Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Small area estimation of average household income based on unit level models for panel data

by Enrico Fabrizi, Maria Rosaria Ferrante and Silvia Pacei

December, 2007

Survey Methodology, December 2007
Vol. 33, No. 2, pp. 187-198
Statistics Canada, Catalogue no. 12-001-X

Small area estimation of average household income based on unit level models for panel data

Enrico Fabrizi, Maria Rosaria Ferrante and Silvia Pacei [1]

Abstract

The European Community Household Panel (ECHP) is a panel survey covering a wide range of topics regarding economic, social and living conditions. In particular, it makes it possible to calculate disposable equivalized household income, which is a key variable in the study of economic inequity and poverty. To obtain reliable estimates of the average of this variable for regions within countries it is necessary to have recourse to small area estimation methods. In this paper, we focus on empirical best linear predictors of the average equivalized income based on "unit level models" borrowing strength across both areas and times. Using a simulation study based on ECHP data, we compare the suggested estimators with cross-sectional model-based and design-based estimators. In the case of these empirical predictors, we also compare three different MSE estimators. Results show that those estimators connected to models that take units' autocorrelation into account lead to a significant gain in efficiency, even when there are no covariates available whose population mean is known.

Key Words: European Community Household Panel; Average equivalized income; Linear mixed models; Empirical best linear unbiased predictor; MSE estimation.

1. Introduction

In recent years, the academic world has taken an increasing interest in the analysis of regional economic disparities that represent a serious challenge to the promotion of national economic growth, and thus to social cohesion. This is particularly true within the European Union, where regional disparities are a distinguishing feature of the economic landscape. This renewed interest in local economies has produced a growing demand for regional statistical information and has stimulated research on income distribution, poverty and social exclusion at the sub-national level.

In the 1990s, Eurostat (the EU's Statistics Bureau) launched the European Community Household Panel (ECHP), an annual panel survey of European households conducted using standardised methods throughout the EU's various member countries (Betti and Verma 2002; Eurostat 2002). The ECHP terminated in 2001, after eight waves. Currently, it is being replaced by the Survey on Income and Living Conditions in the Community (EU-SILC), which resembles the ECHP in many ways, but for which no data has yet been published. The ECHP panel survey covered a wide range of topics and, in particular, it made it possible to calculate disposable equivalized household income, which constitutes a key variable in the study of economic equity and poverty.

The ECHP was designed to provide reliable estimates for large areas within countries called NUTS1 (NUTS stands for the "Nomenclature of Territorial Units for Statistics" which is defined according to certain principles described on the EUROSTAT web site http://europa.eu.int/comm/eurostat/ramon/nuts/home_regions_en.html). Unfortunately NUTS1 correspond to areas (five groups of Administrative Regions in the Italian case) that are too large to effectively measure local area income disparity or to provide useful information for the purposes of regional governance. Therefore, to obtain estimates for a finer geographic detail, a small area estimation method has to be used and the problem is to select an appropriate and effective method.

In this paper, in order to combine information from past surveys, related auxiliary variables and small areas, we consider several possible extensions of the well-known unit level nested error regression model (see Battese, Harter and Fuller 1988) for the estimation of the average of household equivalized income. Using ECHP panel survey data, we illustrate how such model could be potentially useful in improving the efficiency of small area estimates by exploiting the correlation of individual household incomes over time.

In section 2, we present a general set-up for small area estimation using panel survey data and briefly review both design-based and model-based small area estimation methods. In this section, we develop empirical best linear unbiased predictors (EBLUP) and their mean squared error (MSE) estimators for selected unit level cross-sectional and time series models using the available theory on EBLUP for small area estimation (see Rao 2003, and Jiang and Lahiri 2006a, for details). We note that cross-sectional and time series models were considered in the small area literature,

1. Enrico Fabrizi, DMSIA, University of Bergamo, via dei Caniana 2, 24127, Bergamo, Italy. E-mail: enrico.fabrizi@unibg.it; M.R. Ferrante, Department of Statistics, University of Bologna, via Belle Arti 41, 40126, Bologna, Italy. E-mail: ferrante@stat.unibo.it; S. Pacei, Department of Statistics, University of Bologna, via Belle Arti 41, 40126, Bologna, Italy. E-mail: pacei@stat.unibo.it.

but only in the context of area level modelling (see Rao and Yu 1994; Ghosh, Nangia and Kim 1996; Datta, Lahiri, Maiti and Lu 1999; Datta, Lahiri and Maiti 2002; Pfeffermann 2002; among others).

In section 3, we briefly review ECHP survey and describe how we use this survey data to conduct a Monte Carlo simulation study to compare different small area estimators and their MSE estimators. In section 4, we report results from the Monte Carlo simulation experiment. We note that the simulation experiment is aimed at evaluating design-based properties of all estimators, even if they are derived as model based predictors. We observed that the EBLUPs perform very well compared to the design-based estimators even though our pseudo-population exhibits signs of non-normality. The non-normality of the pseudo-population, however, seems to affect the efficiency of the MSE estimators. In our simulation, the Taylor series (see Prasad and Rao 1990; Datta and Lahiri 1999, among others) and the parametric bootstrap (see Butar and Lahiri 2003) MSE estimators are found out to be more sensitive to the non-normality than the jackknife method of Jiang, Lahiri and Wan (2002). We end the paper with a few concluding remarks.

2. The small area estimation methods considered

To describe sample data, let y_{dti} denote the value of a study variable for the i^{th} unit belonging to the d^{th} small area for time t $(d=1,\ldots,m;\ t=1,\ldots,T;\ i=1,\ldots,n_{dt})$. Moreover, let $\mathbf{x}_{dti}$ be the vector of covariates' values associated with each y_{dti} (and whose first element is equal to 1), and let $\mathbf{X}_s=\{\mathbf{x}'_{dti}\}$ be the $n\times p$ matrix of covariates' values for the whole sample $(n=\sum_{d,t} n_{dt})$. Let us suppose that we are interested in predicting small area means for the target variable at final time $T: \bar{Y}_{dT}, (d=1,\ldots,m)$. Let us also suppose that the vectors of mean population values of covariates are known for time T; we denote these vectors by $\bar{\mathbf{X}}'_{dT}\,(d=1,\ldots,m)$.

2.1 Design-based estimators

A first solution to the small area estimation problem is to use direct estimators, that is, estimators employing only y values obtained from the area (and time) which the parameter refers to. The simplest of direct estimators of the population mean is the weighted mean. We denote this direct estimator as $\bar{y}_{dT,\,\text{DIR}}\,(d=1,\ldots,m)$ and we will be using it as a benchmark in the following sections.

Synthetic estimators may be generally defined as unbiased estimators for a larger area with acceptable standard errors. They are used to calculate estimates for small areas, under the hypothesis that small areas have the same characteristics as larger ones. Moreover, when information about auxiliary variables is available, a particular synthetic estimator, the regression estimator, may be obtained by fitting a regression model to all sample data. Note that the synthetic estimator is area specific with respect to the auxiliary variables but not with respect to the study variable.

