levels for the J random effects. The model (2) can then be written in matrix form as

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{Z} \mathbf{u} + \mathbf{e} \quad (3)$$

in which $\mathbf{y} = \{y_{dtgi}\}$ is a vector of order $N = \mathbf{N}'\mathbf{1}_D$ with $\mathbf{N} = (N_1, \dots, N_d, \dots, N_D)'$ and $\mathbf{1}_D = (1, \dots, 1)'$ of order $D \times 1$ where $N_d = \mathbf{N}'_d \mathbf{1}_T$ being: $\mathbf{N}_d = (N_{d1}, \dots, N_{dt}, \dots, N_{dT})'$; N_{dt} is the population size of small area d at time t ; $\mathbf{1}_T = (1, \dots, 1)'$ is of order $T \times 1$. The symbol $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_t, \dots, \boldsymbol{\beta}'_T)$ denote the vector of regression coefficients; the $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_d, \dots, \mathbf{X}'_D)'$ is a known matrix of order $N \times [(T \times Q) + 1]$ where: $\mathbf{X}_d = \text{Blockdiag}(\mathbf{X}_{dt})$ for $t = 1, \dots, T$ is of order $N_d \times Q$ in which $\mathbf{X}_{dt} = \text{Blockdiag}\{\mathbf{X}_{dtg}\}$ for $g = 1, \dots, G$, being $\mathbf{X}_{dtg} = (\mathbf{x}_{dtg,1}, \dots, \mathbf{x}_{dtg,P})'$ of order $N_{dtg} \times P$ and $\mathbf{x}_{dtg,p} = (x_{dtg,p,1}, \dots, x_{dtg,p,N_{dtg}})$ for $p = 1, \dots, P$. The generic row of $\mathbf{X}_{dt}$ referred to unit $dtgi$ is represented by the following Q dimensional vector $\mathbf{x}'_{dtgi} = (\delta_1 \ddot{\mathbf{x}}'_{dtgi}, \dots, \delta_{g'} \ddot{\mathbf{x}}'_{dtgi}, \dots, \delta_G \ddot{\mathbf{x}}'_{dtgi})$ in which $\delta_{g'}$ ($g' = 1, \dots, G$) is an indicator variable being $\delta_{g'} = 1$ if $g' = g$ and $\delta_{g'} = 0$ otherwise and $\ddot{\mathbf{x}}'_{dtgi} = (x_{dtgi,1}, \dots, x_{dtgi,P})$ contains the values of the auxiliary covariates $(x_1, \dots, x_P)$ for unit $dtgi$. Considering the population units hierarchically ordered by small area d , time t , group g and unit i , it is possible to obtain an useful equivalent formulation of matrix $\mathbf{X}$. It is important to note that the introduction of the group structure represents a general framework in the definition of the fixed part of the model that allow to deal with many important cases. By means of this structure it is possible to define one different and denote respectively $\boldsymbol{\beta}_{tg}$ ($g = 1, \dots, G$) for each *subpopulation* U_{tg} related to a given partition of the population of interests U_t . In this context it is necessary that, for each possible partition of the population U_t that may be considered, must be known the population totals of the auxiliary variables for each *subpopulation* U_{dtg} ($g = 1, \dots, G$) referred to small area d (U_{dtg} is a partition of the population U_{dt}). This type of characterization of the fixed effects of the linear mixed model (3) play the same role of *model group* concept utilized in the context of *generalized regression estimator* (Estevao, Hidiroglou and Särndal, 1995). A particular case of interest is the *poststratified* one, that is obtained when $P = 1$; in this case we have $Q = G$, $\ddot{\mathbf{x}}'_{dtgi} \equiv x_{dtgi,1} \equiv x_{dtgi}$, $\mathbf{x}'_{dtgi} = (\delta_1 x_{dtgi}, \dots, \delta_{g'} x_{dtgi}, \dots, \delta_G x_{dtgi})$. The $\mathbf{Z} = (\mathbf{Z}_{u_1}, \dots, \mathbf{Z}_{u_j}, \dots, \mathbf{Z}_{u_J})$ is the incidence matrix of order $N \times A$ being $\mathbf{Z}_{u_j}$, of order $N \times A_{u_j}$, the incidence matrix for the random effect vector $\mathbf{u}_j$. Furthermore, we assume that the vectors $\mathbf{e} = \{e_{dtgi}\}$ of order $N \times 1$ and $\mathbf{u} = \{\mathbf{u}_1, \dots, \mathbf{u}_J\}$ are independent with expectations $E_m(\mathbf{e}) = \mathbf{0}$ and $E_m(\mathbf{u}) = \mathbf{0}$ and covariance matrices $V_m(\mathbf{e}) = \mathbf{R} = \sigma_e^2 \mathbf{W}$, in which $\mathbf{W}$ is a diagonal matrix of known constants, and $V_m(\mathbf{u}) = \mathbf{G} = \sigma_e^2 \boldsymbol{\Omega} = \sigma_e^2 \text{Blockdiag}(\varphi_{u_j} \rho_{u_j} \boldsymbol{\Omega}_{u_j})$ where $\varphi_{u_j} = \sigma_{u_j}^2 / \sigma_e^2$, $\sigma_{u_j}^2$, ρ_{u_j} and $\boldsymbol{\Omega}_{u_j}$ are referred to random effects u_j ($j = 1, \dots, J$) being $\boldsymbol{\Omega}_{u_j}$ a known symmetric matrix of order A_{u_j} . The covariance matrix of $\mathbf{y}$, of order N , is then $V_m(\mathbf{y}) = \mathbf{V} = \mathbf{V}(\boldsymbol{\delta}) = \sigma_e^2 (\mathbf{W} + \mathbf{Z} \boldsymbol{\Omega} \mathbf{Z}') = \sigma_e^2 \boldsymbol{\Sigma}$ in which $\boldsymbol{\delta} = (\sigma_e^2, \boldsymbol{\sigma}_u^2, \boldsymbol{\rho}_u)'$ is the unknown parameter vector and $\boldsymbol{\Sigma} = \mathbf{W} + \mathbf{Z} \boldsymbol{\Omega} \mathbf{Z}' = \mathbf{W} + \sum_j \varphi_{u_j} \rho_{u_j} \mathbf{Z}_{u_j} \boldsymbol{\Omega}_{u_j} \mathbf{Z}'_{u_j}$.

