CrossMark

REGULAR ARTICLE

On longitudinal moving average model for prediction of subpopulation total

Tomasz Żądło

Abstract In the paper the empirical best linear unbiased predictor of the subpopulation total is proposed under some longitudinal model where both temporal and spatial moving average models of profile specific random components are taken into account. Two estimators of the mean square error of the predictor are proposed as well. Considerations are supported by two Monte Carlo simulation studies and the case study.

Keywords Survey sampling · Subpopulation total · Longitudinal data · Linear mixed model

Mathematics Subject Classification (2000) 62D05

1 Introduction

In the survey sampling estimation or prediction of population characteristics is usually the key issue but subpopulations (domains) characteristics are of interest as well. What is more, in many cases we are looking for possibilities of increasing the accuracy, especially when the sample size in the domain of interest in the period of interest is small. Such domains are called small areas. In the case of the longitudinal data we can "borrow strength" from different periods and/or domains and use the information on spatial and temporal correlation. In the paper some unit-level longitudinal model

T. Żądło (✉)
Department of Statistics, University of Economics in Katowice, 50, 1 Maja Street,
40-287 Katowice, Poland
e-mail: tomasz.zadlo@ue.katowice.pl

is proposed which is a special case of the Linear Mixed Model (LMM) with two random components which obey assumptions of spatial moving average model and the temporal MA(1) model.

Verbeke and Molenberghs (2000, p. 24) or Hedeker and Gibbons (2006, p. 115) propose a longitudinal model which is a special case of the Linear Mixed Model with profile-specific random components, where the profile is defined as a vector of random variables for a population element in different periods. Here we define the profile as a vector of random variables for observations of an element in some domain what allows to take the possibility of population changes in time into account. Hence, the profile is not element specific but element and domain specific. In mentioned books the assumptions are made only for the sampled elements while we make assumptions for all of population elements. What is more, the authors assume profiles to be independent, while here they are spatially correlated.

In many papers small area predictors are derived under both area-level and unit-level models where the spatial correlation is taken into account but assuming that all data refer to single time point (Molina et al. 2009; Petrucci and Salvati 2006; Petrucci et al. 2005; Pratesi and Salvati 2008; Chandra et al. 2007). The models are special cases of the Linear Mixed Model where one of the random components obeys the assumption of the SAR(1) process between subpopulations (what means that we assume the same realization of the random component for all of the population elements which belong to the same domain). What is more, Salvati et al. (2009) propose the spatial M-quantile predictor which occurred slightly more accurate than other predictors for contaminated data in their simulation studies.

If longitudinal data are studied many predictors are considered especially based on area-level models. Rao and You (1994) and Esteban et al. (2012) assume longitudinal area-level models with time effects under the assumption of the AR(1) model and independent area-level effects. In Marhuenda et al. (2013) the area-level model with AR(1) time effects and SAR(1) area effects is proposed. Singh et al. (2005) using the Kalman filtering approach propose a spatio-temporal model. Ugarte et al. (2009) study semiparametric models combining both non-parametric trends and small area random effects using P-spline regression.

Saei and Chambers (2003) propose many small area methods for longitudinal data as a part of the EURAREA project. In the sections devoted to both unit-level and area-level models they consider independent area effects together with independent or autocorrelated time effects. Models with time varying area effect are studied as well. The unit-level model with spatially correlated area effects is also considered but for one period.

Molina et al. (2010a) in the European Project SAMPLE propose inter alia many area and unit-level models and predictors. In the chapter 7 they study longitudinal area-level models with time varying area effects assuming the independence of the effects between domains and the AR(1) model across time instants (independence of time varying area effects is also considered). They also propose partitioned versions of the model, where domains are divided into two groups and parameters of the distribution of the time varying area effects differ between these groups. In the chapter 8 they consider area-level time-space models which are special cases of the Linear Mixed Model with three random components, including assumptions of the AR(1) and the SAR(1)

processes for random components. In the chapter 9 they consider unit-level models with independent and correlated time-effects. In one of the models they assume three random components including independent area effects and time varying area effect which obeys assumptions of the AR(1) model across time instants and independence across areas.

In this paper we propose some longitudinal model and we derive empirical best linear unbiased predictor under the model together with its MSE estimators. The main differences between the proposed approach and proposals presented in other papers are as follows:

– random components in our model are profile specific while in other papers area effects or time effects or time varying area affects are assumed, what means that in our case we do not assume that realizations of random components are the same within domains or within time instants or vary only between domains and time periods,
– in this paper we use the spatial moving average model to describe spatial dependence instead of the first order spatial autoregressive model SAR(1),
– here we use the first order temporal moving average model to describe temporal autocorrelation instead of the first order autoregressive model,
– spatial dependence is assumed at the low aggregation level—between profiles instead of domains,
– temporal autocorrelation is assumed at the low aggregation level—within profiles instead of within domains,
– in the model changes of population and changes of domains' affiliation in time are taken into account.

2 Basic notations

Longitudinal data for periods $t = 1, \ldots, M$ are considered. In the period t the population of size N_t is denoted by Ω_t. The population in the period t is divided into D disjoint subpopulations (domains) Ω_{dt} of size N_{dt}, where $d = 1, \ldots, D$. Let the set of population elements for which observations are available in the period t be denoted by s_t and its size by n_t. The set of subpopulation elements for which observations are available in the period t is denoted by s_{dt} and its size by n_{dt}. The d^*th domain of interest in the period of interest t^* will be denoted by $\Omega_{d^*t^*}$. Let $\Omega_{rdt} = \Omega_{dt} \setminus s_{dt}, N_{rdt} = N_{dt} - n_{dt}$, $\bigcup_{t=1}^{M} \Omega_t = \Omega, \bar{\bar{\Omega}} = N, \bigcup_{t=1}^{M} \Omega_{dt} = \Omega_d, \bar{\bar{\Omega}}_d = N_d, \bigcup_{t=1}^{M} \Omega_{rdt} = \Omega_{rd}, \bar{\bar{\Omega}}_{rd} = N_{rd}$, $\bigcup_{t=1}^{M} s_t = s, \bar{\bar{s}} = n, \bigcup_{t=1}^{M} s_{dt} = s_d, \bar{\bar{s}}_d = n_d$.

Let M_{id} denotes the number of periods when the ith population element belongs to the dth domain and m_{id}—the number of periods when the ith population element (which belongs to the dth domain) is observed. Let $M_{rid} = M_{id} - m_{id}$. It is assumed that the population may change in time and that one population element may change its domain affiliation in time. Hence, sets of population elements Ω_d (where $d = 1, 2, \ldots, D$) may overlap. Values of the variable of interest are realizations of random variables Y_{idj} for the ith population element which belongs to the dth domain in the period t_{ij}, where $i = 1, 2, \ldots, N, j = 1, 2, \ldots, M_{id}, d = 1, 2, \ldots, D$. The vector

$\mathbf{Y_{id}} = \left[Y_{idj}\right]_{M_{id} \times 1}$ will be called profile and the vector $\mathbf{Y_{sid}} = \left[Y_{idj}\right]_{m_{id} \times 1}$ will be called sample profile. Let the vector $\mathbf{Y_{rid}} = \left[Y_{idj}\right]_{M_{rid} \times 1}$ be profile for nonobserved realizations of random variables.

The proposed approach may be used to predict the domain total for any (past, current and future) periods but under assumption that values of the auxiliary variables and the division of the population into subpopulations in the period of interest are known.