For instance, if we consider only those observations from the last wave $(t=T)$, the simple regression model would be given by:

$$y_{dTi}=\mathbf{x}'_{dTi}\,\beta+e_{dTi}$$
$$E(e_{dTi})=0,\quad E(e^2_{dTi})=\pi^2.$$

To take account of the complexity of the sampling design, the weighted least squares estimate $\tilde{\beta}_w$ of β may be obtained, and thus the synthetic regression estimator will be given by:

$$\bar{y}_{dT,\,\text{RSYN}}=\bar{\mathbf{X}}'_{dT}\,\tilde{\beta}_w\;\; d=1,\ldots,m. \tag{1}$$

Synthetic estimators usually display very low variances, but they may be severely biased whenever the model holding for the whole sample does not properly fit area-specific data. Composite estimators are weighted averages of a direct and a synthetic estimator. We consider the composite estimator:

$$\bar{y}_{dT,\,\text{COMP}}=\phi_{dT}\bar{y}_{dT,\,\text{DIR}}+(1-\phi_{dT})\bar{y}_{dT,\,\text{RSYN}}, \tag{2}$$

where

$$\phi_{dT}=\frac{\text{MSE}_D(\bar{y}_{dT,\,\text{RSYN}})}{\text{MSE}_D(\bar{y}_{dT,\,\text{DIR}})+\text{MSE}_D(\bar{y}_{dT,\,\text{RSYN}})}$$

and MSE_D signifies that the mean square error is evaluated in relation to the randomization distribution. This choice of ϕ_{dT} leads to composite estimators $\bar{y}_{dT,\,\text{COMP}}$ that are approximately optimal in terms of MSE_D (see Rao 2003, section 4.3). In practice, the quantities in the formula for ϕ_{dT}'s are unknown and may be estimated from the data. Unbiased and consistent estimators can be obtained for $\text{MSE}_D(\bar{y}_{dT,\,\text{DIR}})=V_D(\bar{y}_{dT,\,\text{DIR}})$ using standard formulas. An approximately design unbiased estimator of $\text{MSE}_D(\bar{y}_{dT,\,\text{RSYN}})$ can be obtained using the formulas discussed in Rao (2003, section 4.2.4). In particular, we calculate the approximation:

$$\text{mse}_D(\bar{y}_{dT,\,\text{RSYN}})\approx(\bar{y}_{dT,\,\text{RSYN}}-\bar{y}_{dT,\,\text{DIR}})^2-v_D(\bar{y}_{dT,\,\text{DIR}}),$$

where mse_D and v_D stand for the estimators of the corresponding MSE_D and V_D. In particular, v_D is the ordinary design unbiased estimator of V_D. We then take its average over d, as usual, in order to obtain a more stable estimator. In fact, one problem with mse_D is that it can even be negative.

Moreover, a modified direct estimator borrowing strength over areas for estimating the regression coefficient can be used to improve estimator reliability. If auxiliary information is available, the generalized regression estimator (GREG),

$$\bar{y}_{dT,\,\mathrm{GREG}} = \bar{\mathbf{X}}'_{dT}\tilde{\beta}_w + \frac{\sum_{j \in s_{dT}} w_j e_j}{\sum_{j \in s_{dT}} w_j}, \tag{3}$$

approximately corrects the bias of the synthetic estimator by means of the term $(\sum_{j \in s_{dT}} w_j)^{-1} \sum_{j \in s_{dT}} w_j e_j$, based on regression residuals e_j.

2.2 Model-based estimators

The model-based estimators we have considered are based on the specification of explicit models for sample data which approximate a hypothetical data-generating process. As a consequence, the problem of estimating $\bar{Y}_{dT}$ comes down to one of prediction. Moreover, mean square errors and other statistical properties of estimators are usually evaluated with respect to the data-generating process. We have focused here on "unit level" models based on models relating y_{dti} to a vector of covariates $\mathbf{x}_{dti}$. The use of explicit models has several advantages, the most important of which being the opportunity to test underlying assumptions.

In the estimation of the small area means or totals of continuous variables, linear mixed models are very often used. A general linear mixed model can be described as follows:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_1 v_1 + \ldots + \mathbf{Z}_s v_s + \mathbf{e}, \tag{4}$$

where $\mathbf{y} = \{y_{dti}\}$ is the n-vector of sample observations, β a $p \times 1$ vector of fixed effects, v_j is a $q_j \times 1$ vector of random effects $(j = 1, \ldots, s)$, $\mathbf{e} = \{e_{dti}\}$ a vector of errors; $\mathbf{X}$ is assumed of rank p, $\mathbf{Z}_i = \{\mathbf{z}'_{dti}\}$ is a $n \times q_j$ matrix of incidence of the j^{th} random effect. We assume that $E(v_j) = 0, V(v_j) = \mathbf{G}_j, E(\mathbf{e}) = 0, V(\mathbf{e}) = \mathbf{R}$ (all expectations are wrt. model (4)) and that $v_1, \ldots, v_s, \mathbf{e}$ are mutually independent.

As a consequence, the variance-covariance matrix of $\mathbf{y}$ is given by:

$$\mathbf{V} = V(\mathbf{y}) = \sum_{j=1}^{s} \mathbf{Z}_j \mathbf{G}_j \mathbf{Z}'_j + \mathbf{R} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R},$$

where $\mathbf{Z} = [\mathbf{Z}_1 | \ldots | \mathbf{Z}_s]$. It is usually assumed that matrices $\mathbf{G}$, $\mathbf{R}$ depend on a k-vector of variance components ψ, and so we can write $\mathbf{V}(\psi) = \mathbf{Z}\mathbf{G}(\psi)\mathbf{Z}' + \mathbf{R}(\psi)$.

Note that at the level of individual observations, the model (4) can be rewritten as $y_{dti} = \mathbf{x}'_{dti}\beta + \mathbf{z}'_{1dti}v_1 + \ldots + \mathbf{z}'_{sdti}v_s + e_{dti}$.

We consider different specifications for linear mixed models, all of which can be viewed as special cases of the general model (4). For the sake of simplicity, we have adopted a unit level notation when describing the models considered. The first model:

$$MM1: \quad y_{dti} = \mathbf{x}'_{dti}\beta + v_d + \alpha_t + e_{dti}, \tag{5}$$

may be obtained from formula (4) setting $s = 2, q_1 = m$, $q_2 = T$ $\mathbf{G}_1 = \sigma_v^2 \mathbf{I}_m, \mathbf{G}_2 = \sigma_\alpha^2 \mathbf{I}_T, \mathbf{R} = \sigma_e^2 \mathbf{I}_n$. It includes independent area and time effects, and therefore area effects are assumed not to evolve over time. This random effects structure corresponds to the assumption of a constant covariance between units that belong to the same area, observed at two different points in time.

The second model:

$$MM2: \quad y_{dti} = \mathbf{x}'_{dti}\beta + \delta_{dt} + e_{dti}, \tag{6}$$

corresponds to the particular case in which $s = 1$, $q_1 = mq, \mathbf{G}_1 = \sigma_\delta^2 \mathbf{I}_{mq}, \mathbf{R} = \sigma_e^2 \mathbf{I}_n$. The effects of interaction between area and time are introduced, that is, we assume there are area effects which are not constant over time.

The third model:

$$MM3: \quad y_{dti} = \mathbf{x}'_{dti}\beta + v_d + \alpha_t^* + e_{dti}, \tag{7}$$

is obtained setting $s = 2, q_1 = m, q_2 = T$ $\mathbf{G}_1 = \sigma_v^2 \mathbf{I}_m, \mathbf{R} = \sigma_e^2 \mathbf{I}_n$, while the generic element $g_2(h, k)$ of $\mathbf{G}_2$ is $g_2(h, k) = \sigma_\alpha^2 \rho_\alpha^{|h-k|}, h, k = 1, \ldots, T$. There are independent area and time effects, just as in *MM*1, but the time effects are assumed to follow an AR(1) process.

