Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 66(1) 2012 75-89

Two Area-level Time Models for Estimating Small Area Poverty Indicators

M.D. Esteban[1], D. Morales[1*], A. Pérez[2] and L. Santamaría[1]
[1]*Centro de Investigación Operativa*
[2]*Departamento de Estudios Económicos y Financieros,*
Universidad Miguel Hernández de Elche, Spain
Received 30 April 2011; Revised 16 September 2011; Accepted 16 September 2011

SUMMARY

This paper deals with small area estimation of poverty indicators. Small area estimators of these quantities are derived from time-dependent area-level linear mixed models. As appropriate auxiliary variables are not always available in the survey data on living conditions, the proposed models using only aggregated data are a good alternative to the unit-level models. The mean squared errors are estimated by explicit formulas. Two simulation experiments designed to analyze the behavior of the introduced estimators are carried out. An application to real data from the Spanish Living Conditions Survey is also given.

Keywords : Area-level models, Small area estimation, Time dependency, Poverty indicators.

1. INTRODUCTION

The European Union (EU) has put in place an strategy aiming at making a decisive impact on the eradication of poverty and also declared 2010 as the year of struggle against poverty. One topic of high interest is therefore the estimation and dissemination of poverty, inequality and life condition indicators. Such indicators can greatly assist in monitoring living conditions and in guiding the implementation of policies that aim at improving the living conditions in the EU member states. Given the growing social, demographic and economic problems, the research community, policy makers and practitioners place great emphasis on the development of efficient, effective and reliable indicators and on the collection of high quality data on life conditions not only at national level but also at regional and at lower geographical levels. The objective of this work is the estimation of these indicators in the Spanish provinces by using a model-based approach.

In most European countries, the estimation of poverty is done by using the Living Conditions Survey (LCS) data. The Spanish LCS (SLCS) uses a stratified two-stage design within each Autonomous Community. As most provinces have a very small sample size, the direct estimates at that level have a low accuracy. The problem is thus that domain sample sizes are too small to carry out direct estimations. This situation may be treated by using small area estimation techniques. Small Area Estimation (SAE) is a part of the statistical science that combines survey sampling and finite population inference with statistical models. See a description of this theory in the monograph of Rao (2003), or in the reviews of Ghosh and Rao (1994), Rao (1999), Pfeffermann (2002) and more recently Jiang and Lahiri (2006).

In this paper we use two time-dependent area-level linear mixed models to obtain small area estimates of poverty indicators. The estimates of the model parameters are obtained by using the residual maximum likelihood (REML) estimation method. These estimates are then used to construct empirical best linear unbiased predictors of poverty indicators by sex of the Spanish provinces. Estimation of the mean squared error (MSE)

[*]*Corresponding author* : D. Morales
E-mail address : d.morales@umh.es

of model-based estimators is an important issue that has no easy solution. In this paper we follow Prasad and Rao (1990) to introduce an approximation of the MSE and the corresponding MSE estimator. The rest of the paper is organized as follows. Section 2 describes the data set and the estimation problem of interest. Section 3 introduces the considered area-level time models and the corresponding model-based estimators of poverty indicators. Section 4 presents two simulation experiments to compare the small area estimators obtained under the two introduced models. Finally, Section 5 applies the proposed methodology to data from the SLCS. The target is to estimate poverty indicators by sex in the Spanish provinces.

2. POVERTY INDICATORS AND DATA DESCRIPTION

Let us consider a finite population P_t partitioned into D domains P_{dt} at the time period t, and denote their sizes by N_t and N_{dt}, $d = 1, ..., D$. Let z_{dtj} be an income variable measured in all the units of the population and let z_t be the poverty line, so that units j in the domain d with $z_{dtj} < z_t$ are considered as poor at the time period t. The main goal of this section is to estimate the poverty incidence (proportion of individuals under poverty) and the poverty gap in Spanish domains. These two measures belongs to the FGT family proposed by Foster *et al.* (1984), given by

$$\overline{Y}_{\alpha,dt} = \frac{1}{N_{dt}} \sum_{j=1}^{N_{dt}} y_{\alpha;dtj} ,$$

$$\text{where} \quad y_{\alpha,dtj} = \left(\frac{z_t - z_{dtj}}{z_t} \right)^{\alpha} I(z_{dtj} < z_t), \tag{2.1}$$

$I(z_{dtj} < z_t) = 1$ if $z_{dtj} < z_t$ and $I(z_{dtj} < z_t) = 0$ otherwise. The proportion of units under poverty in the domain d and the period t is thus $Y_{0;dt}$ and the poverty gap is $Y_{1;dt}$.

Following the standards of the Spanish Statistical Office, the Poverty Threshold is fixed as the 60% of the median of the normalized incomes in Spanish households. This threshold is an exogenous and overall country estimate and therefore it is considerably stable. The aim of normalizing the household income is to adjust for the varying size and composition of households. The total number of normalized household members is calculated by giving a weight of 1.0 to the first adult, 0.5 to the second and each subsequent person aged 14 and over and 0.3 to each child aged under 14 in the household. The *normalized* size of the household h is the sum of the weights assigned to each person. So the total number of normalized household members is

$$H_{dth} = 1 + 0.5(H_{dth \geq 14} - 1) + 0.3H_{dth < 14}$$

where $H_{dth \geq 14}$ is the number of people aged 14 and over and $H_{dth < 14}$ is the number of children aged under 14. The normalized net annual income of a household is obtained by dividing its net annual income by its normalized size. The Spanish poverty thresholds (in euros) in 2004-06 are $z_{2004} = 6098.57$, $z_{2005} = 6160.00$ and $z_{2006} = 6556.60$ respectively. These are the z_t-values appearing in (2.1).

We use data from the SLCS corresponding to years 2004-2006. The SLCS is the Spanish version of the "European Statistics on Income and Living Conditions" (EU-SILC), which is one of the statistical operations that have been harmonized for EU countries. We consider $D = 104$ domains obtained by crossing 52 provinces with 2 sexes. The quartiles of the distribution of the domain sample sizes are $q_0 = 17$, $q_1 = 170$, $q_2 = 293$, $q_3 = 640$, $q_4 = 2113$ in 2004, 13, 149, 251, 530, 1494 in 2005 and 18, 129, 233, 481, 1494 in 2006, so they are too small for employing direct estimators in all the domains.

The direct estimator of the total

$$Y_{\alpha,dt} = \sum_{j=1}^{N_{dt}} y_{\alpha,dtj} \text{ is}$$

$$\hat{Y}_{\alpha,dt}^{dir} = \sum_{j \in S_{dt}} w_{dtj} y_{\alpha,dtj}$$

where S_{dt} is the domain sample at the time period t and the w_{dtj} 's are the official calibrated sampling weights which take into account for non response. The estimated domain size is

$$\hat{N}_{dt}^{dir} = \sum_{j \in S_{dt}} w_{dtj}$$

The direct estimator of the domain mean $\overline{Y}_{\alpha,dt}$ is $\overline{y}_{\alpha,dt} = \hat{Y}_{\alpha,dt}^{dir} \big/ \hat{N}_{dt}^{dir}$. These estimates are used as responses in the area-level time model. Their design-based variances can be approximated by

$$\hat{V}_\pi\left(\hat{Y}^{dir}_{\alpha,dt}\right) = \sum_{j\in S_{dt}} w_{dtj}(w_{dtj}-1)(y_{\alpha,dtj}-\bar{y}_{\alpha,dt})^2$$

and $\hat{V}_\pi(\bar{y}_{\alpha,dt}) = \hat{V}\left(\hat{Y}^{dir}_{\alpha,dt}\right)\Big/\hat{N}^2_{dt}$. (2.2)

The last formulas are obtained from Särndal *et al.* (1992), pp. 43, 185 and 391, with the simplifications $w_{dtj} = 1/\pi_{dtj}$, $\pi_{dtj,dtj} = \pi_{dtj}$ and $\pi_{dti,dtj} = \pi_{dti}\pi_{dtj}$, $i \neq j$ in the second order inclusion probabilities.

As we are interested in the cases $y_{dtj} = y_{\alpha,dtj}$, $\alpha = 0, 1$, we select the direct estimates of the poverty incidence and poverty gap at domain d and time period t (i.e. $\bar{y}_{0;dt}$ and $\bar{y}_{1;dt}$ respectively) as target variables for the time dependent area-level models 0 and 1 described in Section 3. The considered auxiliary variables are the known domain means of the category indicators of the following variables:

- AGE: Age groups are *age1-age5* for the intervals ≤ 15, 16 – 24, 25 – 49, 50 – 64 and ≥ 65.
- EDUCATION: Highest level of education completed, with 4 categories denoted by *edu*0 for Less than primary education level, *edu*1 for Primary education level, *edu*2 for Secondary education level and *edu*3 for University level.
- CITIZENSHIP: with 2 categories denoted by *cit*1 for Spanish people and *cit*2 for Non Spanish people.
- LABOR: Labor situation with 4 categories taking the values *lab*0 for Below 16 years, *lab*1 for Employed, *lab*2 for Unemployed and *lab*3 for Inactive.