3 Superpopulation model

Special cases of the general or the generalized mixed linear models are widely used in different areas including for example genetics (e.g. Bernardo 1996), insurance (e.g. Wolny 2009) and statistical image analysis (e.g. Demidenko 2004, chapter 12), We consider superpopulation models used for longitudinal data (compare Verbeke and Molenberghs, 2000; Hedeker and Gibbons, 2006) which are special cases of the LMM. The following model is assumed:

$$\mathbf{Y_d} = \mathbf{X_d}\beta_\mathbf{d} + \mathbf{Z_d}\mathbf{v_d} + \mathbf{e_d}, \tag{1}$$

where $\mathbf{Y_d} = col_{1 \leq i \leq N_d}(\mathbf{Y_{id}})$, where $\mathbf{Y_{id}}$ is a random vector of size $M_{id} \times 1$, $\mathbf{X_d} = col_{1 \leq i \leq N_d}(\mathbf{X_{id}})$, where $\mathbf{X_{id}}$ is a known matrix of size $M_{id} \times p$, $\mathbf{Z_d} = diag_{1 \leq i \leq N_d}(\mathbf{Z_{id}})$, where $\mathbf{Z_{id}}$ is a known vector of size $M_{id} \times 1$ (e.g. vector of 1's), $\mathbf{v_d} = col_{1 \leq i \leq N_d}(v_{id})$, where v_{id} is a random component and $\mathbf{v_d}(d = 1, 2 \ldots, D)$ are assumed to be independent, $\mathbf{e_d} = col_{1 \leq i \leq N_d}(\mathbf{e_{id}})$, where $\mathbf{e_{id}}$ is a random component vector of size $M_{id} \times 1$ and $\mathbf{e_{id}}$ ($i = 1, 2, \ldots, N; d = 1, 2, \ldots, D$) are assumed to be independent, $\mathbf{v_d}$ and $\mathbf{e_d}$ are assumed to be independent.

What is more, the vector of random components $\mathbf{v_d}$ obeys assumptions of the spatial moving average process, i.e.

$$\mathbf{v_d} = \lambda_{(sp)}\mathbf{W}_d\mathbf{u_d} + \mathbf{u_d}, \tag{2}$$

where $\mathbf{W}_d$ is the spatial weight matrix for profiles $\mathbf{Y_{id}}$, $\mathbf{u_d} \sim (\mathbf{0}, \sigma_u^2\mathbf{I_{N_d}})$. Hence, $\mathbf{v_d} \sim (\mathbf{0}, \mathbf{R_d})$, where $\mathbf{R_d} = \sigma_u^2\mathbf{H_d}$ and $\mathbf{H_d} = \mathbf{I_{N_d}} + \lambda_{(sp)}(\mathbf{W_d} + \mathbf{W_d^T}) + \lambda_{(sp)}^2\mathbf{W_d}\mathbf{W_d^T}$. Moreover, elements of $\mathbf{e_{id}}$ obey assumptions of MA(1) temporal process, i.e.

$$e_{idt} = \varepsilon_{idt} - \lambda_{(t)}\varepsilon_{idt-1}. \tag{3}$$

where $\varepsilon_{tdj} \sim (0, \sigma_\varepsilon^2)$. Hence,

$$\mathrm{Cov}(e_{tdj}, e_{t'd'j'}) = \begin{cases} \sigma_\varepsilon^2(1 + \lambda_{(t)}^2) & \text{if } t = t' \\ -\sigma_\varepsilon^2\lambda_{(t)} & \text{if } |t - t'| = 1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

and then $\mathbf{e_{id}} \sim (\mathbf{0}, \Gamma_{\mathbf{id}})$, where elements of $\Gamma_{\mathbf{id}}$ are given by (4).

Variance-covariance matrices of $\mathbf{Y_d}$ (where $d = 1, 2, \ldots, D$) are functions of unknown parameters $\delta = \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_u^2 & \lambda_{(t)} & \lambda_{(sp)} \end{bmatrix}$.

If the population changes in time, new elements of the population or observations of the population element after the change of its domain affiliation form a new profile $\mathbf{Y_{id}}$. It means that observations of the new population element will be temporally correlated within the profile and spatially correlated with other population elements within the subpopulation. If the population element changes its domain affiliation its new observations will be temporally correlated (but temporally uncorrelated with old observations) and spatially correlated with other population elements within a new subpopulation (but spatially uncorrelated with elements of the previous subpopulation).

To explain the idea of the model let us suppose that we study a population of households divided into domains according to the type of the household (what includes the criterion of the number of persons who belong to the household). Let the variable of interest be expenditures on some goods and let us consider the problem of prediction of the expenditures for the domains. Based on the model we assume that expenditures of two households of the same type (i.e. which belong to the same domain) are spatially correlated (where the distance may be measured in geographical or economic sense). Moreover, we assume that expenditures of each household are temporally autocorrelated assuming the MA(1) model. The assumption of the MA(1) model (which belongs to the class of short memory time series models) implies that non-zero covariances are assumed for lags which equal 1 (for periods t and $t - 1$). The assumption is more realistic than the assumption of the temporal independence and in the case of fast changes in the economy and in the economic situation of households it does not have to be treated as strong. Let us consider a situation when the type of household is changed e.g. from the household which consists of two persons (a couple) into the household which consists of three persons (a couple and a child). Hence, we assume that the temporal correlation is broken. Moreover, the household is not longer spatially correlated with households of the previous type but it becomes spatially correlated with households of the new type.

4 Best linear unbiased predictor

Let $\hat{\boldsymbol{\beta}}_{d*} = \left(\mathbf{X_{sd*}^T V_{ssd*}^{-1} X_{sd*}} \right)^{-1} \mathbf{X_{sd*}^T V_{ssd*}^{-1} Y_{sd*}}$, where $\mathbf{X_{sd*}}$ is a known matrix of auxiliary variables of size $\sum_{i=1}^{n_{d*}} m_{id*} \times p$, $\mathbf{Y_{sd*}}$ is a $\sum_{i=1}^{n_{d*}} m_{id*} \times 1$ vector of random variables Y_{idj}, $\mathbf{V_{ssd*}^{-1}} = \left(\sigma_u^2 \mathbf{Z_{sd*} H_{d*} Z_{sd*}^T} + diag_{1 \leq i \leq n_{d*}} (\Gamma_{\mathbf{ssid*}}) \right)^{-1}$, where $\mathbf{Z_{sd}} = diag_{1 \leq i \leq n_d} (\mathbf{Z_{sid}})$, $\mathbf{Z_{sid}}$ is a known vector of size $m_{id} \times 1$ (e.g. the vector of 1s), $\Gamma_{\mathbf{ssid}}$ is a submatrix obtained from $\Gamma_{\mathbf{id}}$ by deleting rows and columns for unsampled observations. Based on the Royall (1976) theorem it is possible to derive the formula of the best linear unbiased predictor (BLUP) of the subpopulation total:

$$\hat{\theta}_{d*t*}^{BLU} = \sum_{i \in s_{d*t*}} Y_{id*t*} + \tilde{\boldsymbol{x}}_{\mathbf{rd*t*}} \hat{\boldsymbol{\beta}}_{\mathbf{d*}} + \gamma_{\mathbf{rd*}}^T \left(\sigma_u^2 \mathbf{Z_{rd*} H_{d*} Z_{sd*}^T} + diag_{1 \leq i \leq N_{rd*}} (\Gamma_{\mathbf{rsid*}}) \right)$$
$$\times \mathbf{V_{ssd*}^{-1}} \left(\mathbf{Y_{sd*}} - \mathbf{X_{sd*}} \hat{\boldsymbol{\beta}}_{\mathbf{d*}} \right), \tag{5}$$

where $\tilde{x}_{\mathbf{rd*t*}}$ is a $1 \times p$ vector of totals of auxiliary variables in Ω_{rd*t*}, $\gamma_{\mathbf{rd*}}$ is a $\sum_{i=1}^{n_{d*}} M_{rid*} \times 1$ vector of ones for observations in Ω_{rd*t*} and zero otherwise. The predictor (5) is the sum of three elements. If $t*$ is the future period then $s_{d*t*} = \emptyset$, $\Omega_{rd*t*} = \Omega_{d*t*}$ and the first element of (5) (given by $\sum_{i \in s_{d*t*}} Y_{id*t*}$) equals zero. Hence, if the domain total of the auxiliary variable is known in the future period as well as the division of the population into subpopulations in the future period is known, then it is possible to use (5) to predict the future domain total of the variable of interest.