The fourth model:

$$MM4: \quad y_{dti} = \mathbf{x}'_{dti}\beta + \delta_{dt}^* + e_{dti}, \tag{8}$$

is similar to model *MM*2 in that it is characterized by time varying area effects, but the further assumption that such effects follow an AR(1) process is also introduced. Thus, provided we order observations by area, with respect to the general formula (4) we have $s = 1, q_1 = mq, \mathbf{G}_1 = \mathrm{diag}(\mathbf{G}_{1d}), \mathbf{R} = \sigma_e^2 \mathbf{I}_n$ where $\mathbf{G}_{1d}, d = 1, \ldots, m$, is a $T \times T$ matrix the generic element $g_{1r}(h,k) = \sigma_\delta^2 \rho_\delta^{|h-k|}, h, k = 1, \ldots, T$.

The last specification:

$$MM5: \quad y_{dti} = \mathbf{x}'_{dti}\beta + v_d + \alpha_t + e_{dti}^*, \tag{9}$$

may be obtained by (4) setting $s = 2, q_1 = m, q_2 = T, \mathbf{G}_1 = \sigma_v^2 \mathbf{I}_m, \mathbf{G}_2 = \sigma_\alpha^2 \mathbf{I}_T$. Provided we order observations by household and time, $\mathbf{R} = \mathrm{diag}(\mathbf{R}_{di})$ where $\mathbf{R}_{di}$ is a $T \times T$ matrix whose generic element is given by $r_{di}(h, k) = \sigma_e^2 \rho_e^{|h-k|}, h, k = 1, \ldots, T$. There are independent area and time effects like in *MM*1, but errors are assumed to be autocorrelated according to an AR(1) process.

In order to evaluate the impact that using past survey waves has on the efficiency of estimator, a cross-sectional linear mixed model (*SMM*) using data from the last wave T only, has been taken as the benchmark:

$$SMM: \quad y_{dTi} = \mathbf{x}'_{dTi}\beta + \vartheta_d + e_{dTi} \tag{10}$$

with $\vartheta_d \overset{\text{IID}}{\sim} N(0, \sigma^2_\vartheta)$, $e_{jdT} \overset{\text{IID}}{\sim} N(0, \sigma^2_e)$.

This is also a particular case of (4) obtained for $s = 2$, $q_1 = m, \mathbf{G}_1 = \sigma^2_v \mathbf{I}_m$ and $\mathbf{R} = \sigma^2_e \mathbf{I}_n$. Note that (10) is the standard nested error regression model of Battese *et al.* (1988).

We also consider the corresponding random error variance linear models (see Rao 2003; section 5.5.2) obtained by replacing $\mathbf{x}'_{dti}\beta$ in formulas (5) - (10) with a general intercept θ. These models will be denoted as *MM*1*, *MM*2*, *MM*3*, *MM*4*, *MM*5*, *SMM**. All the assumptions made regarding random effects and residuals remain unchanged. This latter group of models enables us to explore the gains in efficiency obtained by exploiting the repetition of the observation on the same unit when no covariates are available at the population level.

In small area estimation, the aim is to predict scalar linear combinations of fixed and random effects of the type $\eta = \mathbf{m}'\beta + \mathbf{k}'v$ where $\mathbf{m}$ and $\mathbf{k}$ are $p \times 1$ and $q \times 1$ vectors respectively, with $q = \sum_j q_j$. The best linear unbiased predictor (BLUP) of η can be obtained by estimating (β, v) minimizing the model MSE among all linear estimators:

$$\tilde{\eta}^{\text{BLUP}}(\psi) = \mathbf{m}'\tilde{\beta}(\psi) + \mathbf{k}'\tilde{v}(\psi). \tag{11}$$

When the variance components in ψ are unknown, they may be estimated from the data and substituted into formula (11), thus obtaining "empirical BLUP" $\tilde{\eta}^{\text{EBLUP}}(\hat{\psi}) = \mathbf{m}'\hat{\beta}(\hat{\psi}) + \mathbf{k}'\hat{v}(\hat{\psi})$ (see Rao 2003, chapter 6, and Jiang and Lahiri 2006b for details).

As far as the estimation of ψ is concerned, a number of methods have been proposed in the literature, such as Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML) which assume the normality of random terms, and the MINQUE proposed by Rao (1972) which is non-parametric. In the present work we have opted for the REML method, thus assuming normality.

2.3. Measures of uncertainty associated with predictors based on linear mixed models

The difficult problem of estimating the MSE of EBLUP estimators, taking the variability of the estimated variance and covariance components into account, has been faced in the small area literature by adopting diverse approaches.

One popular method is based on the Taylor series approximation of MSE under normality (Prasad and Rao 1990; Datta and Lahiri 1999). More recently, due to the advent of high-speed computers and powerful software, resampling methods have been proposed. For instance, Butar and Lahiri (2003) introduce a parametric bootstrap method based on the assumption of normality, but analytically less onerous than the Taylor series method. Jiang *et al.* (2002) discuss a general jackknife method, which requires a distributional assumption weaker than normality (posterior linearity). We aim to empirically compare the performance of these three estimators within a context where the number of areas is moderate and the assumption of normality may not hold perfectly true. The following is a short description of the three estimation approaches.

Let us define $\text{MSE}[\tilde{\eta}^{\text{EBLUP}}(\hat{\psi})] = E(\tilde{\eta}^{\text{EBLUP}}(\hat{\psi}) - \eta)^2$, where expectation refers to model (4). It is possible to show that, under normality,

$$\begin{aligned}&\text{MSE}[\tilde{\eta}^{\text{EBLUP}}(\hat{\psi})]\\ &\quad = g_1(\psi) + g_2(\psi) + E(\tilde{\eta}^{\text{EBLUP}} - \tilde{\eta}^{\text{BLUP}})^2\end{aligned} \tag{12}$$

where $g_1(\psi) = \mathbf{k}'(\mathbf{G} - \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{G})\mathbf{k}'$ and $g_2(\psi) = \mathbf{d}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{d}$, with $\mathbf{d} = \mathbf{m}' - \mathbf{k}'\mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{X}$ (see Rao 2003, chapter 3). Using the following approximation, based on a Taylor series argument

$$\begin{aligned}&E(\tilde{\eta}^{\text{EBLUP}} - \tilde{\eta}^{\text{EBLUP}})^2\\ &\quad \approx \operatorname{tr}[(\partial\mathbf{b}'/\partial\psi)\mathbf{V}(\partial\mathbf{b}'/\partial\psi)'\bar{\mathbf{V}}(\psi)] = g_3(\psi)\end{aligned}$$

where $\mathbf{b}' = \mathbf{k}'\mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}$, a second order approximation to (12) can be found:

$$\text{MSE}[\tilde{\eta}^{\text{EBLUP}}(\hat{\psi})] \approx g_1(\psi) + g_2(\psi) + g_3(\psi). \tag{13}$$

Note that here $\approx$ means that the omitted terms are of order $o(m^{-1})$. An asymptotically unbiased estimator of (13), based on Prasad and Rao (1990), is given by

$$\text{mse}_{\text{PR}}(\tilde{\eta}^{\text{EBLUP}}) = g_1(\hat{\psi}) + g_2(\hat{\psi}) + 2g_3(\hat{\psi}). \tag{14}$$

Datta and Lahiri (1999) show that, under normality and REML or ML estimation of ψ, $\text{mse}_{\text{PR}}(\tilde{\eta}^{\text{EBLUP}})$ estimates $\text{MSE}[\tilde{\eta}^{\text{EBLUP}}(\hat{\psi})]$ with a bias of order $o(m^{-1})$.