3. AREA-LEVEL LINEAR TIME MODEL

This section introduces the two area-level time models that will be used on the estimation of domain poverty indicators. The main model (model 1) in this paper is

$$y_{dt} = \mathbf{x}_{dt}\beta + u_{dt} + e_{dt}, \quad d = 1, ..., D, \quad t = 1, ..., m_d, \quad (3.1)$$

where y_{dt} is a direct estimator of the indicator of interest for area d and time instant t, $\mathbf{x}_{dt}$ is a vector containing the aggregated (population) values of p auxiliary variables, the random vectors $(u_{d1}, ..., u_{dm_d})$, $d = 1, ..., D$, are i.i.d. AR(1), with variance and auto-correlation parameters σ_u^2 and ρ respectively, the errors e_{dt}'s are independent N(0, σ_{dt}^2) with known σ_{dt}^2's, and the u_{dt}'s and the e_{dt}'s are independent. Along the paper we also consider a simpler model (model 0) obtained by restricting model (3.1) to $\rho = 0$. For the sake of brevity we skip formulas for model 0. In matrix notation, the model 1 is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad (3.2)$$

where
$$\mathbf{y} = \underset{1\le d\le D}{\text{col}}(\mathbf{y}_d),\ \mathbf{y}_d = \underset{1\le t\le m_d}{\text{col}}(y_{dt}),$$
$$\mathbf{u} = \underset{1\le d\le D}{\text{col}}(\mathbf{u}_d),\ \mathbf{u}_d = \underset{1\le t\le m_d}{\text{col}}(u_{dt}),$$
$$\mathbf{e} = \underset{1\le d\le D}{\text{col}}(\mathbf{e}_d),\ \mathbf{e}_d = \underset{1\le t\le m_d}{\text{col}}(e_{dt}),$$
$$\mathbf{X} = \underset{1\le d\le D}{\text{col}}(\mathbf{X}_d),\ \mathbf{X}_d = \underset{1\le d\le m_d}{\text{col}}(\mathbf{x}_{dt}),$$
$$\mathbf{x}_{dt} = \underset{1\le i\le p}{\text{col}'}(x_{dti}),\ \beta = \underset{1\le i\le p}{\text{col}}(\beta_i),$$
$$\mathbf{Z} = \mathbf{I}_{M\times M},\ M = \sum\nolimits_{d=1}^{D} m_d .$$

We assume that $\mathbf{u} \sim N(\mathbf{0},\mathbf{V}_u)$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{V}_e)$ are independent with covariance matices

$$\mathbf{V}_u = \sigma_u^2\Omega(\rho),\ \Omega(\rho) = \underset{1\le d\le D}{\text{diag}}(\Omega_d(\rho)),$$
$$\mathbf{V}_e = \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_{ed}),\ \mathbf{V}_{ed} = \underset{1\le t\le m_d}{\text{diag}}\left(\sigma_{dt}^2\right),$$

where the variances σ_{dt}^2 are known and

$$\Omega_d = \Omega_d(\rho) = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \cdots & \rho^{m_d-2} & \rho^{m_d-1} \\ \rho & 1 & \ddots & & \rho^{m_d-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{m_d-2} & & \ddots & 1 & \rho \\ \rho^{m_d-1} & \rho^{m_d-2} & \cdots & \rho & 1 \end{pmatrix}_{m_d\times m_d}$$

The Best Linear Unbiased (BLU) estimators and predictors of β and $\mathbf{u}$ are

$$\tilde{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

and $$\tilde{\mathbf{u}} = \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\beta}), \quad (3.3)$$

where

$$\text{var}(\mathbf{y}) = \mathbf{V} = \sigma_u^2 \underset{1\le d\le D}{\text{diag}} (\Omega_d(\rho)) + \mathbf{V}_e$$

$$= \underset{1\le d\le D}{\text{diag}} \left(\sigma_u^2\Omega_d(\rho) + V_{ed}\right) = \underset{1\le d\le D}{\text{diag}} (\mathbf{V}_d).$$

As the variance components are unknown, $\tilde{\beta}$ and $\tilde{\mathbf{u}}$ are not real estimators and predictors. In this paper, variance componentes are estimated by the the restricted (residual) maximum likelihood (REML) method. By plugging these estimators in Formulas (3.3) one obtain the Empirical BLUE (EBLUE) and BLUP (EBLUP) respectively, and they are denoted by $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$.

The loglikelihood of the REML method is

$$l_{reml}\left(\sigma_u^2, \rho\right) = -\frac{M-p}{2} \log 2\pi + \frac{1}{2} \log |\mathbf{X}'\mathbf{X}|$$

$$-\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2} \mathbf{y}'\mathbf{P}\mathbf{y},$$

where

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}, \mathbf{PVP} = \mathbf{P}, \mathbf{PX} = 0.$$

Let $\theta = (\theta_1, \theta_2) = \left(\sigma_u^2, \rho\right)$, $\mathbf{V}_1 = \frac{\partial \mathbf{V}}{\partial \sigma_u^2} = \underset{1\le d\le D}{\text{diag}} (\Omega_d(\rho))$,

$$\dot{\Omega}_d = \frac{\partial \Omega_d(\rho)}{\partial \rho} \text{ and } \mathbf{V}_2 = \frac{\partial \mathbf{V}}{\partial \rho} = \sigma_u^2 \underset{1\le d\le D}{\text{diag}} \left(\dot{\Omega}_d(\rho)\right),$$

then

$$\mathbf{P}_a = \frac{\partial P}{\partial \theta_a} = -\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_a}\mathbf{P} = -\mathbf{P}\mathbf{V}_a\mathbf{P},\ a = 1, 2.$$

By taking partial derivatives of l_{reml} with respect to θ_a and θ_b, we get the scores

$$S_a = \frac{\partial l_{reml}}{\partial \theta_a} = -\frac{1}{2}\text{tr}(\mathbf{P}\mathbf{V}_a) + \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{y},\ a = 1, 2.$$

By taking again partial derivatives with respect to θ_a and θ_b, taking expectations and changing the sign, we get the Fisher information matrix components

$$\text{F}_{ab} = \frac{1}{2}\text{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_b),\ a, b = 1, 2.$$

To calculate the REML estimate we apply the Fisher-scoring algorithm with the updating formula

$$\theta^{k+1} = \theta^k + \mathbf{F}^{-1}(\theta^k)\mathbf{S}(\theta^k),$$

where **S** and **F** are the column vector of scores and the Fisher information matrix respectively. The REML estimator of β is

$$\hat{\beta} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}.$$

The asymptotic distributions of the REML estimators of θ and β are

$$\hat{\theta} \sim N_3(\theta, \mathbf{F}^{-1}(\theta)),\ \hat{\beta} \sim N_p(\beta, (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}).$$

Asymptotic confidence intervals, at the level $1-\alpha$, for θ_a and β_j are

$$\hat{\theta}_a \pm z_{\alpha/2}v_{aa}^{1/2},\ a = 1, 2,\ \hat{\beta}_j \pm z_{\alpha/2}q_{jj}^{1/2},\ j = 1, ..., p,$$

where $\hat{\theta} = \theta^{\kappa}$, $\mathbf{F}^{-1}(\theta^{\kappa}) = (v_{ab})_{a,b=1,2}$, $(\mathbf{X}'\mathbf{V}^{-1}(\theta^{\kappa})\mathbf{X})^{-1} = (q_{ij})_{i,j=1,...,p}$, κ is the final iteration in the Fisher-scoring algorithm and z_α is the α-quantile of the N(0, 1) distribution. If $\hat{\beta}_j = \beta_0$ is observed, then the asymptotic p-value for testing $H_0 : \beta_j = 0$ is

$$p = 2P_{H_0}\left(\hat{\beta}_j > |\beta_0|\right) = 2P\left(\text{N}(0,1) > |\beta_0| / \sqrt{q_{jj}}\right).$$

We are interested in predicting $\mu_{dt} = \mathbf{x}_{dt}\beta + u_{dt}$ with the EBLUP $\hat{\mu}_{dt} = \mathbf{x}_{dt}\hat{\beta} + \hat{u}_{dt}$. If we do not take into account the error, e_{dt}, this is equivalent to predict $y_{dt} = \mathbf{a}'\mathbf{y}$, where $\mathbf{a} = \underset{1\le l\le D}{\text{col}} (\delta_{dl}\mathbf{a}_l)$ and $\mathbf{a}_l = \underset{1\le k\le m_l}{\text{col}} (\delta_{tk})$. Note that **a** is a vector with a "1" in the cell $t + \sum_{l=1}^{d-1} m_l$ and 0's in the remaining cells. The population mean $\bar{Y}_{dt}$ is estimated by means of $\hat{\bar{Y}}_{dt}^{eblup} = \hat{\mu}_{dt}$. Following Prasad and Rao (1990) and Das *et al.* (2004), the mean squared error (MSE) of $\hat{\bar{Y}}_{dt}^{eblup}$ takes the form

$$MSE\left(\hat{\bar{Y}}_{dt}^{eblup}\right) = g_1(\theta) + g_2(\theta) + g_3(\theta),$$

where

$$\theta = \left(\sigma_u^2, \rho\right),$$

$$g_1(\theta) = \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a},$$

$$g_2(\theta) = [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{ZTZ}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}$$
$$[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{ZTZ}'\mathbf{a}],$$
$$g_3(\theta) \approx \text{tr}\left\{(\nabla\mathbf{b}')\,\mathbf{V}(\nabla\mathbf{b}')'E\left[(\hat{\theta}-\theta)(\hat{\theta}-\theta)'\right]\right\}$$

and $\mathbf{Q} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$, $\mathbf{T} = \mathbf{V}_u - \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{V}_u$, $\mathbf{b}' = \mathbf{a}'\mathbf{Z}\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}$.