The MSE of the BLUP given by (5) is as follows:

$$MSE_\xi(\hat{\theta}_{d*t*}^{BLU}) = Var_\xi(\hat{\theta}_{d*t*}^{BLU} - \theta) = g_1(\delta) + g_2(\delta), \tag{6}$$

where

$$g_1(\delta) = \gamma_{\mathbf{rd*}}^T \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{rd*}}^T + diag_{1 \le i \le N_{rd*}}(\Gamma_{\mathbf{rrid*}}) \right) \gamma_{\mathbf{rd*}}$$

$$- \gamma_{\mathbf{rd*}}^T \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{sd*}}^T + diag_{1 \le i \le N_{rd*}}(\Gamma_{\mathbf{rsid*}}) \right)$$

$$\times \mathbf{V}_{\mathbf{ssd*}}^{-1} \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{sd*}}^T + diag_{1 \le i \le N_{rd*}}(\Gamma_{\mathbf{rsid*}}) \right)^T \gamma_{\mathbf{rd*}}, \tag{7}$$

$$g_2(\delta) = \gamma_{\mathbf{rd*}}^T \left(\mathbf{X}_{\mathbf{rd*}} - \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{sd*}}^T + diag_{1 \le i \le N_{rd*}}(\Gamma_{\mathbf{rsid*}}) \right) \mathbf{V}_{\mathbf{ssd*}}^{-1} \mathbf{X}_{\mathbf{sd*}} \right)$$

$$\times \left(\mathbf{X}_{\mathbf{sd*}}^T \mathbf{V}_{\mathbf{ssd*}}^{-1} \mathbf{X}_{\mathbf{sd*}} \right)^{-1} \left(\mathbf{X}_{\mathbf{rd*}} - \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{sd*}}^T + diag_{1 \le i \le N_{rd*}}(\Gamma_{\mathbf{rsid*}}) \right) \mathbf{V}_{\mathbf{ssd*}}^{-1} \mathbf{X}_{\mathbf{sd*}} \right)^T \gamma_{\mathbf{rd*}}, \tag{8}$$

where $\mathbf{Z}_{\mathbf{rd}} = diag_{1 \le i \le N_{rd}}(\mathbf{Z}_{\mathbf{rid}})$, $\mathbf{Z}_{\mathbf{rid}}$ is a known vector of size $M_{rid} \times 1$ (e.g. the vector of 1 s), $\Gamma_{\mathbf{rrid}}$ is a submatrix obtained from $\Gamma_{\mathbf{id}}$ by deleting rows and columns for sampled observations, $\Gamma_{\mathbf{rsid}}$ is a submatrix obtained from $\Gamma_{\mathbf{id}}$ by deleting rows for sampled observations and columns for unsampled observations.

5 Empirical best linear unbiased predictor

Let the unknown variance parameters in (5) be replaced by their maximum likelihood (ML) or restricted maximum likelihood (REML) estimates under normality. We obtain the two-stage predictor called EBLUP. It remains unbiased under some weak assumptions (inter alia symmetric but not necessarily normal distribution of random components for the model assumed for the whole population). The proof is presented by Żądło (2004) for the empirical version of Royall (1976) BLUP and it is based on the results presented by Kackar and Harville (1981) for the empirical version of the BLUP proposed by Henderson (1950).

The problem of MSE estimation based on the Taylor expansion is considered in many papers on small area estimation but for the empirical version of BLUP proposed by Henderson (1950). The first proposal of the MSE estimator of the empirical version of the BLUP proposed by Henderson (1950) was presented by Kackar and Harville (1984) but they did not prove asymptotic unbiasedness of their MSE

estimator. The landmark paper on the topic is the paper written by Prasad and Rao (1990). They assume inter alia (as in this paper) independence of random variables for elements of population from different domains and that estimators of variance components are unbiased (what is not true for ML and REML estimators). They consider three special cases of the linear mixed model: Fay and Herriot (1979) model, the nested error regression model and the random regression coefficient model. To derive the MSE estimator they use three approximations. They prove that two of them are of order $o(D^{-1})$ for all of the three considered models. They also prove that the third approximation is of order $o(D^{-1})$ but only for the Fay and Herriot (1979) model. Unbiasedness of estimators of variance components is not assumed by Datta and Lahiri (2000). They assume the linear mixed model with block-diagonal variance covariance matrix (as in this paper) and they prove that the bias of their MSE estimator for ML and REML estimators of variance components is of order $o(D^{-1})$. But the proof is valid if the variance-covariance matrix is a linear combination of variance components. Das et al. (2004) consider a different asymptotic set-up. The bias of their MSE estimator is of order $o(m_*^{-2})$ where $m_* = min_{1 \leq k \leq q} m_k$ where $m_k = \left\| \mathbf{Z}_k^T \mathbf{P} \mathbf{Z}_k \right\|_2$, $\mathbf{Z}_k$ is the design matrix for the kth random effect factor, $\mathbf{P} = \mathbf{V}_{\mathbf{ss}}^{-1} - \mathbf{V}_{\mathbf{ss}}^{-1} \mathbf{X}_{\mathbf{s}} (\mathbf{X}_{\mathbf{s}}^T \mathbf{V}_{\mathbf{ss}}^{-1} \mathbf{X}_{\mathbf{s}})^{-1} \mathbf{X}_{\mathbf{s}}^T \mathbf{V}_{\mathbf{ss}}^{-1}$, $\mathbf{V}_{\mathbf{ss}}^{-1} = diag_{1 \leq d \leq D} \mathbf{V}_{\mathbf{ssd}}^{-1}$ and $\mathbf{X}_{\mathbf{s}} = col_{1 \leq d \leq D} \mathbf{X}_{\mathbf{sd}}$.

In the previous paragraph the problem of the MSE estimation was considered but for the empirical version of Henderson (1950) BLUP while in this paper empirical version of the BLUP proposed by Royall (1976) is studied. Using our notation Royall (1976) derived the BLUP of domain characteristic defined as a linear combination $\gamma^T \mathbf{Y_d} = \gamma^T (\mathbf{X_d} \beta_{\mathbf{d}} + \mathbf{Z_d} \mathbf{v_d} + \mathbf{e_d})$ while Henderson (1950) derived BLUP of $\gamma^T (\mathbf{X_d} \beta_{\mathbf{d}} + \mathbf{Z_d} \mathbf{v_d})$, where γ is a known vector. Hence, the problem studied by Henderson (1950) may be treated as a special case of the problem considered by Royall (1976). The MSE estimator of the empirical version of Royall (1976) BLUP is proposed by Żądło (2009). He presented proof (under some regularity conditions) that the bias of derived MSE estimator is of order $o(D^{-1})$. The proof is a direct generalization of the results presented by Datta and Lahiri (2000) for the empirical version of Henderson (1950) BLUP. MSE estimators presented below are special cases of the estimators derived by Żądło (2009) where it assumed that the variance-covariance matrix is a linear combination of unknown variance parameters. For the proposed model (1) the assumption is not met what means that the order of approximation of MSE given by the equation (9) and the order of the bias of the MSE estimators presented below [see (10) and (11)] are not proven to be $o(D^{-1})$.