Butar and Lahiri (2003) propose a parametric bootstrap estimation of (13) under the assumption of normality. We adapt their estimator to the models we are analysing, assuming the following bootstrap model:

$$\begin{aligned}&\text{i)}\quad \mathbf{y}^* \mid \mathbf{v}^* \overset{\text{ind}}{\sim} N[\mathbf{X}\hat{\beta} + \mathbf{Z}\mathbf{v}^*, \mathbf{R}(\hat{\psi})]\\ &\text{ii)}\quad \mathbf{v}^* \overset{\text{ind}}{\sim} N[\mathbf{0}, \mathbf{G}(\hat{\psi})]\end{aligned} \tag{15}$$

where $\mathbf{v} = (v'_1, \ldots v'_s)'$. The parametric bootstrap is then used twice, once to estimate the first two terms of (13), thus correcting the bias of $g_1(\hat{\psi}) + g_2(\hat{\psi})$, and once to estimate $g_3(\psi)$.

The following estimator of (13) is proposed:

$$\begin{aligned}\text{mse}_{\text{BL}}(\tilde{\eta}^{\text{EBLUP}}) &= 2[g_1(\hat{\psi}) + g_2(\hat{\psi})] - E_B[g_1(\hat{\psi}^*) + g_2(\hat{\psi}^*)] \\ &\quad + E_B[\tilde{\eta}(\mathbf{y}, \hat{\beta}(\hat{\psi}^*), \hat{\psi}^*) - \tilde{\eta}(\mathbf{y}, \hat{\beta}(\hat{\psi}), \hat{\psi})] \quad (16)\end{aligned}$$

where $\hat{\psi}^*$ is the same as $\hat{\psi}$ except that it is calculated on $\mathbf{y}^*$ instead of $\mathbf{y}$, and E_B is the expected value with regard to the bootstrap model (15).

The bootstrap estimator (16) does not require the analytical derivation of $g_3(\hat{\psi})$ which can be rather laborious when **G** and **R** have complicated structures.

Jiang *et al.* (2002) introduced a general jackknife estimator for the variance of empirical best predictors in linear and non-linear mixed models with *M*-estimation. In the problem we are investigating here, the estimator they propose can be written as:

$$\begin{aligned}\text{mse}_{\text{JLW}}(\tilde{\eta}^{\text{EBLUP}}) &= g_1(\hat{\psi}) - \frac{m-1}{m}\sum_{j=1}^{m}[g_1(\hat{\psi}_{-j}) - g_1(\hat{\psi})] \\ &\quad + \frac{m-1}{m}\sum_{j=1}^{m}(\tilde{\eta}_{-j}^{\text{EBLUP}} - \tilde{\eta}^{\text{EBLUP}})^2 \quad (17)\end{aligned}$$

where $\hat{\psi}_{-j}$ is the estimate of ψ calculated by using all data except those from the j^{th} area. Similarly, $\tilde{\eta}_{-j}^{\text{EBLUP}} = \tilde{\eta}^{\text{EBLUP}}(\mathbf{y}_{-j}, \hat{\beta}(\hat{\psi}_{-j}), \hat{\psi}_{-j})$.

It is worth pointing out that, on the basis of the simulation results reported in Jiang *et al.* (2002), mse_{JLW} is deemed to be more robust than mse_{PR} with regard to departures from the assumption of normality, which can also be expected to be crucial for mse_{BL}.

3. The simulation study based on the European Household Community Panel data

The target population of the ECHP survey consists of all resident households of a large subset of the EU member countries. Although general survey guidelines were issued by Eurostat, a certain degree of flexibility was allowed, so there are some differences in the sampling design across countries. As far as Italy is concerned, the survey is based on a stratified two stage design, in which strata were formed by grouping the PSUs (municipalities) according to geographic region (NUTS2) and demographic size. For more details of the survey, see Eurostat (2002).

The ECHP deals with unit non-response, sample attrition and new entries using weighting and imputation. As attrition could lead to biased estimates of income if it does not appear at random, the effect of poverty on dropout propensity has been investigated (Rendtel, Behr and Sisto 2003; Vandecasteele and Debels 2004), and the results of these studies show that in the case of some countries, including Italy, this effect disappears under the control of weighting variables.

We have focused our attention on the eight ECHP waves available for Italy (1994-2001). Given that our aim is to assess whether the use of several successive observations of the same household could be profitable for the purposes of small area estimation, we have overlooked the problem of attrition and only considered those households that participate to the survey for all waves.

Our target variable is disposable, post-tax household income at the time of the last wave (2001). In studies of poverty and inequality, income is often equivalized according to an equivalence scale in order to avoid comparison problems caused by differences in the composition of households. We consider the widely-used modified OECD scale, also adopted by Eurostat (2002) in its publications on income, poverty and social exclusion. According to this scale, equivalized income is calculated by dividing disposable household income by the number (k) of "equivalent adults", defined as $k = 1 + 0.5a + 0.3c$, where a is the number of adults other then the "head of the household" and c is the number of children aged 13 or less. In general, the equivalized income can be perceived as the amount of income that an individual, living alone, should dispose of in order to attain the same level of economic wellbeing he/she enjoys in his/her household.

Of the many covariates available in the bountiful ECHP questionnaire, we have chosen only those for which area means were available from the 2001 Italian Census results. Thus the chosen covariates are: the percentage of adults; the percentage of employed; the percentage of unemployed; the percentage of people with a high/medium/low level of education in the household; household typology (presence of children, presence of aged people, *etc.*); the number of rooms per-capita and the tenure status of the accommodation (rented, owned *etc.*).

As we have said, the aim of this paper is to compare the performance of different estimators in the controlled environment of a simulation exercise. A number of works in the literature have compared small area estimators using Monte Carlo experiments in which samples are drawn from synthetic populations based either on Censuses (Falorsi, Falorsi and Russo 1994; Ghosh *et al.* 1996) or on the replication of sample units' records (Falorsi, Falorsi and Russo 1999; Lehtonen, Särndal and Veijanen 2003; Singh, Mantel and Thomas 1994). Since household income is not measured by the Italian Census (nor is it given by the results of other Censuses conducted by EU countries), we treated the ECHP survey data as the pseudo-population and then draw samples using stratified probability proportional to size sampling, the size variable being given by survey weights. This solution may not be as good as that of using data from a real Census population, but it is hopefully more realistic than generating population values of household income from a parametric model.

Monte Carlo samples of 1,000 (roughly 15% of the actual ECHP sample size) were drawn from the synthetic population by stratified random sampling without replacement, with strata given by the 21 NUTS2 regions. Thus these regions are treated as planned domains (as in the ECHP) for which sample size in the small areas is established beforehand, so that the sampling fractions reflect the over-sampling of smaller regions exactly as they do in the actual ECHP sampling design. The region-specific sample sizes we obtained range from 14 to 112, being on average equal to 48. Therefore in our simulation $n = 1{,}000$, the number of small areas corresponds to that of the Italian Regions $(m = 21)$ and the number of points in time corresponds to the ECHP available waves $(T = 8)$.