2.2. The BLUP of Parameter of Interest

Let's introduce now the $D \times 1$ vector of parameters of interest expressed as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_D)' = \mathbf{L} \mathbf{y}$ where $\mathbf{L}$ is a matrix of order $(T \times D) \times N$ related to the nature of parameter of interest and $\boldsymbol{\theta}_d = (\theta_{d1}, \dots, \theta_{dT})'$; for example, in the case of θ_{dt} is equal to the total Y_{dt} of the variable of interest y , the $[(d-1) \times T + t]$ -th row of $\mathbf{L}$ is an area indicator vector with $N - N_{dt}$ elements equal to 0 and N_{dt} elements equal to 1.

Let's denote with the subscripts r and s the sampled and non sampled parts of the vectors and matrices above introduced; after the sample of size n is selected we can partition the model (3) as

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_s \\ \mathbf{y}_r \end{bmatrix} = \begin{bmatrix} \mathbf{X}_s \\ \mathbf{X}_r \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{Z}_s \\ \mathbf{Z}_r \end{bmatrix} \mathbf{u} + \begin{bmatrix} \mathbf{e}_s \\ \mathbf{e}_r \end{bmatrix} = \begin{bmatrix} \mathbf{X}_s \boldsymbol{\beta} + \mathbf{Z}_s \mathbf{u} + \mathbf{e}_s \\ \mathbf{X}_r \boldsymbol{\beta} + \mathbf{Z}_r \mathbf{u} + \mathbf{e}_r \end{bmatrix} \quad (4)$$

and the parameter $\boldsymbol{\theta} = \mathbf{L}_s \mathbf{y}_s + \mathbf{L}_r \mathbf{y}_r$ with $\mathbf{L} = [\mathbf{L}'_s, \mathbf{L}'_r]'$ and $\mathbf{y} = [\mathbf{y}'_s, \mathbf{y}'_r]'$.

In the case in which $\boldsymbol{\delta}$ is known the Best Linear Unbiased Predictor (BLUP) of $\boldsymbol{\theta}$ is expressed by (Henderson, 1963)

$$\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \dots, \tilde{\theta}_D)' = \mathbf{L}_s \mathbf{y}_s + \mathbf{L}_r (\mathbf{X}_r \tilde{\boldsymbol{\beta}} + \mathbf{Z}_r \tilde{\mathbf{u}}) = \mathbf{L}_s \mathbf{y}_s + \tilde{\boldsymbol{\psi}} \quad (5)$$

in which $\tilde{\boldsymbol{\psi}} = \tilde{\boldsymbol{\psi}}(\boldsymbol{\delta}) = \mathbf{L}_r (\mathbf{X}_r \tilde{\boldsymbol{\beta}} + \mathbf{Z}_r \tilde{\mathbf{u}})$, that is the predictor of $\boldsymbol{\psi} = \mathbf{L}_r (\mathbf{X}_r \boldsymbol{\beta} + \mathbf{Z}_r \mathbf{u})$, being $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'_s \boldsymbol{\Sigma}_s^{-1} \mathbf{X}_s)^{-1} \mathbf{X}'_s \boldsymbol{\Sigma}_s^{-1} \mathbf{y}_s$ the Generalized Least Squares Estimator (GLSE) of $\boldsymbol{\beta}$ and $\tilde{\mathbf{u}} = \boldsymbol{\Omega} \mathbf{Z}'_s \boldsymbol{\Sigma}_s^{-1} (\mathbf{y}_s - \mathbf{X}_s \tilde{\boldsymbol{\beta}})$ the BLUP of $\mathbf{u}$, where $\mathbf{V}_s = \sigma_e^2 \boldsymbol{\Sigma}_s$ is the covariance matrix of order n of the vector $\mathbf{y}_s$.