Applying results presented by Żądło (2009) under the model (1) for ML and REML estimators of δ we obtain:

$$MSE_\xi(\hat{\theta}_{d*t*}^{EBLU}(\hat{\boldsymbol{\delta}})) \approx g_1(\delta) + g_2(\delta) + g_3^*(\delta), \tag{9}$$

where $g_3^*(\delta)$ is given by (13) in the "Appendix".

Applying results presented by Żądło (2009) under the model (1) the MSE estimator for REML estimators of δ is given by:

$$\widehat{MSE}_\xi \left(\hat{\theta}_{d*t*}^{EBLU}(\hat{\pmb{\delta}}) \right) = g_1(\hat{\pmb{\delta}}) + g_2(\hat{\pmb{\delta}}) + 2g_3^*(\hat{\pmb{\delta}}) \tag{10}$$

and for ML estimators of δ by:

$$\widehat{MSE}_\xi \left(\hat{\theta}_{d*t*}^{EBLU}(\hat{\pmb{\delta}}) \right) = g_1(\hat{\pmb{\delta}}) + g_2(\hat{\pmb{\delta}}) + 2g_3^*(\hat{\pmb{\delta}}) - \mathbf{B}_{\hat{\pmb{\delta}}}^T(\hat{\pmb{\delta}}) \frac{\partial g_1(\hat{\pmb{\delta}})}{\partial \delta}, \tag{11}$$

where for the proposed model (1) $g_1(\hat{\pmb{\delta}})$, $g_2(\hat{\pmb{\delta}})$, $g_3^*(\hat{\pmb{\delta}})$ are given by (7), (8), (13) where δ is replaced by $\hat{\pmb{\delta}}$. $\mathbf{B}_{\hat{\pmb{\delta}}}^T(\hat{\pmb{\delta}})$ and $\frac{\partial g_1(\hat{\pmb{\delta}})}{\partial \delta}$ are given by $\mathbf{B}_{\hat{\pmb{\delta}}}^T(\delta)$ and $\frac{\partial g_1(\delta)}{\partial \delta}$ where δ is replaced by $\hat{\pmb{\delta}}$. $\mathbf{B}_{\hat{\pmb{\delta}}}^T(\delta)$ for ML and REML estimators of δ are given in the Appendix by (14) and (15) respectively. The elements of $\frac{\partial g_1(\delta)}{\partial \delta}$ are given in the "Appendix" by (16)–(19).

In the simulation study the proposed MSE estimator will be compared with delete-one-domain jackknife MSE estimator proposed by Chen and Lahiri (2002). For the proposed model (1) it is given by

$$\widehat{MSE}_\xi^{jack}(\hat{\theta}_{d*t*}^{EBLU}(\hat{\pmb{\delta}})) = b_{d*t*}(\hat{\pmb{\delta}}) - \frac{D-1}{D} \sum_{d=1}^{D} \left(b_{d*t*}(\hat{\pmb{\delta}}_{-d}) - b_{d*t*}(\hat{\pmb{\delta}}) \right)$$

$$+ \frac{D-1}{D} \sum_{d=1}^{D} \left(\hat{\theta}_{d*t*}^{EBLU}(\hat{\pmb{\delta}}_{-d}) - \hat{\theta}_{d*t*}^{EBLU}(\hat{\pmb{\delta}}) \right)^2, \tag{12}$$

where $\hat{\pmb{\delta}}_{-d}$ is an estimator given by the same formula as $\hat{\pmb{\delta}}$ but based on the data without the dth domain, $b_{d*t*}(\hat{\pmb{\delta}}) = g_1(\hat{\pmb{\delta}}) + g_2(\hat{\pmb{\delta}})$, $g_1(\hat{\pmb{\delta}})$, $g_2(\hat{\pmb{\delta}})$ are given by (7) and (8) respectively, where δ is replaced by $\hat{\pmb{\delta}}$, $b_{d*t*}(\hat{\pmb{\delta}}_{-d})$ is given by $b_{d*t*}(\hat{\pmb{\delta}})$, where $\hat{\pmb{\delta}}$ is replaced by $\hat{\pmb{\delta}}_{-d}$, $\hat{\theta}_{d*t*}^{EBLU}(\hat{\pmb{\delta}}_{-d})$ is given by (5) where δ is replaced by $\hat{\pmb{\delta}}_{-d}$.

It is known, that parametric bootstrap distribution approximates the true distribution of the EBLUP very well—see the proof presented by Chatterjee et al. (2008). Hence, it is also possible to use the parametric bootstrap method to estimate the MSE of the EBLUP. The problem for unit-level models in small area estimation is considered inter alia by González et al. (2007), González et al. (2008). In each iteration of both jackknife and bootstrap methods we need to estimate parameters of the model (what is time-consuming). Because the number of iterations in the delete-one-domain jackknife procedure for the data considered in the Sects. 6 and 7 is several times smaller than in the bootstrap method we will use the jackknife method to estimate the MSE in the Monte Carlo simulation studies.

6 Monte Carlo simulation study: artificial data

The simulation study was conducted using R package (R Development Core Team 2013). It is based on artificial longitudinal data from $M = 3$ periods. The population size in each period equals $N = 400$ elements. The population consists of $D = 20$ domains (subpopulations) each of size 10 elements. The balanced panel sample is

considered—in each period the same 40 elements are observed. The sample sizes in $D = 20$ domains are: 1 for seven domains, 2 for six domains and 3 for seven domains. Model parameters are estimated using restricted maximum likelihood method—we wrote restricted likelihood function for the model using R language and then we use *constrOptim* function available in *stats* R package to find the maximum. The number of iterations in Monte Carlo simulation study is $L = 2000$. In the simulation study the simulation MSE of the EBLUP is computed as $L^{-1} \sum_{l=1}^{L} (\hat{\theta}_{d*t*}^{EBLU(l)}(\hat{\delta}) - \theta_{d*t*}^{(l)})^2$, the simulation bias of the EBLUP as $L^{-1} \sum_{l=1}^{L} (\hat{\theta}_{d*t*}^{EBLU(l)}(\hat{\delta}) - \theta_{d*t*}^{(l)})$ and the simulation bias of the MSE estimator as $L^{-1} \sum_{l=1}^{L} (\widehat{MSE}_{\xi}^{(l)}(\hat{\theta}_{d*t*}^{EBLU(l)}(\hat{\delta})) - MSE_{\xi}(\hat{\theta}_{d*t*}^{EBLU}(\hat{\delta})))$, where $\hat{\theta}_{d*t*}^{EBLU(l)}(\hat{\delta})$, $\theta_{d*t*}^{(l)}$ and $\widehat{MSE}_{\xi}^{(l)}(\hat{\theta}_{d*t*}^{EBLU(l)}(\hat{\delta}))$ are values of the EBLUP, the domain total and the MSE estimator computed in lth iteration of the simulation study.