The distribution of equivalized household income in our pseudo-population (that is the distribution obtained by weighted estimation from the ECHP sample data) is characterized by an overall mean of 22,547 Euros and a coefficient of variation of 0.59. The distribution is positively skewed (even though skewness is not extreme: skewness coefficient $\gamma_1 = \mu_3 / \sigma^3 \cong 2.5$) and kurtosis $(\kappa = \mu_4 / \sigma^4 \cong 14.3)$. The difference between mean and median is 9% of the mean. An interesting feature is given by the large disparities among administrative regions (that are the small areas of interest in our study). The mean of the equivalized household income ranges from 16,604 to 27,011, that is the most affluent area has a mean equivalent income 62% higher than the poorest one. Also the coefficient of variation (ranging from 0.28 to 0.84), skewness (γ_1 ranging from 0.1 to 4.6) and kurtosis (κ ranging from -0.7 to 32.9) show that the distribution of our target variable is quite a bit different in different areas.

To motivate the selected specifications of the random effects part of the considered linear mixed models (see section 2.2), an approach often recommended in textbooks (see Verbeke and Molenberghs 2000, chapter 9) has been followed: first we fit a standard OLS regression to our data using all available covariates; then we analyse the resulting residuals as a guide to identifying the random effects. This preliminary analysis has been conducted separately on several random samples of size 1,000, drawn up according to the replication design described above.

The adjusted R^2 of the OLS regression is close to 0.35 in every observed sample. This rather low figure is the result of the nature of the phenomenon under study (household income is not easy to predict), the information contained in the survey and the constraint represented by the need to include only those covariates for which the population total can be obtained from the Census.

Figure 1 contains "box and whiskers" plots of the residuals by area and wave constructed for one of the Monte Carlo sample (very similar findings may be observed in every sample). Analysis of the plots suggests that there is within-area and within-wave correlation, and thus the need to specify models including area and wave effects. From an analysis of residuals, it is less clear whether the inclusion of interaction effects (that is time varying area effects) would be beneficial or not.

Moreover the residuals show a degree of autocorrelation, the average of the autocorrelation coefficient calculated over all individual residual histories being 0.27. Even though this autocorrelation level is not very high, for the sake of completeness we decided to also take into consideration models with autocorrelated errors or random effects. After having tested various different autocorrelation structures (ARMA(p, q), General Linear, *etc*.), we found that the auto-regressive process of order 1 provides the best fit to our data.

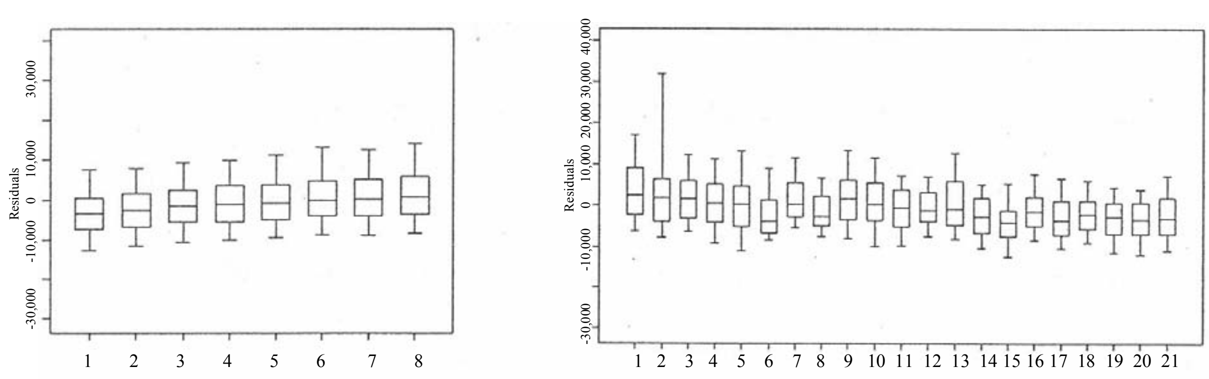

Figure 1 Box and whiskers plot of residuals by wave (left) and area (right)

The apparent skewness of residuals also suggests that the normality assumption for errors does not hold exactly. We maintain this assumption for all the models we specify, and we use REML estimators for variance components. In fact, we may expect departures from normality to have a slight impact on point values of predictors. BLUP formulas can be derived without normality; moreover, there are sound reasons for us to expect REML (and ML) estimators of ψ to perform well even if normality does not hold (see Jiang, 1996, for details). Departures from normality may have more a serious impact on MSE estimation, and this is a problem we are going to be looking at in section 4.2 below.

4. Results

4.1 Point estimators

All computations involved in the simulation exercise described in section 3 were carried out using SAS version 9.1 for Windows. EBLUP estimators are obtained using Proc MIXED, and the generation of samples is based on Proc SURVEYSELECT.

Given that the primary goal of Small Area Estimation is the precise estimation of area-specific parameters, we first evaluated how well the described estimators perform when predicting individual area values. Moreover, we also evaluated the amount of over-shrinkage connected with each estimator. In fact, small area estimates should reflect (at least approximately) the variability in the underlying area parameters taken as a whole.

We note that our simulation experiment is aimed at evaluating design-based properties of the estimators, that is, the population from which the random samples are generated is held fixed.

For the evaluation of the estimators' performance, we adopted an approach that is commonly found in the literature (see Rao 2003; section 7.2.6), using two indicators, the Average Absolute Relative Bias (AARB) and the Average Relative Mean Square Error (ARMSE):

$$\begin{aligned} \text{AARB} &= m^{-1}\sum_{d=1}^{m}\left|R^{-1}\sum_{r=1}^{R}\left(\frac{\tilde{y}_{dT(r)}}{\bar{Y}_{dT}}-1\right)\right| \\ \text{ARMSE} &= m^{-1}\sum_{d=1}^{m}\left\{R^{-1}\sum_{r=1}^{R}\left(\frac{\tilde{y}_{dT(r)}}{\bar{Y}_{dT}}-1\right)^2\right\} \end{aligned} \qquad (18)$$

where $\tilde{y}_{dT(r)}$ is the estimate for area d, time T and replicated sample r, while $\bar{Y}_{dT}$ is the population mean being estimated. Note that AARB measures the bias of an estimator, whereas ARMSE measures its accuracy. The number of replications R is set at 500, a figure large enough to obtain stable Monte Carlo estimates of expected values and variances, frequently used in simulation studies on small area estimation (Heady, Higgins and Ralphs 2004; EURAREA Consortium 2004).

The gain in efficiency connected to each small area estimator is evaluated using the ratio of its ARMSE to the ARMSE of certain estimators we use as benchmarks. In particular, all estimators are compared with the weighted mean $\bar{y}_{dT,\,\text{DIR}}$ and we denote this ratio as AEFF_{Dir}. Moreover, EBLUP estimators associated with models (5) - (9), which use data from previous waves, are compared with the EBLUP estimator associated with the cross-sectional model (10), in order to assess the gain in efficiency deriving from the use of past waves. In this case the ratio is denoted as $\text{AEFF}_{\text{Sect}}$.