An alternative expression of the BLUP of parameter $\boldsymbol{\theta}$ may be obtained by means of theorem 1 and theorem 2 given below.

Theorem 1. Let $\mathbf{A}$ and $\mathbf{D}$ be nonsingular matrices of order m and n and $\mathbf{B}$ be $m \times n$ matrix; then $(\mathbf{A} + \mathbf{BDB}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} [\mathbf{B}' \mathbf{A}^{-1} \mathbf{B} + \mathbf{D}^{-1}]^{-1} \mathbf{B}' \mathbf{A}^{-1}$ (Rao, 1973, pag 33).

Using this theorem, for $\mathbf{A} = \mathbf{W}_s$, $\mathbf{B} = \mathbf{Z}_s$ and $\mathbf{D} = \boldsymbol{\Omega}$, the inverse of $\boldsymbol{\Sigma}_s$ may be expressed as

$$\boldsymbol{\Sigma}_s^{-1} = (\mathbf{W}_s + \mathbf{Z}_s \boldsymbol{\Omega} \mathbf{Z}'_s)^{-1} = \mathbf{W}_s^{-1} - \mathbf{W}_s^{-1} \mathbf{Z}_s \mathbf{T}_s^* \mathbf{Z}'_s \mathbf{W}_s^{-1} \quad (6)$$

where $\mathbf{T}_s^* = (\mathbf{Z}'_s \mathbf{W}_s^{-1} \mathbf{Z}_s + \boldsymbol{\Omega}^{-1})^{-1}$ is a matrix of order A .

Theorem 2. For $\mathbf{W}_s$ and $\boldsymbol{\Omega}$ nonsingular matrices of order n and D above introduced we have (Chambers and Saei, 2002)

$$\boldsymbol{\Omega} \mathbf{Z}'_s \boldsymbol{\Sigma}_s^{-1} = \mathbf{T}_s^* \mathbf{Z}'_s \mathbf{W}_s^{-1} \quad (7)$$

In order to show the above result we start with the product of the two matrices $\mathbf{Z}'_s$ and $\boldsymbol{\Sigma}_s^{-1}$, obtaining $\mathbf{Z}'_s \boldsymbol{\Sigma}_s^{-1} = \mathbf{Z}'_s \mathbf{W}_s^{-1} - \mathbf{Z}'_s \mathbf{W}_s^{-1} \mathbf{Z}_s \mathbf{T}_s^* \mathbf{Z}'_s \mathbf{W}_s^{-1}$. Using the inverse of matrix $\mathbf{T}_s^*$ it is possible to write $\mathbf{Z}'_s \mathbf{W}_s^{-1} \mathbf{Z}_s = [(\mathbf{T}_s^*)^{-1} - \boldsymbol{\Omega}^{-1}]$ that introduced into the last expression gives $\mathbf{Z}'_s \boldsymbol{\Sigma}_s^{-1} = \mathbf{Z}'_s \mathbf{W}_s^{-1} - [(\mathbf{T}_s^*)^{-1} - \boldsymbol{\Omega}^{-1}] \mathbf{T}_s^* \mathbf{Z}'_s \mathbf{W}_s^{-1}$; by some algebra we obtain finally $\mathbf{Z}'_s \boldsymbol{\Sigma}_s^{-1} = \boldsymbol{\Omega}^{-1} \mathbf{T}_s^* \mathbf{Z}'_s \mathbf{W}_s^{-1}$ and by product for $\boldsymbol{\Omega}$ result (7) is obtained.

Using formula (7) the equation of $\tilde{\mathbf{u}}$ above introduced became (Chambers and Saei, 2003)

$$\tilde{\mathbf{u}} = \mathbf{T}_s^* \mathbf{Z}'_s \mathbf{W}_s^{-1} (\mathbf{y}_s - \mathbf{X}_s \tilde{\boldsymbol{\beta}}) \quad (8)$$

and consequently we obtain an alternative formulation of BLUP (5).

2.3. The EBLUP of Parameter of Interest

The BLUP estimator $\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}(\tilde{\boldsymbol{\delta}})$ given by (5) depends on the variance parameters $\tilde{\boldsymbol{\delta}}$, which are unknown in practical applications. Replacing $\tilde{\boldsymbol{\delta}}$ by an estimator $\hat{\tilde{\boldsymbol{\delta}}} = \hat{\tilde{\boldsymbol{\delta}}}(\mathbf{y}_s)$ we obtain a two stage estimator $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\hat{\tilde{\boldsymbol{\delta}}})$, that is depends on $\hat{\tilde{\boldsymbol{\delta}}}$, which is referred to as the Empirical Best Linear Unbiased Predictor (EBLUP). The two stage estimator $\hat{\boldsymbol{\theta}}$ is unbiased for $\boldsymbol{\theta}$ provided (a) $E_m(\hat{\boldsymbol{\theta}})$ is finite; (b) $\hat{\tilde{\boldsymbol{\delta}}}$ is an even translation invariant estimator of $\tilde{\boldsymbol{\delta}}$, that is, $\hat{\tilde{\boldsymbol{\delta}}}(-\mathbf{y}_s) = \hat{\tilde{\boldsymbol{\delta}}}(\mathbf{y}_s)$ and $\hat{\tilde{\boldsymbol{\delta}}}(\mathbf{y}_s - \mathbf{X}\mathbf{a}) = \hat{\tilde{\boldsymbol{\delta}}}(\mathbf{y}_s)$ for all $\mathbf{y}_s$ and $\mathbf{a}$; (c) the distributions of $\mathbf{u}$ and $\mathbf{e}$ are both symmetric around $\mathbf{0}$ (not necessarily normal). A proof that $\hat{\boldsymbol{\theta}}$ is unbiased, due to Kackar and Harville (1981), uses the following results:

$$(i) \quad \hat{\tilde{\boldsymbol{\delta}}}(\mathbf{y}_s) = \hat{\tilde{\boldsymbol{\delta}}}(\mathbf{Z}_s \mathbf{u} + \mathbf{e}_s) = \hat{\tilde{\boldsymbol{\delta}}}(-\mathbf{Z}_s \mathbf{u} - \mathbf{e}_s) \quad , \quad (ii) \quad \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \boldsymbol{\varepsilon}(\mathbf{u}, \mathbf{e}_s) - \mathbf{L}_r \mathbf{Z}_r \mathbf{u}$$

where $\boldsymbol{\varepsilon}(\mathbf{u}, \mathbf{e}_s)$ is an odd function of $\mathbf{u}$ and $\mathbf{e}_s$, that is, $\boldsymbol{\varepsilon}(-\mathbf{u}, -\mathbf{e}_s) = -\boldsymbol{\varepsilon}(\mathbf{u}, \mathbf{e}_s)$. Result (ii) implies $E_m(\boldsymbol{\varepsilon}(\mathbf{u}, \mathbf{e}_s)) = E_m(\boldsymbol{\varepsilon}(-\mathbf{u}, -\mathbf{e}_s)) = -E_m(\boldsymbol{\varepsilon}(\mathbf{u}, \mathbf{e}_s))$ or $E_m(\boldsymbol{\varepsilon}(\mathbf{u}, \mathbf{e}_s)) = 0$, so that $E_m(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{0}$. Kackar and Harville (1981) have also shown that standard procedures for estimating $\tilde{\boldsymbol{\delta}}$ yield even translation invariant estimators; in particular Maximum Likelihood (ML), Restricted Maximum Likelihood (REML) and the method of fitting constants, also called Henderson's method 3 (for details of these methods see Harville, 1977; Searle *et al.*, 1992; Cressie, 1992; Rao, 1997; Rao, 2003; Chambers and Saei, 2003).

In conclusion the EBLUP of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = \mathbf{L}_s \mathbf{y}_s + \mathbf{L}_r (\mathbf{X}_r \hat{\boldsymbol{\beta}} + \mathbf{Z}_r \hat{\mathbf{u}}) = \mathbf{L}_s \mathbf{y}_s + \hat{\boldsymbol{\psi}} \quad (9)$$

in which $\hat{\psi} = \hat{\psi}(\hat{\delta}) = \mathbf{L}_r(\mathbf{X}_r\hat{\beta} + \mathbf{Z}_r\hat{u})$ being $\hat{\beta}$ the Best Linear Unbiased Estimator (BLUE) of β and $\hat{u}$ the EBLUP of u .

2.4. MSE of EBLUP

An essential measure of the quality of a statistically estimator is its *mean square error* (MSE) and so it is important that any small area estimate be accompanied by an estimate of its MSE. In this section we describe the main aspects of the methodology for estimating the MSE of $\hat{\theta}$, that is the MSE of $\hat{\psi}$, included in (9), which is the predictor of ψ .

The error in the EBLUP estimator $\hat{\psi}$ may be decomposed as

$$\hat{\psi} - \psi = [\tilde{\psi} - \psi] + [\hat{\psi} - \tilde{\psi}]$$

therefore

$$\text{MCPE}(\hat{\psi}) = \text{MCPE}(\tilde{\psi}) + E_m[\hat{\psi} - \tilde{\psi}][\hat{\psi} - \tilde{\psi}]' + 2E_m[\tilde{\psi} - \psi][\hat{\psi} - \tilde{\psi}]'. \quad (10)$$

where MCPE is the *mean cross product error*. Under normality of the random effects u and e , the cross-product term in (10) is zero provided $\tilde{\theta}$ is translation invariant so that

$$\text{MCPE}(\hat{\psi}) = \text{MCPE}(\tilde{\psi}) + E_m[\hat{\psi} - \tilde{\psi}][\hat{\psi} - \tilde{\psi}]'. \quad (11)$$