The estimator of $MSE\left(\hat{\bar{Y}}_{dt}^{eblup}\right)$ is

$$mse\left(\hat{\bar{Y}}_{dt}^{eblup}\right) = g_1(\hat{\theta}) + g_2(\hat{\theta}) + 2g_3(\hat{\theta}). \quad (3.4)$$

After straightforward algebra, the following domain-decomposed formulas are obtained

$$g_1(\theta) = \sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{a}_d - \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d$$

$$g_2(\theta) = [\,\mathbf{a}_d'\mathbf{X}_d - \sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d + \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d\,]\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)^{-1}$$
$$[\,\mathbf{X}_d'\mathbf{a}_d - \sigma_u^2\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{a}_d + \sigma_u^4\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\,]$$

$$g_3(\theta) \approx \text{tr}\left\{\begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22}\end{pmatrix}\begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22}\end{pmatrix}^{-1}\right\},$$

where F_{ab} is the generic component of the REML Fisher information matrix, which was calculated when obtaining the updating equation of the Fisher-scoring algorithm and

$$q_{11} = \frac{\partial\mathbf{b}'}{\partial\sigma_u^2}\underset{1\le l\le D}{\text{diag}}(\mathbf{V}_l)\left(\frac{\partial\mathbf{b}'}{\partial\sigma_u^2}\right)'$$
$$= \mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d - 2\sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d + \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d$$

$$q_{12} = \frac{\partial\mathbf{b}'}{\partial\sigma_u^2}\underset{1\le l\le D}{\text{diag}}(\mathbf{V}_l)\left(\frac{\partial\mathbf{b}'}{\partial\rho}\right)'$$
$$= \sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d - \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d - \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d + \sigma_u^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d$$

$$q_{22} = \frac{\partial\mathbf{b}'}{\partial\rho}\underset{1\le l\le D}{\text{diag}}(\mathbf{V}_l)\left(\frac{\partial\mathbf{b}'}{\partial\rho}\right)'$$
$$= \sigma_u^4\mathbf{a}_d'\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d - 2\sigma_u^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d + \sigma_u^8\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d$$

4. SIMULATION EXPERIMENTS

Two simulation experiments for analyzing the behavior of the EBLUP and its mean squared error estimator are presented in this section. The scope of the simulations is to investigate when it is worthwhile and what is gained when using the more complicated model 1 with correlation parameter instead of the simplified model 0 restricted to $\rho = 0$. For $d = 1, ..., D$, $t = 1, ..., m_d$, the explanatory and target variables are

$$x_{dt} = (b_{dt} - a_{dt})U_{dt} + a_{dt},\ U_{dt} = \frac{t}{m_d + 1},$$
$$a_{dt} = 1,\ b_{dt} = 1 + \frac{1}{D}\ (m_d(d-1) + t),$$
$$y_{dt} = \beta_1 + \beta_2 x_{dt} + u_{dt} + e_{dt},\ \beta_1 = 0,\ \beta_2 = 1,$$

where $e_{dt} \sim \text{N}\left(0, \sigma_{dt}^2\right)$,

$$\sigma_{dt}^2 = \alpha_0 + \frac{(\alpha_1 - \alpha_0)(m_d(d-1) + t - 1)}{M - 1},\ \alpha_0 = 0.8$$

and $\alpha_1 = 1.2$. For $d = 1, ..., D$, the random vectors $(u_{d1}, ..., u_{dm_d})$ are generated from the joint normal distribution of an AR(1) stochastic process with parameters σ_u^2 and ρ. This is done as follows :

$$u_{d1} = (1 - \rho^2)^{-1/2}\,\varepsilon_{d1},\ u_{dt} = \rho u_{dt-1} + \varepsilon_{dt},\quad t = 2, ..., m_d,$$

where $\varepsilon_{dt} \sim \text{N}(0, \sigma_u^2)$, $d = 1, ..., D$; $t = 1, ..., m_d$, and $\sigma_u^2 = 1$.

The first simulation experiment is dedicated to investigate the gain of efficiency achieved by the EBLUP based on model 1 (EBLUP1) with respect to the EBLUP based on model 0 (EBLUP0) as a function of the correlation parameter ρ. Data are simulated from model 0 when $\rho = 0$ and are simulated from model 1 when $\rho = 0.25, 0.5, 0.75$. The experiment has the following steps :

1. For ρ = 0, 1/4, 1/2, 3/4, repeat $K = 10^4$ times ($k = 1, ..., K$)

   1.1. Generate a sample of size $m = \sum_{d=1}^{D} m_d$. Calculate $\mu_{dt}^{(k)} = \beta_1 + \beta_2 x_{dt} + u_{dt}^{(k)}$.

   1.2. Calculate $\hat{\beta}_1^{(k,0)}, \hat{\beta}_2^{(k,0)}, \hat{\sigma}_u^{2(k,0)}$, and EBLUP0 $\hat{\mu}_{dt}^{(k,0)}$ by using REML method under model 0.

   1.3. Calculate $\hat{\beta}_1^{(k,1)}, \hat{\beta}_2^{(k,1)}, \hat{\sigma}_u^{2(k,1)}, \hat{\rho}^{(k,1)}$ and EBLUP1 $\hat{\mu}_{dt}^{(k,1)}$ by using REML method under model 1.

2. For $d = 1, ..., D$; $t = 1, ..., m_d$, calculate

$$BIAS_{dt}^{(a)} = \frac{1}{K}\sum_{k=1}^{K}\left(\hat{\mu}_{dt}^{(k,a)} - \mu_{dt}^{(k)}\right)$$

$$MSE_{dt}^{(a)} = \frac{1}{K}\sum_{k=1}^{K}\left(\hat{\mu}_{dt}^{(k,a)} - \mu_{dt}^{(k)}\right)^2, \; a = 0, 1$$

$$BIAS^{(a)} = \frac{1}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} BIAS_{dt}^{(a)}$$

$$MSE^{(a)} = \frac{1}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} MSE_{dt}^{(a)}, \; a = 0, 1.$$

Mean squared errors $MSE^{(0)}$ and $MSE^{(1)}$ are presented in the Table 4.1 (left). Biases $BIAS^{(0)}$ and $BIAS^{(1)}$ are presented in the Table 4.1 (right). In the Fig. 4.1 the MSE_{dm_d}-values are plotted for $D = 100$, $m_d = 5$ and $\rho = 0$ (top-left), $\rho = 0.25$ (top-right), $\rho = 0.5$ (bottom-left) and $\rho = 0.75$ (bottom-right). They are labeled by $MSE_d^{(0)}$ and $MSE_d^{(1)}$ respectively. In the Fig. 4.2 the $BIAS_{dm_d}$-values are plotted for $D = 100$, $m_d = 5$ with the same configuration as in the Fig. 4.1. They are labeled by $BIAS_d^{(0)}$ and $BIAS_d^{(1)}$ respectively.

When the true model is model 0, the best results in MSE are obtained if we work all the time under the assumption that $\rho = 0$. However if we use the EBLUP derived under the incorrect model 1 the increase of MSE is almost negligible. This can be appreciated in the two first rows of the Table 4.1 (left) and on the Fig. 4.1. If we look at the bias, no increment is observed for incorrectly using model 1.

When the true model is model 1 and the correlation parameter is small ($\rho = 0.25$), there is almost no difference in MSE or BIAS by using the true model or the incorrect model 0. If the correlation parameter is of medium size ($\rho = 0.5$) there is a clear increase of MSE and BIAS by using the incorrect model. Finally if the correlation parameter is high ($\rho = 0.75$) the use of the incorrect model produce severe increases of MSE and BIAS.