As far as the evaluation of the degree of shrinkage is concerned, we have compared the empirical standard deviation of population area values:

$$\text{ESD} = \sqrt{m^{-1}\sum_{d=1}^{m}(\bar{Y}_{dT}-\bar{\bar{Y}}_{T})^2},$$

where $\bar{\bar{Y}}_{T}$ is the mean of the population values of the m areas at time T, with the empirical standard deviation of the estimated area values, which in the case of a simulation study is given by:

$$\text{esd} = R^{-1}\sum_{r=1}^{R}\left[\sqrt{m^{-1}\sum_{d=1}^{m}(\tilde{y}_{dT(r)}-\bar{\tilde{y}}_{T(r)})^2}\right],$$

where $\bar{\tilde{y}}_{T(r)}$ is the mean of the estimated values for the m areas at time T in the simulation run r. The comparison is carried out using the indicator

$$\text{RESD} = \frac{\text{esd}}{\text{ESD}} - 1 \qquad (19)$$

which tells us how the empirical standard deviation associated with one estimator differs from that of the population.

Table 1 contains the percentage values of AARB, ARMSE, AEFF and RESD obtained for the direct estimator, the design-based estimators given in (2) and (3) and the EBLUP estimators derived from models (5) - (10).

All estimators perform significantly better than $\bar{y}_{dT,\,\text{DIR}}$ in terms of ARMSE, leading to less than 100% AEFF_{Dir} values. We can also see that design-based estimators are worse than EBLUP estimators in terms of ARMSE, and that the gain in efficiency demonstrated by AEFF_{Dir} is particularly high in some cases (in excess of 50%). This result highlights the superior accuracy of the model-based estimators in question.

Table 1 Performance indicators - auxiliary information is available

Model	AARB%	ARMSE%	$AEFF_{Dir}$%	$AEFF_{Sect}$%	RESD%
DIR	0.0	0.787	100.0	-	15.6
COMP	2.7	0.552	70.1	-	-9.8
GREG	0.2	0.543	68.2	-	10.0
SMM	2.3	0.377	47.7	100.0	-8.7
*MM*1	3.1	0.358	45.3	95.0	2.4
*MM*2	2.4	0.427	54.1	113.4	-4.4
*MM*3	2.6	0.380	48.3	101.2	4.7
*MM*4	2.6	0.429	54.2	113.6	-8.0
*MM*5	2.9	0.318	40.4	84.7	-7.7

The most reliable EBLUP estimator is the one associated with the *MM*5 model, with independent area and time effects and residuals autocorrelated according to an AR(1) process, leading to a gain in efficiency of about 60% compared with the direct estimator. This is followed by the EBLUP estimator associated with model *MM*1, which differs from the previous one only because of the absence of autocorrelated residuals.

In terms of bias, the GREG estimator gives the smallest value of AARB, as would be expected (Särndal, Swensson and Wretman 1992, chapter 7; Veijanen, Lehtonen and Särndal 2005). This is followed by the remaining estimators, all of which reveal a similar value for AARB. Of the EBLUP estimators, those associated with the *MM*1 and *MM*5 models are more efficient in terms of ARMSE, but they are slightly more biased than the one associated with the *SMM*. This is probably due to the fact that we limit our evaluation of performance to the last wave; for this data subset we would expect the fit of the regression underlying *SMM*, based on the last wave only, to be better than the one based on the whole data set. As far as EBLUP estimators are concerned, the $\text{AEFF}_{\text{Sect}}$ column shows how the gain in efficiency of the predictors, based on borrowing strength over time, is positive in some cases and negative in others. Models *MM*2 and *MM*4 (see formulas (6) and (8)), where effects of interaction between area and time are present, are apparently inadequate because the predictors associated with both models perform rather poorly. The performance of the predictor associated with *MM*3 (see (7)) is also slightly worse than that of the predictor associated with the cross-sectional model: this rather surprising result is probably due to the low number of waves, which does not allow for an effective estimation of the correlation coefficient between consecutive time effects.

As we have already said, the estimator associated with model *MM*5 is the one that performs the best: it is considerably more efficient than the one associated with *SMM*, with an $\text{AEFF}_{\text{Sect}}$ of roughly 85% representing a gain in efficiency of about 15% due to consideration of more than one wave. The EBLUP estimator associated with *MM*1 also turns out to be more efficient than the one associated with *SMM*, but in this case the gain is one of only 5%. These results confirm the fact that household level data at several consecutive points in time may be employed, via certain kinds of longitudinal model, to produce more efficient estimates.

Moving on to the indicator for shrinkage reported in the last column of the table, we can see that the direct estimator overestimate the standard deviation of the population of area means, by 15%. The same effect, albeit somewhat attenuated, is observed for the GREG estimator, whose standard deviation is over-inflated by 10%. On the contrary, the COMP estimator tends to “shrink” the estimates towards the centre of the distribution, leading to a reduction in the standard deviation of area means of about 10% with respect to the population. These results are in line with those obtained by other authors comparing the same kinds of estimator (Heady *et al*. 2004; Spjøtvoll and Thomsen 1987). The results obtained for EBLUP estimators are more encouraging, as the calculated percentage difference is always less then 10% in absolute terms. Hence, in this respect all EBLUP estimators seem to be acceptable. Moreover, we may expect that the BLUP estimators are under-dispersed compared to the corresponding population parameters. In this case, the indicator RESD assumes positive values for some longitudinal EBLUP estimators because it is calculated only on the last wave, while longitudinal models are aimed to predict $m \times T$ parameters.

Table 2 summarizes the results regarding those EBLUP estimators associated with random error variance models, as described in the last paragraph of section 2. When no auxiliary variables are included in the models, the advantage of “borrowing strength” over time and area is singled out independently of the advantage associated with covariates.

As expected, the improvements in efficiency measured by AEFF_{Dir} are smaller than those shown in Table 1, although the reductions in ARMSE remain significant. The ranking of those predictors associated with the various random effects specification remains the same as the one presented in Table 1, the predictor associated with the $MM5^*$ model resulting the most efficient, as shown by ARMSE%. The gain in efficiency associated with this latter estimator compared with the direct estimator is about 43%.

Table 2 Performance indicators - auxiliary information is unavailable

Model	AARB%	ARMSE%	AEFF_{Dir}%	$\text{AEFF}_{\text{Sect}}$%	RESD%
SMM^*	2.7	0.575	72.8	100.0	-7.6
$MM1^*$	2.9	0.556	70.3	96.6	7.5
$MM2^*$	2.8	0.639	80.8	111.0	-3.0
$MM3^*$	3.7	0.574	72.6	99.7	8.6
$MM4^*$	3.5	0.691	87.2	119.8	-6.7
$MM5^*$	3.0	0.445	56.2	77.2	-6.3

With regard to bias, the EBLUP estimators obtained from those models with no covariates tend to be more biased than the corresponding ones with covariates.

The analysis of the $\text{AEFF}_{\text{Sect}}$ column shows that the reduction in ARMSE allowed for by some of those models borrowing strength over time, is larger than in the case where covariates are included, as it reaches 22% in the best example of the $MM5^*$ model.

This last result is really encouraging. In fact, within the context of Small-Area Estimation, the absence of any known totals of covariates in the population can be very limiting when trying to obtain reliable estimates. The observed ARMSE reduction connected to the consideration of more waves in a panel survey show that estimates may be improved "borrowing strength" over time, when it is not possible to exploit auxiliary information.

With regard to the results of shrinkage, they may be considered acceptable also in this case, and one can see a relationship between the results obtained for EBLUP estimators derived from analogous models with or without covariates.