It is clear from (11) that the $\text{MCPE}(\hat{\psi})$ is always larger than $\text{MCPE}(\tilde{\psi})$, under normality. The common practice of approximating $\text{MCPE}(\hat{\psi})$ by $\text{MCPE}(\tilde{\psi})$ could therefore lead to significant underestimation, especially in cases where $\tilde{\psi}$ varies with δ to a significant extent and where the variability of $\hat{\delta}$ is not small. Referring to $\text{MCPE}(\tilde{\psi})$ we note that

$$\tilde{\psi} = \tilde{\psi}^* + \mathbf{D}(\tilde{\beta} - \beta)$$

where:

$$\tilde{\psi}^* = \mathbf{L}_r\mathbf{X}_r\beta + \mathbf{L}_r\mathbf{Z}_r[\mathbf{\Omega}\mathbf{Z}_s'\Sigma_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\beta)]$$

$$\mathbf{D} = \mathbf{L}_r\mathbf{X}_r + \mathbf{L}_r\mathbf{Z}_r\mathbf{\Omega}\mathbf{Z}_s'\Sigma_s^{-1}\mathbf{X}_s.$$

Since $\tilde{\psi}^* - \psi$ and $\mathbf{D}(\tilde{\beta} - \beta)$ are uncorrelated, it is possible to show that

$$\text{MCPE}(\tilde{\psi}) = \text{MCPE}(\tilde{\psi}^*) + V_m[\mathbf{D}(\tilde{\beta} - \beta)] = \mathbf{M}_1 + \mathbf{M}_2$$

where

$$\mathbf{M}_1 = \sigma_e^2 \mathbf{L}_r' \mathbf{X}_r (\mathbf{\Omega} - \mathbf{\Omega} \mathbf{Z}_s' \mathbf{\Sigma}_s^{-1} \mathbf{Z}_s \mathbf{\Omega}) \mathbf{X}_r' \mathbf{L}_r'$$

$$\mathbf{M}_2 = \sigma_e^2 \mathbf{D} (\mathbf{X}_s' \mathbf{\Sigma}_s^{-1} \mathbf{X}_s)^{-1} \mathbf{D}'$$

The values of $\text{MSE}(\tilde{\boldsymbol{\psi}})$ and consequently of those of $\text{MSE}(\tilde{\boldsymbol{\theta}})$ are included into the principal diagonal of $\text{MCPE}(\tilde{\boldsymbol{\psi}})$. Indicating with $g_{1,i}(\boldsymbol{\delta})$ and $g_{2,i}(\boldsymbol{\delta})$ the i -th element of the principal diagonal respectively of matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ we have $\text{MCPE}(\tilde{\boldsymbol{\psi}}_i) = \text{MSE}(\tilde{\boldsymbol{\psi}}_i) = g_{1,i}(\boldsymbol{\delta}) + g_{2,i}(\boldsymbol{\delta})$.

Considering now the last term of (11), we observe that it is generally intractable except in special cases (Peixoto and Harville, 1986). It is therefore necessary to obtain an approximation to this term; an heuristic approximation, due to Kackar and Harville (1984), may be illustrated in the following way. By a Taylor approximation, we have

$$\hat{\boldsymbol{\psi}} - \tilde{\boldsymbol{\psi}} \approx (\partial \tilde{\boldsymbol{\psi}} / \partial \boldsymbol{\delta}') [\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}] = \mathbf{F} [\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}]$$

with $\mathbf{F} = \partial \tilde{\boldsymbol{\psi}} / \partial \boldsymbol{\delta}$. It is useful now for simplicity to consider the generic element of the above formula

$$\hat{\boldsymbol{\psi}}_i - \tilde{\boldsymbol{\psi}}_i \approx (\partial \tilde{\boldsymbol{\psi}}_i / \partial \boldsymbol{\delta}) (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) = \mathbf{v}_i' [\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}] \quad i = 1, \dots, (D \times T)$$

in which $\mathbf{v}_i = \partial \tilde{\boldsymbol{\psi}}_i / \partial \boldsymbol{\delta}$ that is the i -th row of matrix $\mathbf{F}$. Furthermore under normality

$$\mathbf{v}_i \approx \partial \tilde{\boldsymbol{\psi}}_i^* / \partial \boldsymbol{\delta} = \mathbf{v}_i^*$$

being $\tilde{\boldsymbol{\psi}}_i^*$ the i -th generic row of vector $\tilde{\boldsymbol{\psi}}^*$. Thus

$$E_m(\hat{\boldsymbol{\psi}}_i - \tilde{\boldsymbol{\psi}}_i)^2 \approx E_m(\mathbf{v}_i' [\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}])^2 \approx E_m(\mathbf{v}_i^{*'} [\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}])^2 \quad (12)$$

Following Prasad and Rao (1990), it is possible to show that

$$E_m(\mathbf{v}_i^{*'} [\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}])^2 \approx \text{tr} \{ E_m(\mathbf{v}_i^* \mathbf{v}_i^{*'}) \bar{\mathbf{V}}(\hat{\boldsymbol{\delta}}) \} = g_{3,i}(\boldsymbol{\delta})$$

where $\bar{\mathbf{V}}(\hat{\boldsymbol{\delta}})$ is the asymptotic covariance of $\hat{\boldsymbol{\delta}}$. In conclusion the $\text{MCPE}(\hat{\boldsymbol{\psi}}_i)$ $i = 1, \dots, (D \times T)$ is given by

$$\text{MCPE}(\hat{\boldsymbol{\psi}}_i) = \text{MSE}(\hat{\boldsymbol{\psi}}_i) = g_{1,i}(\boldsymbol{\delta}) + g_{2,i}(\boldsymbol{\delta}) + g_{3,i}(\boldsymbol{\delta}) \quad (13)$$