In the simulation data are generated based on the model (1) assuming arbitrary chosen parameters: different values of $\lambda_{(sp)}$ and $\lambda_{(t)}$, $\sigma_u^2 = 1$, $\sigma_\varepsilon^2 = 1$, $\forall_d \beta_d = \beta = 100$, $\mathbf{X_{id}} = \mathbf{Z_{id}} = [1]_{M_{id} \times 1}$. The spatial weight matrix (denoted by $\mathbf{W_d}$) is row-standardized neighborhood matrix (each population element has two neighbors). In the simulation study three predictors are considered:

– spatial BLUP (SBLUP) given by (5) where variance parameters are assumed to be known,
– spatial EBLUP (SEBLUP), given by (5) where variance components are replaced by REML estimates,
– BLUP under the assumption that $\lambda_{(sp)} = 0$ and $\lambda_{(t)} = 0$ (BLUPind). The BLUPind under the model and for balanced panel sample (all realizations of random variables within one profile are observed or are not observed) does not depend on unknown model parameters.

Because we are mainly interested in the spatial effect in the simulation we assumed $\lambda_{(t)} = \{-0.5, 0.5\}$ and $\lambda_{(sp)} = \{-0.9, -0.6, 0.6, 0.9\}$. In our opinion the comparison of accuracies of the SEBLUP and its simplified version (under assumption of the lack of spatio-temporal correlation of random effects and components) is crucial because the predictor is the natural alternative of the SEBLUP. What is important, the comparison measures the effect of including spatio-temporal correlation. Additional comparison between mean squared errors of SEBLUP and SBLUP is also important because it allows to measure the loss of accuracy due to the estimation of model parameters.

In each figure squares denote values of some statistic for one out of $D = 20$ domains and the black squares denote the mean values of the statistic over $D = 20$ domains. Hence, we do not present only the mean values of the considered statistics but their whole distribution [as e.g. simulation results presented by Białek (2014)]. In the Fig. 1 it is shown that ratios of mean squared errors of BLUPind and SEBLUP for all of domains and different values of $\lambda_{(t)}$ and $\lambda_{(sp)}$ are from 1.004 to 1.131. It means that the maximum gain in accuracy due to the inclusion of spatio-temporal correlation is 13.1%. Because we compare the MSE of BLUPind and the MSE of SEBLUP (not SBLUP) the decrease of accuracy due to the estimation of model parameters is taken into account.

What is important, the decrease of accuracy due to the estimation of model parameters presented in the Fig. 2 for all of domains and different values of $\lambda_{(t)}$ and $\lambda_{(sp)}$
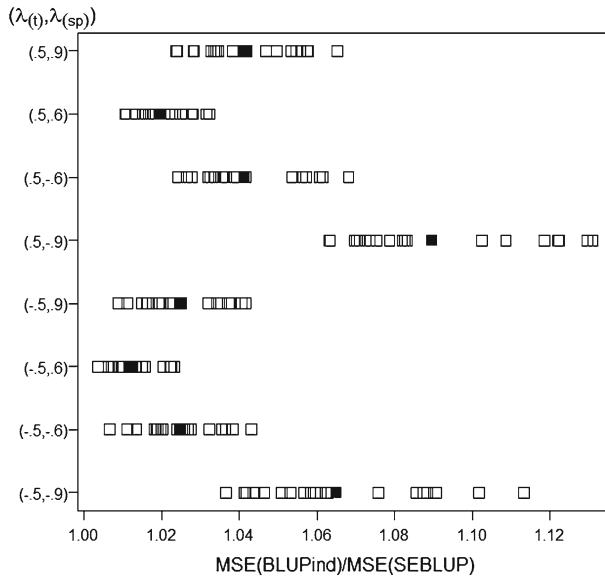
Fig. 1 Effect of including spatio-temporal correlation for different values $\lambda_{(t)}$ and $\lambda_{(sp)}$ (model parameters are estimated)
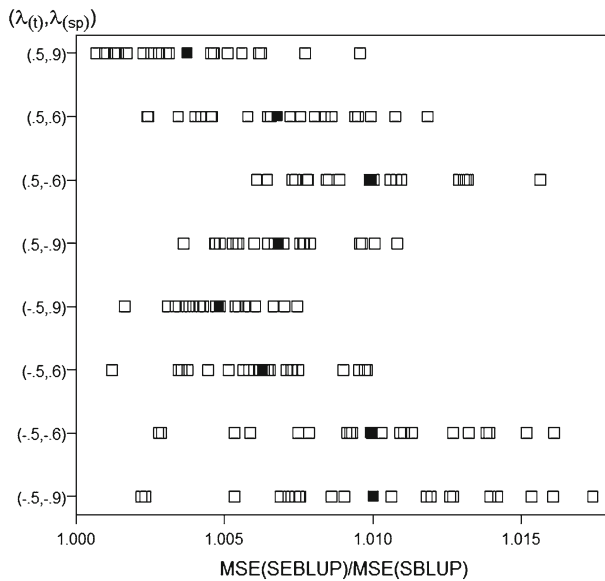

Fig. 2 Effect of estimation of model parameters for different values $\lambda_{(t)}$ and $\lambda_{(sp)}$

is very small—from 0.1 to 1.7 %. It means that its influence on results presented in the Fig. 1 is not large.

Approximately unbiasedness of the MSE estimator (10) is not proven but the biases presented in the Fig. 3 are not high—for $D = 20$ domains and for different values of
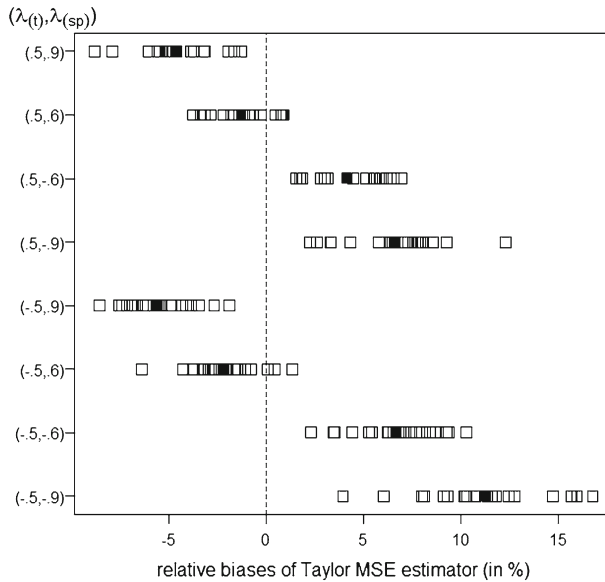
Fig. 3 Biases of MSE estimator based on Taylor expansion for different values $\lambda_{(t)}$ and $\lambda_{(sp)}$
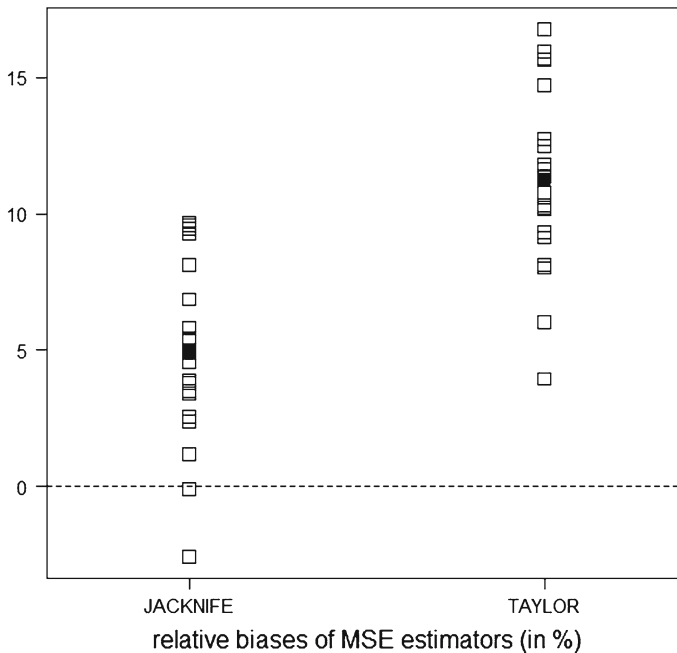
Fig. 4 Biases of MSE estimators (in %) for $\lambda_{(t)} = -0.5$ and $\lambda_{(sp)} = -0.9$