Table 4.1. *MSE*'s (left) and *BIAS*'s (right) of EBLUP0 and EBLUP1 for $D = 100$

		m_d				m_d			
ρ	a	2	5	10	20	2	5	10	20
0.00	0	0.5086	0.5026	0.5003	0.4996	0.00078	–0.00011	0.00053	–0.00001
0.00	1	0.5138	0.5046	0.5014	0.5001	0.00078	–0.00011	0.00053	–0.00001
0.25	0	0.5263	0.5204	0.5185	0.5176	0.00079	–0.00011	0.00053	–0.00002
0.25	1	0.5214	0.5074	0.5026	0.5007	0.00078	–0.00011	0.00052	–0.00001
0.50	0	0.6263	0.6189	0.6183	0.6193	–0.00020	–0.00133	0.00196	0.00103
0.50	1	0.5457	0.5133	0.5052	0.5015	–0.00021	–0.00132	0.00193	0.00104
0.75	0	1.2021	1.1903	1.1930	1.1971	–0.00030	–0.00130	0.00197	0.00106
0.75	1	0.5953	0.5230	0.5029	0.4935	–0.00032	–0.00129	0.00192	0.00106

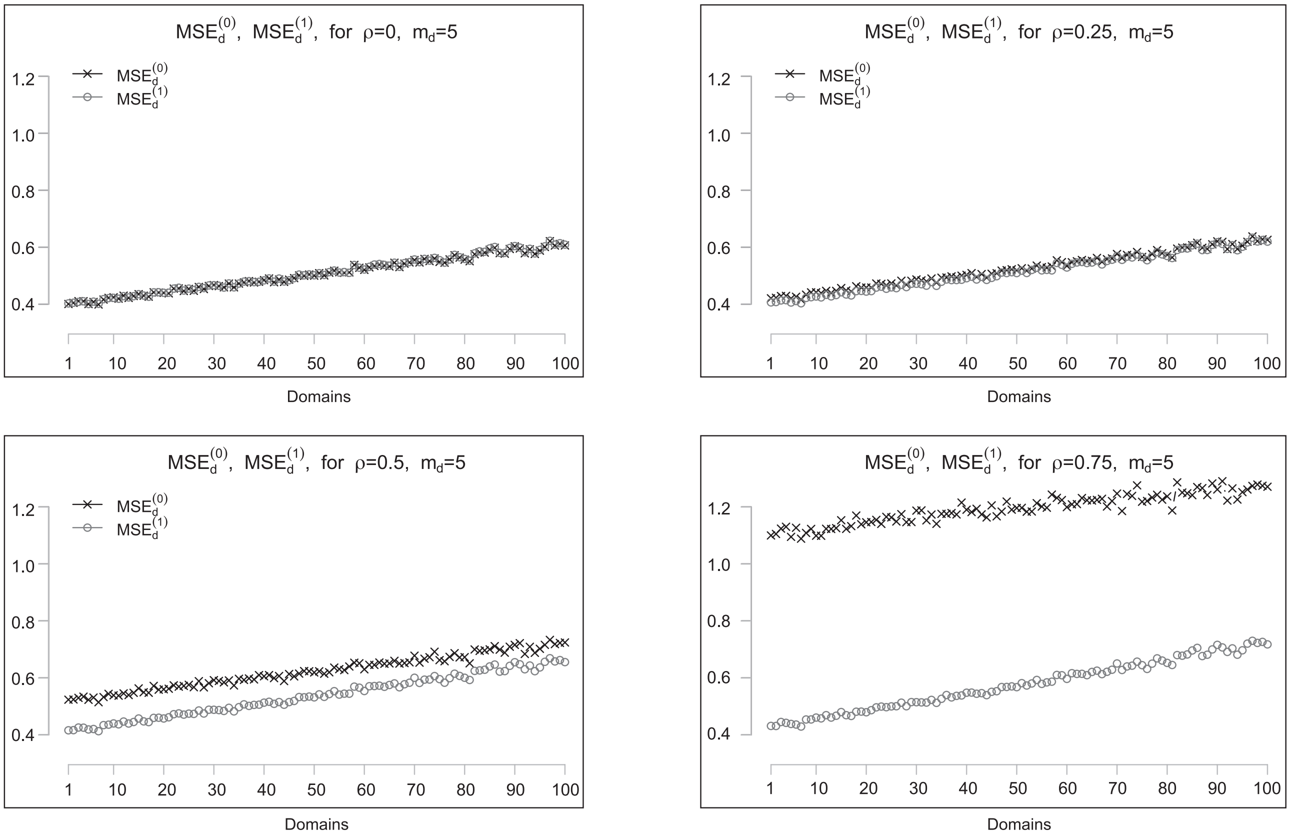

Fig. 4.1. MSE_{dm_d}'s of EBLUP0 and EBLUP1 for D = 100, m_d = 5.

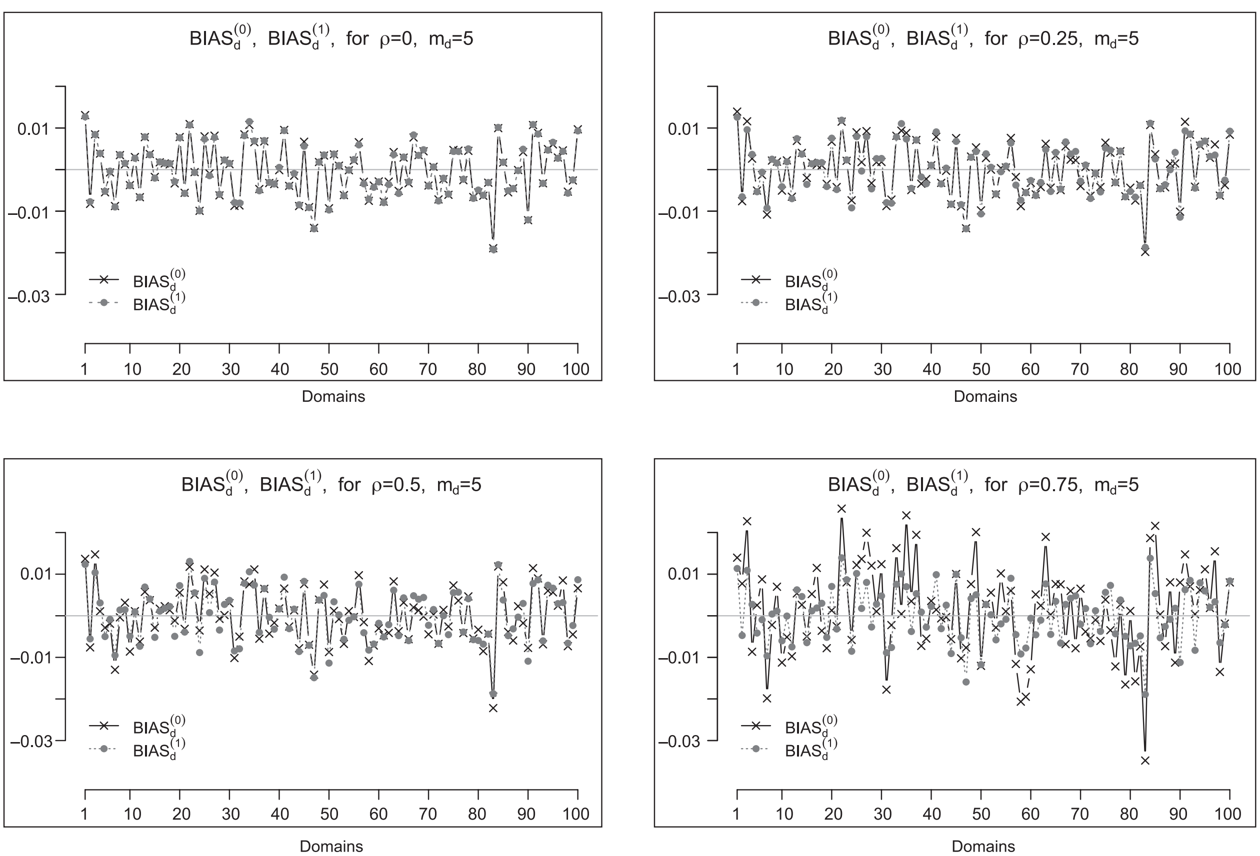

Fig. 4.2. $BIAS_{dm_d}$'s of EBLUP0 and EBLUP1 for D = 100, m_d = 5.

The second simulation experiment takes the empirical MSEs, $MSE_{dt}^{(a)}$, obtained in the first experiment and includes the following additional steps:

1.4 Calculate $mse\left(\hat{\mu}_{dt}^{(k,0)}\right)$ and $mse\left(\hat{\mu}_{dt}^{(k,1)}\right)$ by applying formula (3.4) restricted to model 0 or unrestricted respectively.

3. For $d = 1, ..., D$; $t = 1, ..., m_d$, calculate

$$B_{dt}^{(a)} = \frac{1}{K}\sum_{k=1}^{K}\left(mse\left(\hat{\mu}_{dt}^{(k,a)}\right) - MSE_{dt}^{(a)}\right)$$

$$E_{dt}^{(a)} = \frac{1}{K}\sum_{k=1}^{K}\left(mse\left(\hat{\mu}_{dt}^{(k,a)}\right) - MSE_{dt}^{(a)}\right)^2, \; a = 0, 1,$$

$$B^{(a)} = \frac{1}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} B_{dt}^{(a)}, \; E^{(a)} = \frac{1}{D}\sum_{d=1}^{D}\sum_{t=1}^{m_d} E_{dt}^{(a)}, \quad a = 0, 1.$$

Mean squared errors $E^{(0)}$ and $E^{(1)}$ are presented in the Table 4.2 (left). Biases $B^{(0)}$ and $B^{(1)}$ are presented in the Table 4.2 (right). For $D = 100$ and $m_d = 5$, in the Fig. 4.3 the B_{dm_d}-values are plotted on the top for $\rho = 0$ and $\rho = 0.75$ and the E_{dm_d}-values are plotted in the bottom for the same values of ρ. We observe that in the case $\rho = 0$ there is no difference between working under the true model 0 or under the incorrect model 1. On the other hand, if $\rho = 0.75$ then we get higher bias and mean squared error in the estimation of the MSE of the EBLUP by working under model 0. Again we conclude that if true model is model 1, then there is a loss of efficiency by using model 0.