For practical application , we need an estimator of $MSE(\hat{\psi}_i)$ as a measure of variability associated with $\hat{\psi}_i$. A naïve approach approximates $MSE(\hat{\psi}_i)$ by $MSE(\tilde{\psi}_i)$ and then substitutes $\hat{\delta}$ for δ . The resulting estimator of MSE is

$$M\tilde{S}E_N(\hat{\psi}_i) = g_{1,i}(\hat{\delta}) + g_{2,i}(\hat{\delta}).$$

Another estimator of MSE is obtained by substituting $\hat{\delta}$ for δ in the MSE approximation (13)

$$M\tilde{S}E(\hat{\psi}_i) = g_{1,i}(\hat{\delta}) + g_{2,i}(\hat{\delta}) + g_{3,i}(\hat{\delta}).$$

in which $E_m(g_{2,i}(\hat{\delta})) \approx g_{2,i}(\delta)$, $E_m(g_{3,i}(\hat{\delta})) \approx g_{3,i}(\delta)$ but $g_{1,i}(\hat{\delta})$ is not the correct estimator of $g_{1,i}(\delta)$ because its bias is generally of the same order as $g_{2,i}(\delta)$ and $g_{3,i}(\delta)$. For details concerning the bias of $g_{1,i}(\hat{\delta})$, $M\tilde{S}E_N(\hat{\psi}_i)$ and $M\tilde{S}E(\hat{\psi}_i)$ see Prasad and Rao (1990) and Rao (2003).

3. Particular Cases

Linear models with spatial correlated area effects and linear models with independent and correlated time effects are special cases of model (2) with $J=1$ and $J=2$ respectively. The methodology concerning these cases may be derived adapting the expressions described in the previous section 2. The first case related to spatial correlation is based on the following model

$$y_{dgi} = \beta_0 + \mathbf{x}_{dgi}' \boldsymbol{\beta} + u_d + e_{dgi}. \quad (14)$$

Furthermore have: $A_1 = D$; $\boldsymbol{\Omega} = \{f_{d,d'}\}$ is a matrix of order D with $f_{d,d'}$ denoting a *distance* (not necessarily Euclidean) between small areas d and d' ; $\boldsymbol{\delta} = (\sigma_e^2, \sigma_u^2, \rho_u)'$; $V_m(\mathbf{u}) = \sigma_u^2 \rho_u \boldsymbol{\Omega}$.

The second case related to spatial and temporal correlation is based on the following model

$$y_{dtgi} = \beta_0 + \mathbf{x}_{dtgi}' \boldsymbol{\beta}_t + u_{1d} + u_{2t} + e_{dtgi}. \quad (15)$$

Furthermore: $\boldsymbol{\Omega} = \sigma_e^2 \text{Blockdiag} \{ \varphi_{u_{1d}} \rho_{u_{1d}} \boldsymbol{\Omega}_{u_{1d}}, \varphi_{u_{2t}} \rho_{u_{2t}} \boldsymbol{\Omega}_{u_{2t}} \}$; $A_1 = D$; $A_2 = T$; $\boldsymbol{\delta} = (\sigma_e^2, \sigma_{u_{1d}}^2, \rho_{u_{1d}}, \sigma_{u_{2t}}^2, \rho_{u_{2t}})'$.

Finally, we suggest for the first case an alternative formulation the BLUP of $\boldsymbol{\theta}$ expressed in terms of area level totals, which is useful either for computer implementation and for showing interesting characteristics of *area level* and *unit level*

models, we derive new expressions of $\tilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{u}}$. Let's start with formula (6) in which, writing

$$\mathbf{W}_s^{-1} = \boldsymbol{\Lambda}_s = \text{Diag}(\lambda_{dgi}), \quad \text{for } d = 1, \dots, D; g = 1, \dots, G; i = 1, \dots, n_{dg}$$

$$\mathbf{Z}_s' \mathbf{W}_s^{-1} \mathbf{Z}_s = \ddot{\boldsymbol{\Lambda}}_s = \text{Diag}(\lambda_{sd.}), \quad \text{for } d = 1, \dots, D$$

for $\lambda_{sd.} = (\lambda_{sd,1}, \dots, \lambda_{sd,n_{dg}}) \mathbf{1}_{n_{dg}}$ being n_{dg} the sample size of group g in area d , we obtain

$$\mathbf{T}_s^* = (\ddot{\boldsymbol{\Lambda}}_s + \boldsymbol{\Omega}^{-1})^{-1}, \quad \mathbf{T}_s^* \mathbf{Z}_s' \mathbf{W}_s^{-1} = (\ddot{\boldsymbol{\Lambda}}_s + \boldsymbol{\Omega}^{-1})^{-1} \mathbf{Z}_s' \boldsymbol{\Lambda}_s. \quad (16)$$