$\lambda_{(sp)}$ and $\lambda_{(t)}$—from ca $-8.8\,\%$ to ca $16.8\,\%$ (with mean ca $1.9\,\%$). In the Fig. 4 biases of two MSE estimators (10) and (12) are compared for $\lambda_{(t)} = -0.5$ and $\lambda_{(sp)} = -0.9$ where we observed (see Fig. 3) the highest bias for the proposed MSE estimator based

on the Taylor expansion. In the Fig. 3 it is shown that the jackknife estimator may give significantly better results although it is not the rule (compare with the Fig. 7 for real data).

7 Monte Carlo simulation study: real data

The second simulation study was also conducted using R package (R Development Core Team 2013) and model parameters are estimated using R as described in the previous section. The number of iterations in Monte Carlo simulation study is $L = 2000$. We consider real data on investments of Polish companies (in million PLN) in $N = 378$ regions called poviats (NUTS 4) in $M = 3$ years 2009–2011. We consider the balanced panel sample—in the first period a sample of size $n = 38$ using (arbitrarily chosen) Midzuno (1952) sampling scheme is selected and the same elements are in the sample in $M = 3$ periods. The population is divided into $D = 28$ domains according to larger regions called voivodships (NUTS 3) and types of poviats (city poviats and land poviats) within voivodships. In 7 out of $D = 28$ domains sample size equals 0. The spatial weight matrix is the row-standardized neighborhood matrix. The neighborhood matrix is constructed based on the 2-nearest neighbors role using auxiliary variable—the number of new companies registered in the poviat. Data are generated based on the model (1) where the values of all of the model parameters are obtained based on the whole population data using REML and assuming that $\forall_d \beta_d = \beta$, $\mathbf{X_{id}} = \mathbf{Z_{id}} = [1]_{M_{id} \times 1}$. We assume $\forall_d \beta_d = \beta$ because for the considered case we have no observations from some of domains in all of periods (what implies that it is not possible to estimate some of β_d's). What is important, the spatial and temporal correlations for the real data are weak: $\lambda_{(t)} = 0.352$ and $\lambda_{(sp)} = -0.396$. In the model-based simulation study we compare accuracies of the following predictors and estimators of the domain total in the last period:

- spatial BLUP (SBLUP), given by (5), where variance parameters are assumed to be known,
- spatial EBLUP (SEBLUP), given by (5), where variance parameters are replaced by REML estimates,
- BLUP under the assumption that $\lambda_{(sp)} = 0$ and $\lambda_{(t)} = 0$ (BLUPind) which under the model and for the balanced panel sample does not depend on unknown model parameters,
- Count Synthetic Estimator (C-SYN), see Rao (2003, p. 46),
- Ratio Synthetic Estimator (R-SYN), see Rao (2003, p. 47), where the auxiliary variable is the number of new companies registered in the poviat in 2011,
- Generalized Regression Estimator (GREG), see Rao (2003, p. 17), where the auxiliary variable used in the calibration equation is the number of new companies registered in the poviat in 2011,
- the longitudinal version of Generalized Regression Estimator (GREG-L), see Rao (2003, p. 17), which differs from GREG in that in the calibration equation three auxiliary variables are used: numbers of new companies registered in poviats in 2009, 2010 and 2011,
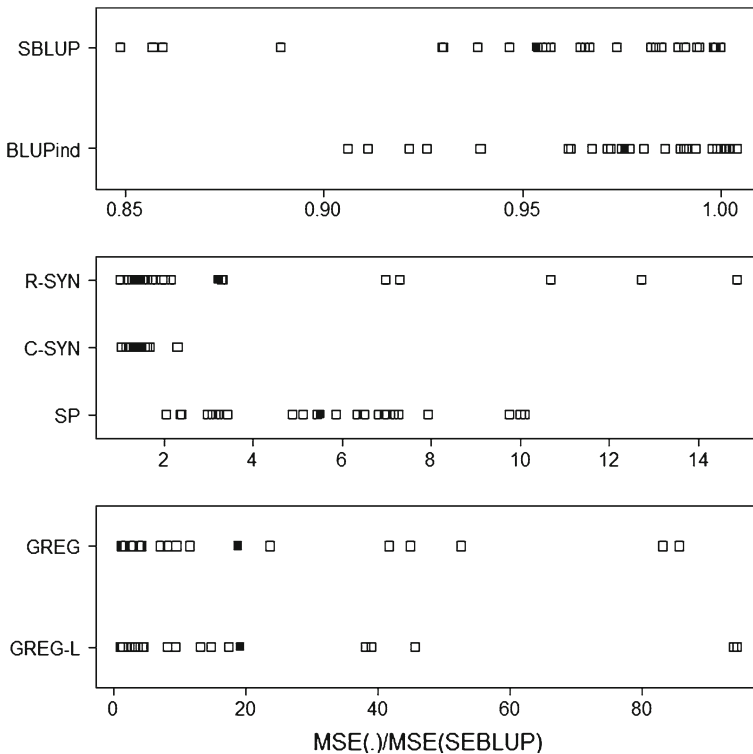
Fig. 5 Ratios of predictors MSEs

– the predictor proposed in the SAMPLE project (SP) which is the EBLUP under the following unit-level model with correlated time effects (Molina et al. 2010a, p. 143): $Y_{idj} = x_{idj}\beta + u_{1,d} + u_{2,dj} + e_{idj}$ where $e_{idj} \sim (0, \sigma_0^2)$, domain specific $u_{1,d}$ are independent and $u_{1,d} \sim (0, \sigma_1^2)$, time-varying area effects $u_{2,dj}$ for $d = 1, 2, \ldots, D$ are independent, but inside domains for $j = 1, 2, \ldots, M$ are AR(1) with parameters denoted by σ_2^2 and $\rho_{(t)}$. The predictor does not take the spatial correlation into account. The temporal autocorrelation is included but on the higher aggregation level—within domains instead of within profiles as in (1). To compute values of the predictor function in R language presented in Molina et al. (2010b, pp. 123–126) is used.

SEBLUP, SBLUP, BLUPind, SP use information on the variable of interest from all of the periods while C-SYN, R-SYN, GREG and GREG-L use information on the variable of interest only from the last period. GREG and GREG-L are direct estimators what means that is it possible to compute their values only for domains with sample sizes greater than zero in the period of interest (in 21 out of $D = 28$ domains in the simulation study).