5. ESTIMATION OF POVERTY INDICATORS

We first consider the linear model

$$\bar{y}_{dt} = \bar{\mathbf{X}}_{dt}\beta + u_{dt} + e_{dt}, \; d = 1, ..., D$$

where $\bar{y}_{dt}$ is the direct estimate of the poverty indicator and $\bar{\mathbf{X}}_{dt}$ is the $1 \times p$ vector containing the population (aggregated) mean values of all the categories (except the last one) of the explanatory variables described in Section 2. Random effects errors are assumed to follow the distributional assumptions of model (3.1) either restricted to $\rho = 0$ (model 0) or without this restriction (model 1). As some of the explanatory variables where not significative, the starting models where simplified to include only the auxiliary variables appearing in Tables 5.1 and 5.2. As the 90% confidence intervals for ρ are 0.8214±0.063 and 0.6862±0.094 for $\alpha = 0$ and $\alpha = 1$ respectively, we recommend model 1 in both cases. Regression parameters and their corresponding p-values are presented in Tables 5.1 and 5.2.

The signs of the regression coefficients appearing in Tables 5.1 and 5.2 give information about how the auxiliary variables are influencing the poverty indicators. We recall that the auxiliary variables are the

Table 4.2. *E*'s (left) and *B*'s (right) of EBLUP0 and EBLUP1 for $D = 100$

		m_d				m_d			
ρ	a	2	5	10	20	2	5	10	20
0.00	0	0.00347	0.00194	0.00140	0.00112	–0.00118	–0.00015	0.00014	–0.00038
0.00	1	0.00350	0.00194	0.00140	0.00112	–0.00086	–0.00018	0.00013	–0.00038
0.25	0	0.00350	0.00202	0.00150	0.00122	–0.00118	–0.00006	–0.00007	–0.00023
0.25	1	0.00352	0.00203	0.00146	0.00118	–0.00116	–0.00047	–0.00007	–0.00059
0.50	0	0.00365	0.00242	0.00195	0.00168	–0.00139	–0.00028	–0.00052	–0.00030
0.50	1	0.00398	0.00222	0.00161	0.00132	–0.00198	–0.00109	–0.00073	–0.00113
0.75	0	0.00465	0.00395	0.00361	0.00336	–0.00307	–0.00209	–0.00232	–0.00190
0.75	1	0.00513	0.00243	0.00173	0.00141	–0.00405	–0.00225	–0.00165	–0.00162

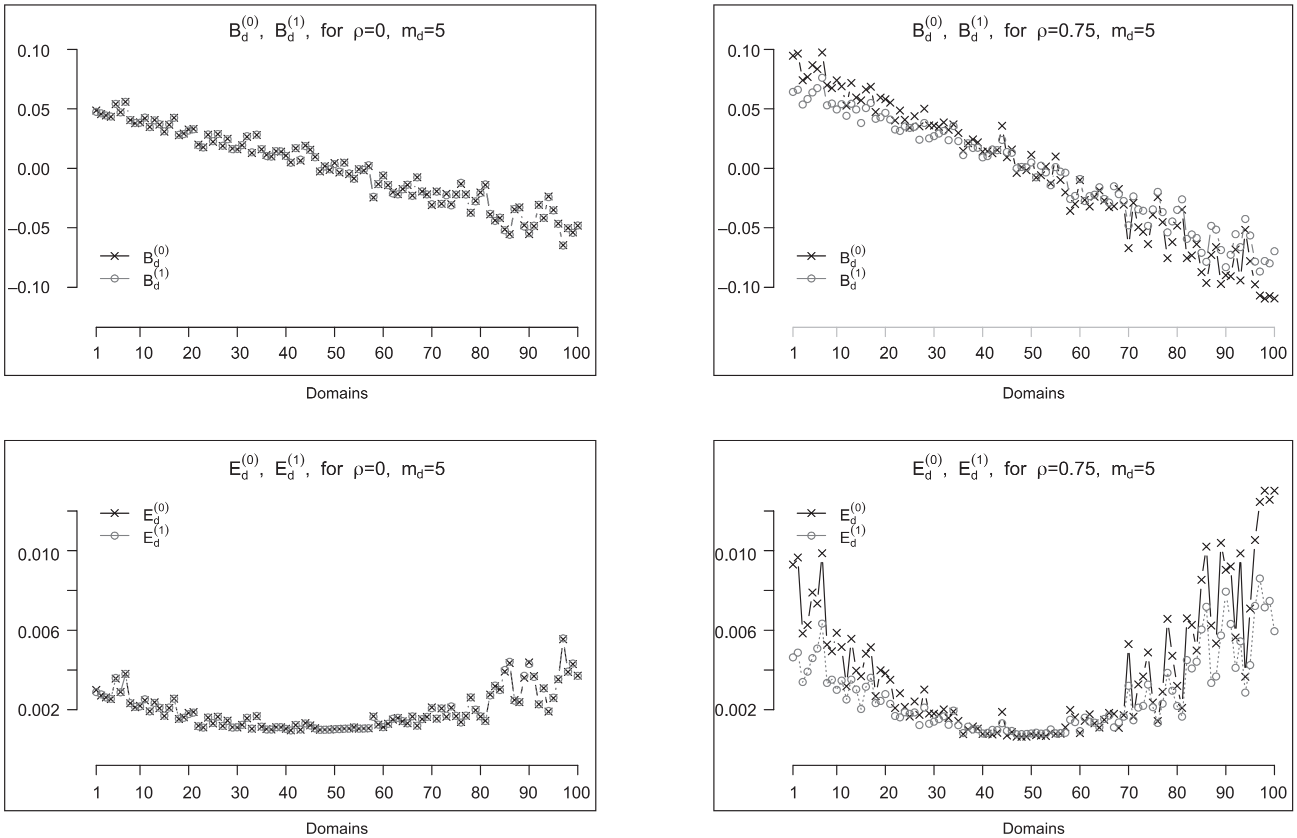

Fig. 4.3. B_{dm_d}'s (top) and E_{dm_d}'s (bottom) of EBLUP0 and EBLUP1 for D = 100, m_d = 5.

Table 5.1. Regression parameters and p-values for $\alpha = 0$.

model 0	age3	age4	age5	edu1	edu2	cit1	lab2
β	−0.3686	−2.8841	−0.3649	0.7470	0.3977	0.4808	0.9206
p	0.0143	$<10^{-35}$	0.0142	$<10^{-10}$	0.0005	$<10^{-11}$	$<10^{-5}$
model 1	age3	age4	age5	edu1	edu2	cit1	lab2
β	−0.209	−2.509	−0.075	0.398	0.200	0.531	0.262
p	0.262	$<10^{-17}$	0.687	0.008	0.196	$<10^{-7}$	0.310

Table 5.2. Regression parameters and p-values for $\alpha = 1$.

model 0	age3	age4	edu0	edu1	cit1	lab2
β	−0.1217	−0.6419	0.3863	0.1132	0.1186	0.2936
p	0.0215	$<10^{-6}$	$<10^{-5}$	$<10^{-5}$	$<10^{-4}$	0.0021
model 1	age3	age4	edu0	edu1	cit1	lab2
β	−0.1455	−0.5574	0.4187	0.0753	0.1254	0.2028
p	0.0131	$<10^{-4}$	$<10^{-4}$	0.0087	0.0004	0.0585

proportion of population in a given category of the variables AGE, EDUCATION, CITIZENSHIP and LABOR. Therefore, by observing the signs of the regression parameters for $\alpha = 0$, we interpret that the poverty proportion tends to be smaller in those domains with larger proportion of population in the subset defined by age greater than 25, education in the category of university studies completed, and non Spanish citizenship (may be because immigrants tends to go to regions with greater richness where it is easier to find job), and with lower proportion of unemployed people. By doing the same exercise with the signs of the regression parameter for $\alpha = 1$, we interpret that poverty gap tends to be greater in those domains with larger proportion of population characterized by youth, absence of studies, Spanish citizenship and unemployment.

Residuals $\hat{e}_{dt} = \overline{y}_{dt} - \overline{\mathbf{X}}_{dt}\hat{\beta} - \hat{u}_{dt}$ of fitted model 1 are plotted against the observed values $\overline{y}_{dt}$ in the Fig. 5.1 for $\alpha = 0$ (top-right) and $\alpha = 1$ (top-left). The dispersion graph shows that EBLUP1 estimates are over and below direct estimates, so that the design unbiased property of the direct estimator is not completely lost by using the model 1. On the right part of the figure we observe that residuals tend to be positive, which means that the model is smoothing the value of the direct estimator larger values. We find that this is an interesting property because it protects us from the presence of outliers in the collection of direct domain estimates.