Introducing, now, the explicit expression of $\boldsymbol{\Sigma}_s^{-1}$ given in (6) the GLSE of $\boldsymbol{\beta}$ may be reformulated as $\tilde{\boldsymbol{\beta}} = \mathbf{B}_s^{-1} \mathbf{b}_s$ with $\mathbf{B}_s = \mathbf{B}_{s,1} - \mathbf{B}_{s,2}$ and $\mathbf{b}_s = \mathbf{b}_{s,1} - \mathbf{b}_{s,2}$, being $\mathbf{B}_{s,1} = \mathbf{X}_{s,\lambda}' \mathbf{X}_s$, $\mathbf{B}_{s,2} = \mathbf{X}_{s,\lambda}' \mathbf{Z}_s \mathbf{T}_s^* \mathbf{Z}_s' \mathbf{X}_{s,\lambda}$, $\mathbf{b}_{s,1} = \mathbf{X}_{s,\lambda}' \mathbf{y}_s$ and $\mathbf{b}_{s,2} = \mathbf{X}_{s,\lambda}' \mathbf{Z}_s \mathbf{T}_s^* \mathbf{Z}_s' \mathbf{y}_{s,\lambda}$ in which $\mathbf{X}_{s,\lambda}' = \mathbf{X}_s' \boldsymbol{\Lambda}_s$.

Keeping in mind the structure of matrix $\mathbf{X}_s$, described in section 2.1, it is possible to express the generic $\mathbf{B}_{s,k}$ and $\mathbf{b}_{s,k}$ for $k = 1, 2$ as functions of aggregated sample data for each group g of area d (for $g=1, \dots, G$; $d=1, \dots, D$):

$$\mathbf{B}_{s,1} = \sum_{d=1}^D [\text{Blockdiag}(\sum_{i=1}^{n_{dg}} \mathbf{x}_{dgi} \mathbf{x}_{dgi}' \lambda_{dgi}), \quad \text{for } g = 1, \dots, G]$$

$$\mathbf{B}_{s,2} = \sum_{d=1}^D [\text{Block}(\mathbf{c}_{sdg,\lambda} \mathbf{x}_{sdg',\lambda}'), \quad \text{for } g = g' = 1, \dots, G]$$

$$\mathbf{b}_{s,1} = \sum_{d=1}^D [\text{Col}(\sum_{i=1}^{n_{dg}} \mathbf{x}_{dgi} y_{dgi} \lambda_{dgi}), \quad \text{for } g = 1, \dots, G]$$

$$\mathbf{b}_{s,2} = \sum_{d=1}^D [\text{Col}(\mathbf{c}_{sdg,\lambda} y_{sd,\lambda}), \quad \text{for } g = 1, \dots, G]$$

where $\mathbf{c}_{dg,\lambda}$ is a vector of order P expressed as $\mathbf{c}_{dg,\lambda} = \sum_{d'} \mathbf{T}_{s,dd'}^* \mathbf{x}_{sd'g,\lambda}$ ($d=1, \dots, D$) being $\mathbf{x}_{sdg,\lambda} = \sum_i \mathbf{x}_{dgi}' \lambda_{dgi}$ ($i = 1, \dots, n_{dg}$). Note that $\mathbf{c}_{dg,\lambda}$ is a linear combination of the vectors $\mathbf{x}_{sdg,\lambda}$ ($g=1, \dots, G$) with weights given respectively by the elements $\mathbf{T}_{s,dd'}^*$ ($d'=1, \dots, D$) of the d -th row of matrix $\mathbf{T}_s^*$, related to area d . It is interesting to note that in the case of model (1), in which it is considered also the group structure:

$$\mathbf{T}_s^* = (\ddot{\mathbf{A}} + \varphi^{-1} \mathbf{I}_D)^{-1} = \text{Diag}(\mathbf{T}_{s,dd}^*) = \text{Diag}(\sigma_u^2[\sigma_u^2 + (\sigma_e^2 / \lambda_{sd})]^{-1}) \quad \text{and} \quad \text{then}$$

$$\mathbf{c}_{dg,\lambda} = \mathbf{T}_{s,dd}^* \mathbf{x}_{sdg,\lambda}.$$

On the basis of the above results, and using some algebra, it is possible to derive the following expression for the BLUP of $\mathbf{u}$ in which the generic element is given by

$$\tilde{u}_d = \sum_{d'=1}^D \mathbf{T}_{s,dd'}^* \sum_{g=1}^G (y_{d'g,\lambda} - \mathbf{x}_{d'g,\lambda}' \hat{\boldsymbol{\beta}}_g)$$

where $y_{d'g,\lambda} = \sum_i y_{dgi} \lambda_{dti}$ ($i = 1, \dots, n_{dg}$). Consequently it is possible to obtain the BLUP of $\boldsymbol{\theta}$ where the generic term is expressed by

$$\tilde{\theta}_d = \sum_{g=1}^G (y_{dg} + \mathbf{x}_{rdg}' \hat{\boldsymbol{\beta}}_g) + (N_d - n_d) \tilde{u}_d$$

being $y_{dg} = \sum_i y_{dgi}$ ($i = 1, \dots, n_{dg}$) and $\mathbf{x}_{rdg} = \sum_i \mathbf{x}_{dgi}$ ($i = 1, \dots, N_{dg} - n_{dg}$).