In the Fig. 5 the accuracy of SEBLUP is compared with other predictors and estimators. Estimators and predictors R-SYN, C-SYN, GREG, GREG-L and SP are several times less accurate than SEBLUP. What is interesting, in 22 out of $D = 28$ domains
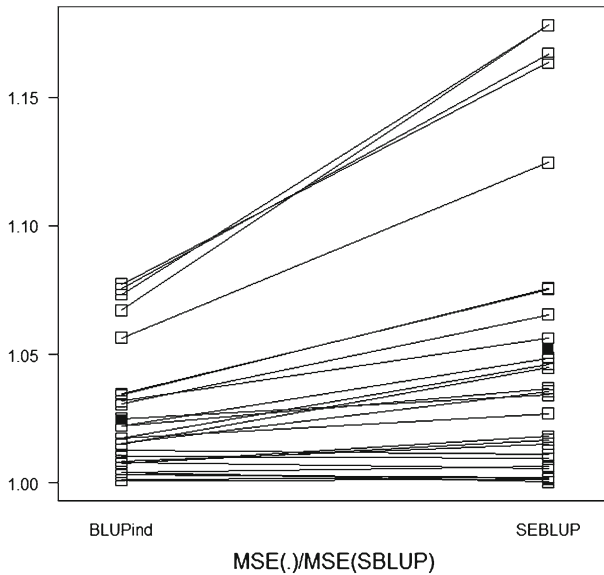
Fig. 6 Effect of including spatio-temporal correlation (assuming that model parameters are known) and effect of estimation of model parameters

SEBLUP is less accurate than BLUPind. The situation is explained in the Fig. 6 (the results for the same domains are matched by lines). The reason is that the gain in accuracy due to the including spatio-temporal correlation (assuming that model parameters are known) measured by ratios MSE(BLUPind)/MSE(SBLUP) is in 22 domains smaller than the increase of MSE due to the estimation of model parameters measured by ratios MSE(SEBLUP)/MSE(SBLUP). It explains the suggestions presented in the previous section that the comparison of SEBLUP and its simplified version (assuming the lack of spatio-temporal correlation) is very important or even crucial.

In the Fig. 7 biases of two MSE estimators (10) and (12) are compared. For the studied case means of absolute biases are similar (see the right part of the Fig. 7). For the jackknife MSE estimators it equals 5.1 % while for the MSE estimator based on the Taylor expansion it equals 4.8 %.

8 Case study: real data

In the previous section we have studied the problem of prediction of total value of investments of Polish companies (in million PLN) in $D = 28$ regions in 2011 in the simulation study. Because we were interested in the gain in accuracy which resulted only from incorporating spatio-temporal correlation we did not use auxiliary information. In this section we will use the same data to show how to choose the appropriate model based on the real data. In this section we will use data on investments of Polish companies in 2009–2011 (the same as in the previous section) and additionally two
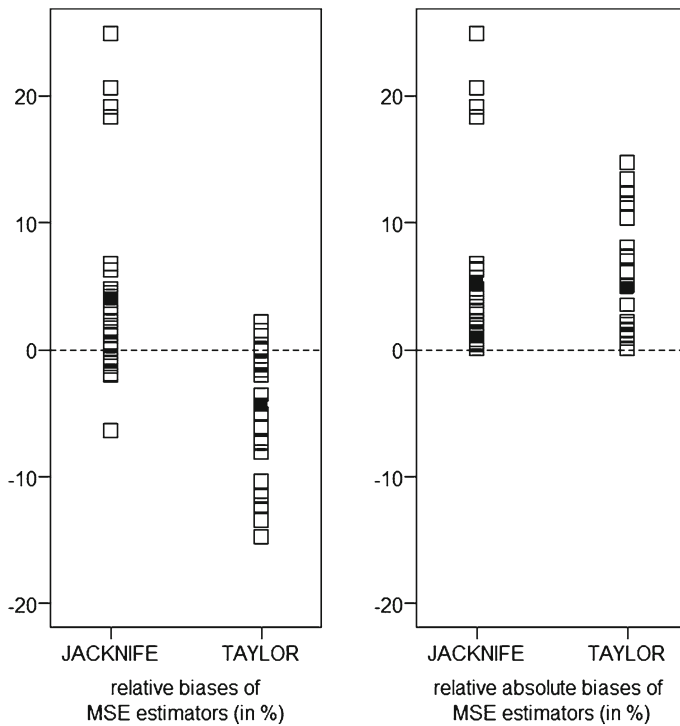
Fig. 7 Biases and absolute biases of MSE estimators

auxiliary variables: the production sold (in million PLN) and fixed assets (in million PLN) but both with one year lag (i.e. for years 2008–2010). The same sample as in the previous section is studied. Firstly, we would like to find the appropriate model for the real data. Is is possible to use the likelihood ratio test to compare two models but if the models are nested (see e.g. Pinheiro and Bates 2000, pp. 83–84). Hence, at the significance level 0.05, we compare our model with two auxiliary variables with its special cases with two auxiliary variables as well but under simplified assumptions on spatio-temporal correlation (obtaining the following p values):

- assuming the independence of random effects and the independence of random components (p value of likelihood ratio test: 1.1×10^{-8})
- assuming the independence of random effects and MA(1) random components (p value of likelihood ratio test: 2.8×10^{-9})
- assuming the spatial moving average model for random effects and independence of random components (p value of likelihood ratio test: 0.0306).

Hence, our model should be preferred comparing with its special cases. Pinheiro and Bates (2000) in chapter 5 suggest using e.g Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) if we would like to compare non nested models. Moreover, the authors present different models available in R which will be compared in this section with the proposed model (1). It is possible to include other models as well but in this case the computations must be conducted using original functions (as

in the case of the proposed model). Pinheiro and Bates (2000) in chapter 5 present special cases of the linear mixed models where different assumptions on correlation structure of random components can be made but assuming the independence of random components within groups defined by the grouping variable used for the random effects. Hence, if we assume profile specific random effects we can define different temporal models for random components within profiles, and if we define time specific random effects we can define different spatial models for random components within domains. Below we use different correlation structures described by Pinheiro and Bates (2000) in chapter 5 including different spatial correlation structures defined in Pinheiro and Bates (2000, p. 232).

In the Table 1 we present the values of the AIC and BIC criteria of the proposed model and other non nested models:

– with independent profile specific random effects and MA(2) random components (model_i_MA2)
– with independent profile specific random effects and AR(1) random components (model_i_AR1)
– with independent profile specific random effects and AR(2) random components (model_i_AR2)
– with independent profile specific random effects and ARMA(1,1) random components (model_i_ARMA)
– with independent profile specific random effects and compound symmetry temporal correlation of random components (model_i_compound_symmetry)
– with independent time specific random effects and independent random components (model_t)
– with independent time specific random effects and exponential spatial correlation of random components (model_t_exponential)
– with independent time specific random effects and gaussian spatial correlation of random components (model_t_gaussian)
– with independent time specific random effects and linear spatial correlation of random components (model_t_linear)
– with independent time specific random effects and rational quadratic spatial correlation of random components (model_t_rational_quadratic)
– with independent time specific random effects and spherical spatial correlation of random components (model_t_spherical)
– with independent time specific random effects and compound symmetry spatial correlation of random components (model_t_compound_symmetry)
– with independent domain specific random effects and independent random components (model_d)

The proposed model has the smallest values of AIC and BIC criteria comparing with other analyzed models. It is worth noting that the values of the criteria for some models are the same what is not unusual—see eg. Pinheiro and Bates (2000, p. 249) where 4 out 5 models with different spatial correlation structures have the same values of AIC and BIC criteria.