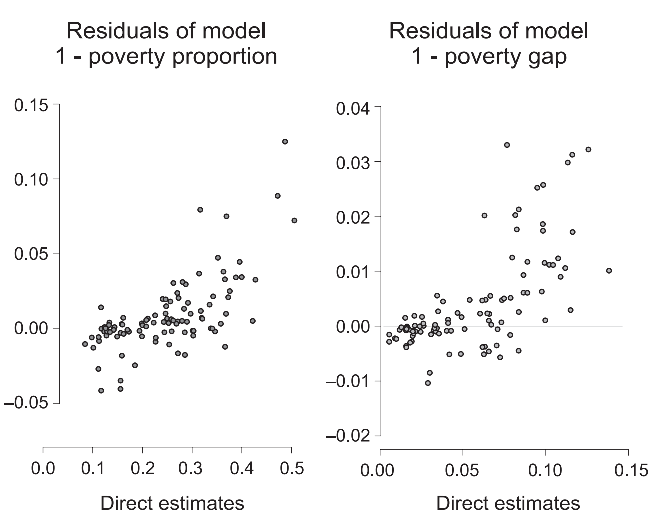

Fig. 5.1. Residuals versus direct estimates.

The three considered estimators of the poverty proportion and gap (direct, EBLUP0 and EBLUP1) are plotted in the Fig. 5.2 for $\alpha = 0$ (top-left) and $\alpha = 1$ (top-right). Their root mean squared error estimates are plotted in the Fig. 5.2 for $\alpha = 0$ (bottom-left) and $\alpha = 1$ (bottom-right). We observe that the EBLUP1 is the one presenting the best results and it is thus the one we recommend. Finally full numerical information is presented in the Table A.1 for the poverty proportion and in the Table A.2 for the poverty gap. In these tables direct, EBLUP0 and EBLUP1 estimates are labeled by dir, eb0 and eb1 respectively.

In Tables A.1 and A.2 the Spanish provinces are listed as follows: 1 Álava, 2 Albacete, 3 Alicante, 4 Almería, 5 Ávila, 6 Badajoz, 7 Baleares, 8 Barcelona, 9 Burgos, 10 Cáceres, 11 Cádiz, 12 Castellón, 13 Ciudad Real, 14 Córdoba, 15 Coruña La, 16 Cuenca, 17 Gerona, 18 Granada, 19 Guadalajara, 20 Guipúzcoa, 21 Huelva, 22 Huesca, 23 Jaén, 24 León, 25 Lérida, 26 La Rioja, 27 Lugo, 28 Madrid, 29 Málaga, 30 Murcia, 31 Navarra, 32 Orense, 33 Asturias (Oviedo), 34 Palencia, 35 Palmas Las, 36 Pontevedra, 37 Salamanca, 38 Santa Cruz de Tenerife, 39 Cantabria (Santander), 40 Segovia, 41 Sevilla, 42 Soria, 43 Tarragona, 44 Teruel, 45 Toledo, 46 Valencia, 47 Valladolid, 48 Vizcaya, 49 Zamora, 50 Zaragoza, 51 Ceuta, 52 Melilla.

In the Fig. 5.3 the Spanish provinces are plotted in 4 colored categories depending on the values of the EBLUP1 estimates in % of the poverty proportions and the poverty gaps, *i.e.*, $p_d = 100.\,\hat{\bar{Y}}^{eblup1}_{0;d,2006}$ and $g_d = 100.\,\hat{\bar{Y}}^{eblup1}_{1;d,2006}$. We observe that the Spanish regions where the proportion of the population under the poverty line is smallest are those situated in the north and east. On the other hand the Spanish regions with higher poverty proportion are those situated in the center-south. In an intermediate position we can find regions that are in the center-north of Spain. If we investigate how far the annual net incomes of population under the poverty line z_{2006} are from z_{2006}, we observe that in the Spanish regions situated in the center-north there exist a distance that is generally lower than the 6% of z_{2006}. However, the cited distance is in general greater than 6% of z_{2006} in the center-south.

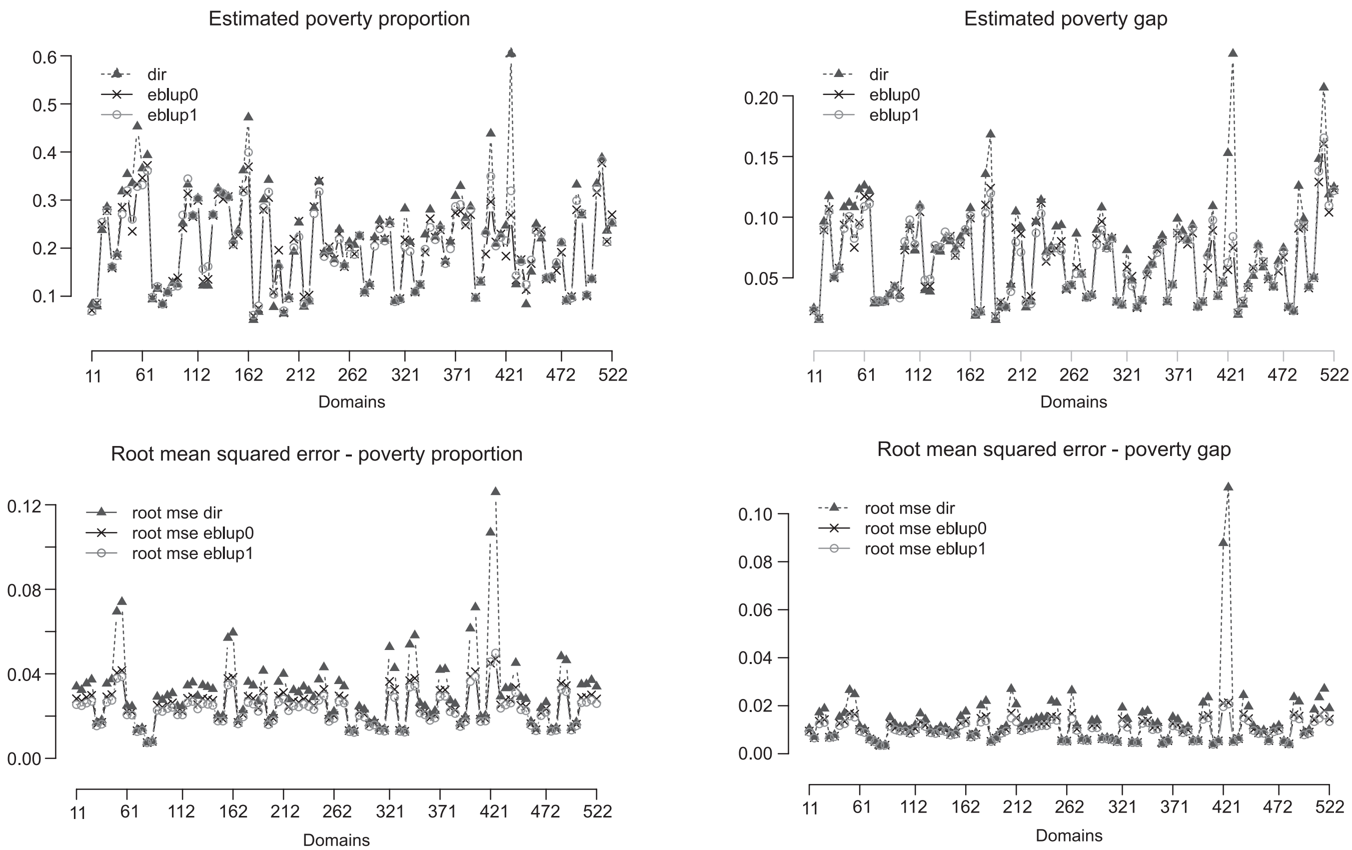

Fig. 5.2. Estimates of poverty proportions and gaps (top) and squared roots of their estimated MSEs (bottom).