4. Conclusions

The sampling surveys carried out by Italian National Institute of Statistics (Istat) are large scale surveys planned to produce reliable estimates of the parameters of interest only for medium/large subpopulation groups at national level, while the reliability of estimates referred to regional levels is guaranteed only for larger subpopulations.

However the usual direct survey estimators referred to small subpopulations (or small areas) — such as subregional areas, specific age-sex groups of people or specific subsets of enterprises identified by cross classification of economic activity and size class — are likely to yield unacceptably standard errors due to unduly small size of the sample in the small areas.

The methods presented in section 3 may represent useful tools for addressing the problems of producing reliable estimates referred to small areas without enlarging sample sizes.

Model (14), that consider *correlated area effects*, may be useful adopted for producing small area estimates for occasional surveys such as *Health Status and Health Service Utilization Survey* conducted every five years by ISTAT.

However the most important sampling surveys on households and enterprises conducted by Istat (e.g. *labour force survey*, *consumer expenditure survey*; *structural business survey*; *short term business surveys*) are repeated surveys producing annually and quarterly estimates of the parameters of interest. In this context the use of model (15) for the construction of the EBLUP of the parameters of interest for each small area may improve the performances of estimators, in terms of MSE, with respect to the case of standard model (1). The gain in reliability is due to the fact that this model is referred to more realistic situations in which the values of the variables of interest are correlated over space and time and thus this correlation may be used to construct small area predictors which borrow strength from the sample units observed in different areas or in the same areas but in different times.

It is important to note that general model (2) and also particular models (14) and (15) consider a correlation structure for each one of the J random effects, but it is not considered a dependent structure between the different random effects; in particular in model (15) it is not assumed that area random effects and time random effects are related each other. Hence an important development may derive from considering more complicated models based on area/time joint effects.

References

- Chambers R., Saei A. (2002) Small Area Estimation: A Review of Methods Based on the Application of Mixed Models, *Report of EURAREA Project*: “Enhancing Small Area Estimation Techniques to Meet European Needs”.
- Chambers R., Saei A. (2003) Small Area Estimation in Generalized Linear Mixed Models, *Report of EURAREA Project*: “Enhancing Small Area Estimation Techniques to Meet European Needs”.
- Cressie N. (1992) REML Estimation in Empirical Bayes Smoothing of Census Undercount, *Survey Methodology*, 18, 75-94.
- Estevao V., Hidioglou M. A. Särndal C.E. (1995) Methodological Principles for a Generalized Estimation Systems at Statistics Canada, *Journal of Official Statistics*, 11, 181-204.
- Fay R. E. (1987) Application of Multivariate Regression to Small Domain Estimation. In *Small Area Estimation* (R. Platek, J. N. K. Rao, C. E Särndal and m. P. Singh, eds.) 91-102, Wiley, New York
- Fuller W. A. (1999) Small Area Statistics for Agriculture from a National Survey – in *Proceedings of the International Statistical Institute 1999*, Helsinki.
- Eurostat (2002) *European social statistics - Income, poverty and social exclusion*, 2nd report.
- Eurostat (2003) *Regions. Nomenclature of territorial units for statistics, NUTS – 2003*, Methods and Nomenclatures series.
- Harville D. A. (1977) Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems, *Journal of the American Statistical Association*, 72, 322-340.
- Henderson C.R. (1963) Selection Index and Expected Genetic Advance, *Statistical Genetics and Plant Breeding*, National Academy of Science, National Research Council, Publication 982, Washington, DC, 141-163.
- Kackar R. N. , Harville D. A. (1981) Ubiasedness of Two-stage Estimation and Prediction Procedures for Mixed Linear Models, *Communication in Statistics*, Series A, 10, 1249-1261.
- Kackar R. N. , Harville D. A (1984) Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models, *Journal of the American Statistical Association*, 79, 853-862.
- Noble A., Haslett S., Arnold G. (2002) Small Area Estimation via Generalized Linear Models, *Journal of Official Statistics*, 18, 1, pp.45-60.
- Peixoto J. L., Harville D. A. (1986) Comparisons of Alternative Predictors Under the Balanced One-Way Random Model, *Journal of the American Statistical Association*, 81, 431-436.

- Rao C. R. (1973) *Linear Statistical Inference and its Applications*, 2nd Wiley, New York.
- Rao P. S. R. S. (1997) *Variance Component Estimation*, Chapman & Hall, London.
- Rao J. N. K. (2003) *Small Area Estimation*, Wiley, New York.
- Särndal C. E., Swensson B., Wretman J. (1992) *Model Assisted Survey Sampling*, Springer-Verlag, New York, Inc.
- Searle S. R., Casella G., McCulloch C. E. (1992) *Variance Components*, Wiley, New York.