4.2 Comparing different estimators of the MSE of EBLUP estimators

In section 2.3 we reviewed three different estimators of the MSE associated with EBLUP estimators. In this section we are going to compare the performances of these three estimators using our simulation exercise. Given that we are focusing on MSE estimation rather than a comparison of EBLUP estimators derived from different models, we only consider the predictor associated with model *MM*5, which emerged as the best performer in the previous section.

Let us denote the predictor of $\bar{Y}_{dT}$ with $\tilde{\eta}_{dT}^{\text{EBLUP}}$ and its mean square error as $\text{MSE}(\tilde{\eta}_{dT}^{\text{EBLUP}})$. The following estimator:

$$\text{mse}_{\text{ACT}}(\tilde{\eta}_{dT}^{\text{EBLUP}}) = \frac{1}{R}\sum_{r=1}^{R}(\tilde{\eta}_{dT(r)}^{\text{EBLUP}} - \tilde{\tilde{\eta}}_{dT}^{\text{EBLUP}})^2 + (\tilde{\tilde{\eta}}_{dT}^{\text{EBLUP}} - \bar{Y}_{dT})^2,$$

where $\tilde{\eta}_{dT(r)}^{\text{EBLUP}}$ is $\tilde{\eta}_{dT}^{\text{EBLUP}}$ calculated on the r^{th} replicated sample and $\tilde{\tilde{\eta}}_{dT}^{\text{EBLUP}} = R^{-1}\sum_{r=1}^{R}\tilde{\eta}_{dT(r)}^{\text{EBLUP}}$, will be used as benchmark for the comparison of the performance of the mean square error estimators described in section 2.3, because the true mean squared error is not known.

As in the case of point estimators, all computations are done using SAS. To determine the Prasad-Rao estimator (14), the output of Proc MIXED's ESTIMATE statement is used with the option KENWARDROGER activated. The sum $g_1(\hat{\psi}) + g_2(\hat{\psi})$ is obtained from the output of Proc MIXED. The KENWARDROGER option allows for the calculation of an MSE inflation factor, described in Kenward and Rogers (1986), which is equivalent to $2g_3(\hat{\psi})$ (see also Rao 2003, section 6.2.7).

The estimator $\text{mse}_{\text{BL}}(\tilde{\eta}_{dT}^{\text{EBLUP}})$ is re-sampling based. Hence the evaluation of its performance with respect to a Monte Carlo exercise requires the implementation of two nested simulations: for each $r\,(r = 1, \ldots, R)$, we run the R_{BOOT} replications needed to approximate expectations with respect to the bootstrap model. To limit the computational burden, we set $R_{\text{BOOT}} = 150$. Butar and Lahiri (2003) propose an analytical approximation of mse_{BL}, but only for models that are not as complex as the one in question.

For both $\text{mse}_{\text{BL}}(\tilde{\eta}_{dT}^{\text{EBLUP}})$ and $\text{mse}_{\text{JLW}}(\tilde{\eta}_{dT}^{\text{EBLUP}})$, we have prepared *ad-hoc* SAS codes using the output of Proc MIXED as inputs.

In order to compare the three MSE estimators, we employ the same measures used to evaluate the performance of point estimators, AARB and ARMSE. As there is usually some concern about the under-estimation of MSE estimators, we are also interested in the sign of any bias associated with the estimators in question. Therefore, in the case of MSE estimators we do not only calculate the average of the absolute values of the estimates obtained for the bias in each region (AARB), but also the average of these estimates without the absolute value (AARB'), so as to better understand whether the given estimators indeed tend to under-evaluate the MSE or not. Hence the calculated measures are:

$$\text{AARB} = m^{-1}\sum_{d=1}^{m}\left|R^{-1}\sum_{r=1}^{R}\left(\frac{\text{mse}_*(\tilde{\eta}_{dT}^{\text{EBLUP}})}{\text{mse}_{\text{ACT}}(\tilde{\eta}_{dT}^{\text{EBLUP}})}-1\right)\right|$$

$$\text{AARB}' = m^{-1}\sum_{d=1}^{m}\left\{R^{-1}\sum_{r=1}^{R}\left(\frac{\text{mse}_*(\tilde{\eta}_{dT}^{\text{EBLUP}})}{\text{mse}_{\text{ACT}}(\tilde{\eta}_{dT}^{\text{EBLUP}})}-1\right)\right\},$$

$$\text{ARMSE} = m^{-1}\sum_{d=1}^{m}\left\{R^{-1}\sum_{r=1}^{R}\left(\frac{\text{mse}_*(\tilde{\eta}_{dT}^{\text{EBLUP}})}{\text{mse}_{\text{ACT}}(\tilde{\eta}_{dT}^{\text{EBLUP}})}-1\right)^2\right\}$$

where the symbol * refers to the considered estimation procedures, that are PR, BL, JLW. Results of the comparisons based on $R = 500$ MC iterations are reported in Table 3.

Table 3 Performance of MSE estimators of $\tilde{\eta}_{dT}^{\text{EBLUP}}$ under model MM5

Estimator	AARB	AARB'	ARMSE
mse_{PR}	0.378	-0.383	0.238
mse_{BL}	0.377	-0.318	0.228
mse_{JLW}	0.337	0.036	0.261

In terms of ARMSE and AARB, the three estimators behave similarly, with no particular one emerging as clearly better than the other two. Nonetheless, the AARB' column clearly shows that mse_{PR} and mse_{BL} systematically underestimate MSE_{ACT}, whereas mse_{JLW} does not. This is probably due to the failure of the normality assumption for error terms. In fact, as we foresaw in section 3, equivalized income is a positively skewed variable, and the regression residuals **e** also appear to be so. Normality is a crucial assumption in the derivation of mse_{PR} and mse_{BL}, while mse_{JLW} could be expected to be more robust in this respect. Our findings are consistent with the theory predictions and simulation results described in Jiang *et al.* (2002). Although Bell (2001) noted that mse_{JLW} may be negative for some data set because of the bias correction, this never happens in our simulations. For all replicated data set we have that the second term in (17) gives a positive, and in most cases substantial contribution to the estimate of the MSE. A discussion of modifications of (17) when it returns negative values can be found in Jiang and Lahiri (2006b).

To conclude then, in the case of the present problem, mse_{JLW} emerges as the most appropriate of the three measures for estimating $\text{MSE}(\tilde{\eta}_{dT}^{\text{EBLUP}})$. This finding could be of importance for any application of normality-based linear mixed models theory to data set in which normality assumptions for error terms do not hold exactly.

We replicated the simulation exercise also for the cross-section model without covariates SMM^*, that is often considered in simulations aimed at the comparison of different estimation methods. To this end we note that for this model the ratio $\hat{\sigma}_e^2/\hat{\sigma}_v^2$ is around 12, leading to a EBLUP predictor characterized by $\gamma_i = \hat{\sigma}_v^2 n_i(\hat{\sigma}_v^2 n_i + \hat{\sigma}_e^2)^{-1}$ ranging from 0.54 to 0.9. We note also that some, but not all, areas are characterized by the presence of outliers (skewness coefficient γ_1 ranges from 0.1 to 4.6).

In this setting MSE estimators show a behavior quite different form that illustrated in the case of model *MM*5. Results are shown in Table 4.