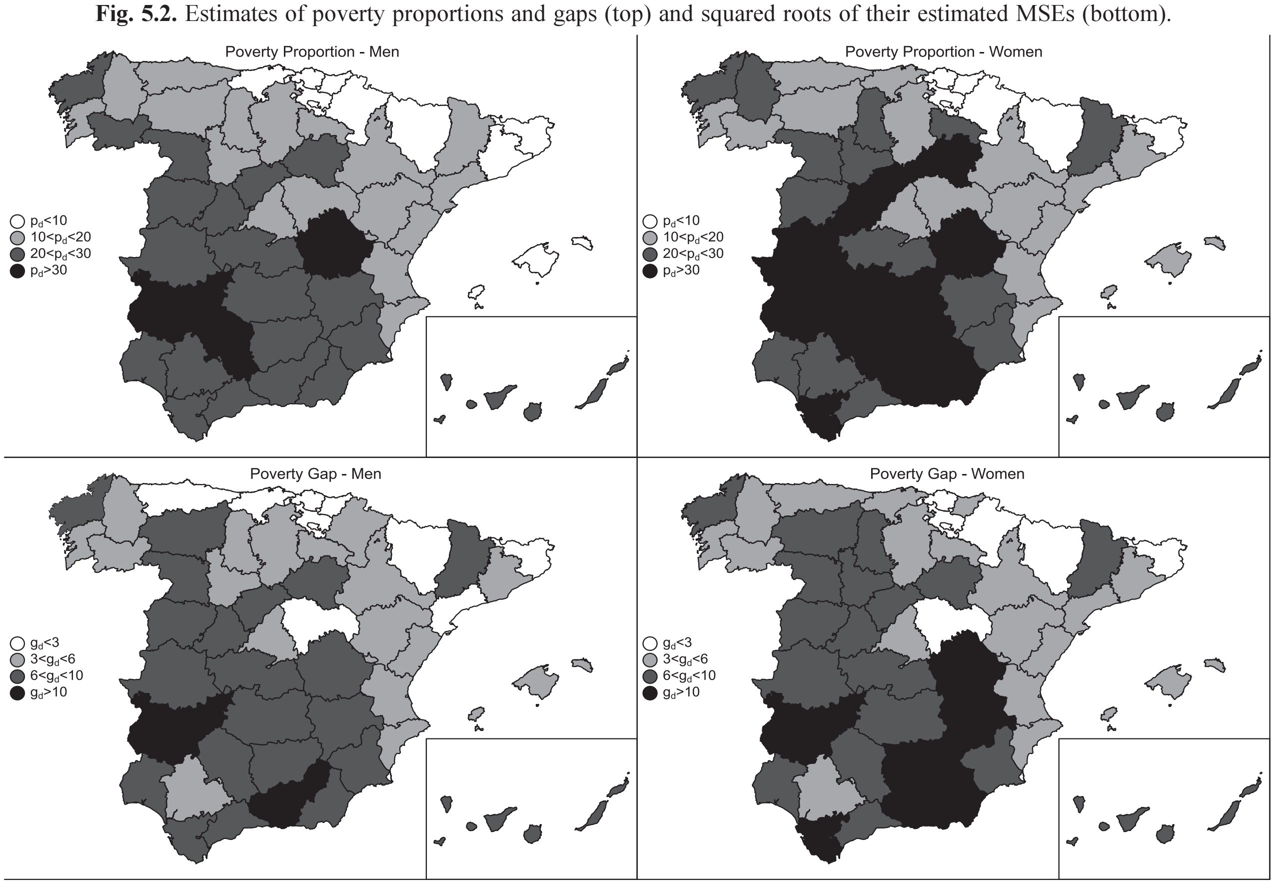

Fig. 5.3. Estimates of Spanish poverty proportions (top) and gaps (bottom) for men (left) and women (right) in 2006.

6. CONCLUSIONS

As poverty indicators are nonlinear, the EBLUP approach based on unit-level model does not give a good estimation procedure. More recently new unit-level model procedures generating censuses have been proposed (see Molina and Rao (2010)). However these methods require data from auxiliary variables in all the units of the population. Instead, area-level models provide an easy-to-apply solution. As they do not need unit-level data, they are a good alternative to generating-censuses unit-level-model methods for estimating poverty indicators. One of the criticism to area-level model is that they loose information by using aggregated data. This loss of information can be reduced by adding time information. For this reason we propose the use of temporal models that borrow strength from time. Two models are introduced and simulation studies are carried out to investigate when it is worthwhile to introduce a time correlation parameter. The methodology is illustrated with an application to SLCS data. The gain of efficiency is illustrated with figures and tables.

The analyzed data may require further modeling for taking into account the possible different behavior of poverty in the subsets of men and women. The new models may require introducing mixed normal distributions or modeling the men and women area effects with different distributions or with the same distribution but different parameters. These are tasks for future investigations.

ACKNOWLEDGEMENTS

The authors are grateful to the European Commission and to the Spanish Government for their economical support under the grants EU-FP7-SSH-2007-1 of SAMPLE project and MTM2009-09473 respectively. The authors also thanks the Instituto Nacional de Estadística for providing the Spanish EU-SILC data.

REFERENCES

Das, K., Jiang, J. and Rao, J.N.K. (2004). Mean squared error of empirical predictor. *Ann. Statist.*, **32**, 818-840.

Foster, J., Greer, J. and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, **52**, 761-766.

Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statist. Sci.*, **9**, 55-93.

Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, **15**, 1-96.

Molina, I. and Rao, J.N.K. (2010). Small area estimation of poverty indicators. *Canad. J. Statist.*, **38(3)**, 369-385.

Pfeffermann, D. (2002). Small area estimation - New developments and directions. *Intern. Statist. Rev.*, **70(1)**, 125-143.

Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *J. Amer. Statist. Assoc.,* **85**, 163-171.

Rao, J.N.K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, **25**, 175-186.

Rao, J.N.K. (2003). *Small Area Estimation*. John Wiley.

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

APPENDIX A: TABLES

Table A.1. Estimated domain poverty proportions and squared root MSE's by sex in 2006.

	men/poverty proportions/women						men / sqrt.mse / women					
d	dir	eb0	eb1	dir	eb0	eb1	dir	eb0	eb1	dir	eb0	eb1
1	0.083	0.071	0.068	0.079	0.086	0.087	0.034	0.028	0.025	0.032	0.027	0.025
2	0.237	0.250	0.254	0.285	0.278	0.278	0.035	0.029	0.026	0.037	0.030	0.027
3	0.160	0.159	0.161	0.189	0.185	0.184	0.017	0.016	0.015	0.018	0.017	0.016
4	0.318	0.285	0.270	0.354	0.315	0.320	0.035	0.029	0.027	0.037	0.030	0.027
5	0.335	0.235	0.260	0.453	0.333	0.328	0.069	0.040	0.038	0.074	0.042	0.039
6	0.366	0.346	0.331	0.393	0.372	0.361	0.025	0.022	0.021	0.025	0.022	0.020
7	0.094	0.096	0.098	0.115	0.117	0.120	0.014	0.013	0.013	0.014	0.014	0.013
8	0.083	0.084	0.083	0.108	0.108	0.108	0.008	0.007	0.007	0.008	0.008	0.008
9	0.127	0.131	0.120	0.124	0.138	0.121	0.029	0.025	0.023	0.028	0.024	0.022
10	0.252	0.242	0.269	0.332	0.313	0.344	0.030	0.026	0.024	0.031	0.026	0.024
11	0.267	0.265	0.268	0.303	0.299	0.303	0.025	0.022	0.021	0.025	0.022	0.021
12	0.122	0.131	0.157	0.122	0.135	0.162	0.034	0.028	0.026	0.036	0.029	0.027
13	0.269	0.268	0.270	0.324	0.311	0.320	0.030	0.025	0.023	0.035	0.028	0.026
14	0.312	0.302	0.314	0.307	0.307	0.306	0.034	0.028	0.026	0.033	0.027	0.025
15	0.216	0.207	0.209	0.237	0.226	0.232	0.020	0.019	0.018	0.020	0.019	0.018
16	0.362	0.320	0.317	0.472	0.369	0.399	0.057	0.038	0.035	0.059	0.038	0.036
17	0.050	0.059	0.061	0.067	0.075	0.080	0.018	0.017	0.016	0.023	0.021	0.020
18	0.301	0.280	0.285	0.342	0.305	0.317	0.036	0.029	0.026	0.034	0.028	0.026
19	0.077	0.108	0.104	0.165	0.196	0.161	0.027	0.024	0.022	0.041	0.032	0.029
20	0.064	0.065	0.070	0.100	0.097	0.095	0.018	0.017	0.016	0.020	0.019	0.018
21	0.192	0.219	0.201	0.253	0.256	0.223	0.036	0.029	0.027	0.040	0.031	0.028
22	0.078	0.098	0.083	0.089	0.102	0.091	0.028	0.024	0.023	0.032	0.027	0.025
23	0.283	0.286	0.272	0.339	0.339	0.318	0.031	0.027	0.024	0.034	0.028	0.026
24	0.192	0.188	0.183	0.193	0.203	0.199	0.032	0.027	0.025	0.029	0.025	0.023
25	0.177	0.177	0.170	0.239	0.230	0.218	0.037	0.030	0.027	0.043	0.033	0.030
26	0.166	0.161	0.163	0.212	0.204	0.202	0.020	0.019	0.017	0.022	0.020	0.019
27	0.207	0.188	0.187	0.225	0.225	0.226	0.037	0.030	0.027	0.034	0.028	0.026
28	0.110	0.108	0.109	0.126	0.123	0.123	0.014	0.013	0.013	0.013	0.013	0.012
29	0.222	0.215	0.204	0.258	0.247	0.240	0.025	0.022	0.020	0.023	0.021	0.020
30	0.219	0.218	0.215	0.256	0.253	0.251	0.017	0.016	0.015	0.018	0.017	0.016

	men/poverty proportions/women						men / sqrt.mse / women					
d	dir	eb0	eb1	dir	eb0	eb1	dir	eb0	eb1	dir	eb0	eb1
31	0.090	0.090	0.089	0.094	0.095	0.093	0.014	0.014	0.013	0.014	0.013	0.013
32	0.282	0.217	0.203	0.213	0.212	0.193	0.053	0.036	0.032	0.043	0.033	0.029
33	0.108	0.109	0.109	0.122	0.124	0.125	0.014	0.013	0.013	0.013	0.013	0.012
34	0.228	0.192	0.198	0.280	0.261	0.243	0.054	0.037	0.034	0.058	0.038	0.034
35	0.224	0.223	0.218	0.246	0.238	0.240	0.026	0.023	0.021	0.025	0.022	0.021
36	0.174	0.171	0.168	0.214	0.210	0.199	0.021	0.019	0.018	0.022	0.020	0.019
37	0.308	0.274	0.287	0.329	0.277	0.291	0.042	0.032	0.029	0.042	0.032	0.029
38	0.263	0.248	0.253	0.286	0.270	0.280	0.027	0.024	0.022	0.026	0.023	0.021
39	0.095	0.098	0.097	0.128	0.132	0.132	0.017	0.016	0.015	0.020	0.018	0.017
40	0.234	0.188	0.231	0.438	0.296	0.349	0.061	0.039	0.036	0.071	0.041	0.039
41	0.209	0.213	0.204	0.228	0.229	0.222	0.020	0.019	0.017	0.020	0.019	0.018
42	0.247	0.183	0.216	0.604	0.268	0.319	0.107	0.045	0.046	0.126	0.047	0.050
43	0.125	0.134	0.143	0.174	0.176	0.172	0.029	0.025	0.024	0.033	0.028	0.025
44	0.083	0.114	0.124	0.151	0.168	0.175	0.033	0.028	0.026	0.045	0.033	0.031
45	0.250	0.239	0.236	0.220	0.236	0.230	0.029	0.025	0.023	0.028	0.024	0.023
46	0.137	0.136	0.138	0.139	0.138	0.141	0.017	0.016	0.016	0.014	0.014	0.013
47	0.165	0.154	0.170	0.210	0.191	0.213	0.024	0.021	0.020	0.027	0.023	0.022
48	0.092	0.091	0.092	0.099	0.099	0.096	0.014	0.013	0.013	0.014	0.014	0.013
49	0.332	0.280	0.299	0.268	0.271	0.273	0.048	0.035	0.033	0.046	0.034	0.032
50	0.101	0.100	0.103	0.136	0.136	0.136	0.014	0.014	0.013	0.017	0.016	0.015
51	0.334	0.316	0.324	0.388	0.377	0.383	0.035	0.029	0.026	0.035	0.029	0.027
52	0.236	0.214	0.212	0.251	0.270	0.253	0.037	0.030	0.028	0.034	0.028	0.026

Table A.2. Estimated domain poverty gaps and squared root MSE's by sex in 2006.

	men/poverty proportions/women						men / sqrt.mse / women					
d	dir	eb0	eb1	dir	eb0	eb1	dir	eb0	eb1	dir	eb0	eb1
1	0.025	0.023	0.023	0.015	0.017	0.017	0.010	0.009	0.009	0.007	0.007	0.007
2	0.096	0.089	0.090	0.117	0.106	0.105	0.017	0.014	0.012	0.019	0.014	0.013
3	0.050	0.050	0.051	0.059	0.058	0.057	0.007	0.007	0.007	0.008	0.007	0.007
4	0.108	0.094	0.091	0.112	0.099	0.101	0.015	0.013	0.012	0.017	0.013	0.012
5	0.108	0.075	0.082	0.123	0.095	0.093	0.027	0.017	0.016	0.025	0.016	0.015
6	0.126	0.117	0.109	0.121	0.116	0.111	0.011	0.010	0.009	0.010	0.009	0.009
7	0.029	0.030	0.032	0.029	0.030	0.031	0.006	0.006	0.006	0.005	0.005	0.005
8	0.031	0.031	0.031	0.036	0.036	0.036	0.003	0.003	0.003	0.004	0.004	0.004

	men/poverty proportions/women						men / sqrt.mse / women					
d	dir	eb0	eb1	dir	eb0	eb1	dir	eb0	eb1	dir	eb0	eb1
9	0.042	0.044	0.043	0.035	0.039	0.033	0.015	0.012	0.011	0.012	0.011	0.010
10	0.075	0.073	0.080	0.093	0.091	0.098	0.011	0.010	0.009	0.011	0.010	0.010
11	0.072	0.075	0.078	0.109	0.104	0.108	0.010	0.009	0.008	0.012	0.010	0.010
12	0.040	0.043	0.048	0.039	0.043	0.049	0.017	0.013	0.013	0.014	0.012	0.011
13	0.073	0.075	0.077	0.072	0.074	0.075	0.010	0.009	0.009	0.010	0.009	0.008
14	0.082	0.084	0.088	0.080	0.084	0.084	0.011	0.010	0.010	0.011	0.010	0.009
15	0.073	0.068	0.069	0.083	0.077	0.079	0.009	0.008	0.008	0.009	0.009	0.008
16	0.088	0.088	0.090	0.107	0.099	0.101	0.016	0.013	0.012	0.018	0.014	0.013
17	0.019	0.021	0.021	0.022	0.023	0.022	0.008	0.007	0.007	0.009	0.008	0.008
18	0.135	0.110	0.103	0.168	0.124	0.120	0.020	0.015	0.013	0.022	0.015	0.014
19	0.015	0.018	0.018	0.026	0.030	0.029	0.005	0.005	0.005	0.007	0.006	0.006
20	0.026	0.026	0.026	0.044	0.042	0.039	0.010	0.009	0.009	0.011	0.010	0.009
21	0.105	0.091	0.079	0.091	0.086	0.071	0.027	0.017	0.015	0.021	0.015	0.013
22	0.026	0.032	0.029	0.030	0.035	0.029	0.011	0.010	0.010	0.013	0.011	0.011
23	0.096	0.096	0.087	0.114	0.112	0.103	0.013	0.011	0.011	0.015	0.012	0.011
24	0.071	0.064	0.067	0.076	0.071	0.073	0.015	0.012	0.012	0.015	0.012	0.012
25	0.092	0.074	0.075	0.093	0.080	0.072	0.022	0.016	0.014	0.021	0.015	0.014
26	0.041	0.040	0.042	0.043	0.044	0.044	0.006	0.005	0.005	0.005	0.005	0.005
27	0.086	0.059	0.053	0.053	0.054	0.053	0.026	0.017	0.015	0.012	0.010	0.010
28	0.034	0.034	0.034	0.036	0.036	0.035	0.006	0.006	0.006	0.006	0.006	0.005
29	0.090	0.083	0.077	0.108	0.096	0.089	0.014	0.012	0.011	0.014	0.012	0.011
30	0.075	0.075	0.074	0.083	0.083	0.082	0.007	0.006	0.006	0.007	0.006	0.006
31	0.030	0.030	0.031	0.027	0.028	0.028	0.006	0.006	0.006	0.005	0.005	0.005
32	0.073	0.059	0.053	0.048	0.051	0.043	0.019	0.014	0.013	0.014	0.012	0.011
33	0.025	0.025	0.026	0.031	0.031	0.032	0.005	0.005	0.005	0.005	0.005	0.005
34	0.056	0.053	0.054	0.061	0.064	0.061	0.017	0.013	0.013	0.018	0.014	0.013
35	0.076	0.073	0.071	0.085	0.081	0.080	0.012	0.011	0.010	0.013	0.011	0.010
36	0.030	0.031	0.031	0.044	0.045	0.044	0.004	0.004	0.004	0.006	0.005	0.005
37	0.099	0.087	0.087	0.089	0.082	0.083	0.015	0.013	0.012	0.014	0.012	0.011
38	0.081	0.078	0.081	0.093	0.087	0.091	0.010	0.009	0.009	0.011	0.010	0.009
39	0.026	0.027	0.026	0.030	0.031	0.030	0.006	0.005	0.005	0.006	0.005	0.005
40	0.070	0.058	0.068	0.109	0.089	0.098	0.021	0.015	0.014	0.023	0.016	0.015
41	0.034	0.036	0.035	0.045	0.048	0.047	0.004	0.004	0.004	0.006	0.005	0.005
42	0.153	0.057	0.063	0.235	0.075	0.084	0.088	0.021	0.020	0.111	0.021	0.021
43	0.019	0.021	0.022	0.028	0.030	0.031	0.005	0.005	0.005	0.007	0.006	0.006
44	0.045	0.042	0.042	0.052	0.058	0.057	0.024	0.016	0.015	0.020	0.015	0.014
45	0.077	0.074	0.077	0.059	0.063	0.064	0.011	0.010	0.010	0.009	0.009	0.008
46	0.051	0.049	0.050	0.043	0.043	0.043	0.010	0.009	0.008	0.006	0.005	0.005
47	0.064	0.055	0.059	0.074	0.067	0.072	0.011	0.010	0.009	0.012	0.010	0.010
48	0.026	0.026	0.026	0.023	0.023	0.023	0.005	0.005	0.005	0.004	0.004	0.004
49	0.126	0.091	0.095	0.099	0.092	0.093	0.024	0.016	0.015	0.022	0.016	0.015
50	0.043	0.042	0.043	0.051	0.050	0.050	0.009	0.008	0.008	0.010	0.009	0.009
51	0.148	0.129	0.138	0.207	0.160	0.165	0.018	0.014	0.013	0.023	0.016	0.015
52	0.119	0.104	0.110	0.125	0.123	0.122	0.027	0.018	0.016	0.019	0.014	0.013