Table 4 Performance of MSE estimators of $\tilde{\eta}_{dT}^{\text{EBLUP}}$ under model SSM*

Estimator	AARB	AARB'	ARMSE
mse_{PR}	0.449	0.262	0.503
mse_{BL}	0.376	0.213	0.376
mse_{JLW}	0.354	0.149	0.335

All estimators overestimate the actual MSE, although mse_{JLW} overestimates less than the other two. From a detailed analysis of results related to individual areas, we have the values of AARB′ (that represents the most apparent difference with the results of Table 3) is driven by severe overestimation of actual MSE in areas characterized by the lowest levels of skewness and kurtosis. For these areas $\hat{\sigma}_e^2$ largely overstates actual variation in the data, thus leading to overestimation of $g_1(\sigma_v^2, \sigma_e^2)$. This is likely to be due to the fact that the failure of normality (the excess of kurtosis) causes the overestimation of σ_e^2. This problem did not appear in the case of model *MM*5 because of the presence of covariates and the AR(1) modeling of individual residuals.

5. Concluding remarks and further developments

The results obtained show that, in general, EBLUP estimators derived from unit level linear mixed model specifications that "borrow strength over time", as well as over areas, provide a significant gain in efficiency compared with both the direct estimator and with other commonly-used design based estimators such as the optimal composite estimator and the GREG estimator. Moreover, the mean squared error of some of the longitudinal EBLUP estimators in question is considerably lower, on average over the areas, than that of the analogous cross-sectional EBLUP estimators. Among the model specifications used to derive EBLUP estimators, those with independent time and area effects, whether inclusive of the autocorrelation of residuals or not, appear the most efficient, offering a gain in efficiency of about 55-60% compared with the direct estimator. These results also hold when covariates are removed; in fact, they offer the chance to obtain reliable small area estimates even in the absence of covariates, provided that repeated observations of the same unit at several points in time are available. Besides the shrinkage effect connected to EBLUP estimators appears moderate, reducing the need for ensemble or multiple estimation (Rao 2003, Chapter 9). With regard to estimation of the MSE of the small area estimators in question, we noted that the jackknife estimator provides the best results being correct, on average, over the areas and thus more robust to any departure from the standard assumptions of linear mixed models. This finding may be of importance to all applications of normality-based linear mixed models theory to data set in which normality assumptions do not hold exactly, as in the case of income data.

Acknowledgements

Research was partially funded by Miur-PRIN 2003 "Statistical analysis of changes of the Italian productive sectors and their territorial structure", coordinator Prof. C. Filippucci. The work of Enrico Fabrizi was partially supported by the grants 60FABR06 and 60BIFF04, University of Bergamo.

We thank ISTAT for kindly providing the data used in this work.

References

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error component model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36.

Bell, W. (2001). Discussion with "Jackknife in the Fay-Herriott model with an example", *Proceeding of the Seminar on Funding Opportunity in Survey Research*, 98-104.

Betti, G., and Verma, V. (2002). Non-monetary or lifestyle deprivation, in EUROSTAT (2002). *Income, Poverty Risk and Social Exclusion in the European Union*, Second European Social Report, 87-106.

Butar, F., and Lahiri, P. (2003). On measures of uncertainty of empirical bayes small-area estimators. *Journal of Statistical Planning and Inference*, 112, 63-76

Datta, G.S., and Lahiri, P. (1999). A unified measure of uncertainty of estimated best linear predictors in small area estimation problems. *Statistica Sinica*, 10, 613-627.

Datta, G.S., Lahiri, P. and Maiti, T. (2002). Empirical bayes estimation of median income of four-person families by state using time series and cross-sectional data. *Journal of Statistical Planning and Inference*, 102, 83-97.

Datta, G.S., Lahiri, P., Maiti, T. and Lu, K.L. (1999). Hierarchical bayes estimation of unemployment rates for the States of the U.S. *Journal of the American Statistical Association*, 94, 1074-1082.

EURAREA CONSORTIUM (2004). EURAREA. Enhancing Small Area Estimation Techniques to meet European Needs, Project Reference Volume, downloadable at http://www.statistics.gov.uk/eurarea/download.asp.

EUROSTAT (2002). European social statistics - Income, poverty and social exclusion. 2nd report.

Falorsi, P.D., Falorsi, S. and Russo, A. (1994). Empirical comparison of small area estimation methods for the Italian Labour Force Survey. *Survey Methodology*, 20, 171-176.

Falorsi, P.D., Falorsi, S. and Russo, A. (1999). Small area estimation at provincial level in the Italian Labour Force Survey. *Journal of the Italian Statistical Society*, 1, 93-109.

Ghosh, M., Nangia, N. and Kim, D. (1996). Estimation of median income of four-person families: A bayesian time series approach. *Journal of the American Statistical Association*, 91, 1423-1431.

Heady, P., Higgins, N. and Ralphs, M. (2004). Evidence-Based Guidance on the Applicability of Small Area Estimation Techniques. Paper presented at the European Conference on Quality and Methodology in Official Statistics, Mainz, Germany, May, 24-26.

Jiang, J. (1996). REML estimation: Asymptotic Behavior and Related Topics. *The Annals of Statistics*, 24, 255-286.

Jiang, J., Lahiri, P. and Wan, S.M. (2002). A unified jackknife theory for empirical best prediction with *M*-estimation. *The Annals of Statistics*, 30, 1782-1810.

Jiang, J., and Lahiri, P. (2006a). Estimation of finite population domain means - A model-assisted empirical best prediction approach. *Journal of the American Statistical Association*, 101, 301-311.

Jiang, J., and Lahiri, P. (2006b). Mixed model prediction and small area estimation. Editor's invited discussion paper, *Test*, 15, 1, 1-96.

Kenward, M.G., and Roger, J.H. (1986). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53, 983-997.

Lehtonen R., Särndal C.-E. and Veijanen A. (2003) The effect of model choice in estimation for domains, including small domains, *Survey Methodology*, 29, 1, 33-44.

Pfeffermann, D. (2002). Small area estimation – New developments and directions. *International Statistical Review*, 70, 125-143.

Prasad, N.G.N., and Rao, J.N.K. (1990). The estimation of mean squared errors of small area estimators. *Journal of the American Statistical Association*, 85, 163-171.

Rao, C.R. (1972). Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association*, 67, 112-115.

Rao, J.N.K. (2003). *Small Area Estimation*, New York: John Wiley & Sons, Inc.

Rao, J.N.K., and You, M. (1994). Small area estimation by combining time series and cross-sectional data. *Canadian Journal of Statistics*, 22, 511-528.

Rendtel, U., Behr, A. and Sisto, J. (2003). Attrition effect in the European Community household panel, CHINTEX PROJECT, European Commission.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Singh, A.C., Mantel, H.J. and Thomas, B.W. (1994). Time series EBLUPS for small areas using survey data. *Survey Methodology*, 20, 1, 33-43.

Spjøtvoll, E., and Thomsen, I. (1987). Application of some empirical bayes methods to small area statistics. *Bulletin of the International Statistical Institute*, 4, 435-450.

Vandecasteele, L., and Debels, A (2004). Modelling attrition in the European Community Household Panel: The effectiveness of weighting, 2nd International Conference of ECHP Users, EPUNet 2004, Berlin, June 24-26.

Veijanen, A., Lehtonen, R. and Särndal, C.-E. (2005). The Effect of Model Quality on Model-Assisted and Model-Dependent Estimators of Totals and Class Frequency for Domains, paper presented at SAE2005 Conference, Challenges in Statistics Production for Domains and Small Areas, August, 28-31 2005, Jyväskylä, Finland.

Verbeke, G., and Molenberg, H. (2000). *Linear Mixed Models for Longitudinal Data*, New York: Springer-Verlag.