We have also compared our model with models with the same variance-covariance matrices as the models presented in the Table 1 but using only one out of two auxiliary

**Table 1** Values of AIC and BIC citeria for different models with two auxiliary variables

Model	AIC	BIC
Proposed model	1424.172	1443.325
Model_i_MA2	1459.464	1478.43
Model_i_AR1	1457.464	1473.721
Model_i_AR2	1459.464	1478.43
Model_i_ARMA	1459.464	1478.43
Model_i_compound_symmetry	1458.758	1475.015
Model_t	1457.252	1470.799
Model_t_exponential	1459.001	1475.258
Model_t_gaussian	1458.97	1475.227
Model_t_linear	1458.969	1475.227
Model_t_rational_quadratic	1459.031	1475.289
Model_t_spherical	1458.969	1475.227
Model_t_compound_symmetry	1459.252	1475.509
Model_d	1455.049	1468.597

variables. these models have also higher values of AIC and BIC criteria than the proposed model. Although the assumed model with only one out of two auxiliary variables has higher values of AIC and BIC criteria the formal test of significance of fixed effects will be conducted as well. In the section we will use permutation tests of fixed effects. The algorithm for testing the jth fixed effect is as follows (Pesarin and Salmaso 2010, p. 45):

1. Based on the original data a test statistic, denoted by $T_0 = T(\mathbf{X})$, is computed, e.g. the test statistic can be defined as log-likelihood (as in this paper).
2. We take a random permutation of jth column of the matrix $\mathbf{X}$ and we obtain a new matrix of auxiliary variables denoted by $\mathbf{X}^*$.
3. Value of the test statistics $T^* = T(\mathbf{X}^*)$ is computed.
4. Steps 2 and 3 are repeated B times and B values of $T^{*b} = T(\mathbf{X}^{*b})$ are computed, where $b = 1, 2, \ldots, B$.
5. We estimate p value as $B^{-1} \sum_{1 \leq b \leq B} I(T^{*b} \geq T_0)$—he fraction of the permutation values not smaller than the the value of the test statistic computed based on the real data.

If is not possible to make computations for all possible permutations, the estimated p value strongly converges to its respective true value (Pesarin and Salmaso 2010, p. 45). In the case study the number of all possible permutations is $(n \times M)! = (38 \times 3)! \approx 2.5 \times 10^{186}$. Hence, p-values will be computed based on $B = 1000$ independent permutations. Let us consider tests of fixed effects for two auxiliary variables (production sold and fixed assets). In both cases p-values of permutation test equal 0, what means that the variables have a significant influence on the variable of interest.

Finally, in the Fig. 8 we present real values of domain totals of investments and the predicted values—values of the empirical version of the proposed predictor given by (5) based on the sample data considered in the section. It should be noted that the
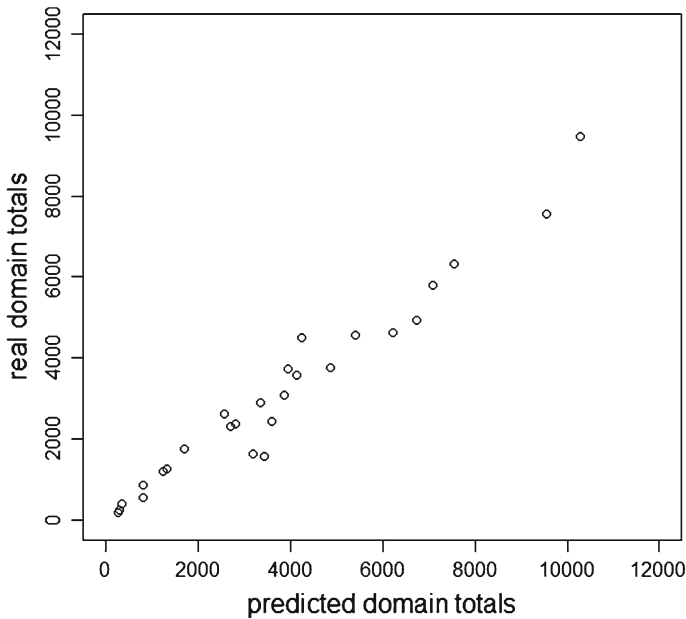
Fig. 8 Real and predicted domain totals of investments (in million PLN) in $D = 28$ domains

values of the predictors are computed based on the following small sample sizes in the period of interest (in 2011):

– zero for 7 out of $D = 28$ domains,
– one for 11 out of $D = 28$ domains,
– two for 5 out of $D = 28$ domains,
– three for 3 out of $D = 28$ domains,
– four for 2 out of $D = 28$ domains.

9 Conclusions

In the paper some special case of the LMM for longitudinal data is proposed. The BLUP of the subpopulation total for the model is derived and MSE estimators of its empirical version are proposed. The accuracy of the proposed predictor and biases of the proposed MSE estimators are analyzed in two Monte Carlo simulation studies based on the artificial and the real data. In the first simulation study based on the artificial data the accuracy of the empirical version of the proposed predictor was better for all of the domains comparing with the predictor derived under the assumption of lack of spatio-temporal correlation. In the second simulation study based on the real data the empirical version of the proposed predictor was even several times more accurate than other predictors and estimators but it was better than the predictor derived under the assumption of lack of spatio-temporal correlation only in 6 out of 28 domains. It resulted from the decrease of the accuracy due to the estimation of model parameters.

In both simulation studies biases of the proposed MSE estimator were small. The considerations are also supported by the case study.

Appendix: Elements of MSE and its estimator

Let us introduce the following notations. Let $\mathbf{A_d^{MA}} = \partial\mathbf{H_d}/\partial\lambda_{(sp)} = (\mathbf{W_d} + \mathbf{W_d^T}) + 2\lambda_{(sp)}\mathbf{W_d}\mathbf{W_d^T}$ and $\mathbf{B_{id}^{MA}} = \partial\Gamma_{\mathbf{id}}/\partial\lambda_{(t)}$. Let $\mathbf{B_{ssid*}^{MA}}$ be submatrix obtained from $\mathbf{B_{id}^{MA}}$ by deleting rows and columns for unsampled observations. Let $\mathbf{B_{rsid*}^{MA}}$ be submatrix obtained from $\mathbf{B_{id}^{MA}}$ by deleting rows for sampled observations and columns for unsampled observations.

The third element of MSE of EBLUP (9) is given by:

$$g_3^*(\delta) = tr\left(\frac{\partial, \mathbf{c_{d*}^T}}{\partial\delta}\mathbf{V_{ssd*}}\left(\frac{\partial, \mathbf{c_{d*}^T}}{\partial\delta}\right)^T \mathbf{I_\delta^{-1}}\right), \tag{13}$$

$$\frac{\partial\,\mathbf{c_{d*}^T}}{\partial\delta} = col_{1\leq k\leq q}\frac{\partial\,\mathbf{c_{d*}^T}}{\partial\delta_k} = col_{1\leq k\leq q}\frac{\partial\gamma_{\mathbf{rd*}}^T\mathbf{V_{rsd*}}\mathbf{V_{ssd*}^{-1}}}{\partial\delta_k},$$

$$\gamma_{\mathbf{rd*}}^T\mathbf{V_{rsd*}}\mathbf{V_{ssd*}^{-1}} = \gamma_{\mathbf{rd*}}^T\left(\sigma_u^2\mathbf{Z_{rd*}}\mathbf{H_{d*}}\mathbf{Z_{sd*}^T} + diag_{1\leq i\leq N_{rd*}}(\Gamma_{\mathbf{rsid*}})\right)$$
$$\times\left(\sigma_u^2\mathbf{Z_{sd*}}\mathbf{H_{d*}}\mathbf{Z_{sd*}^T} + diag_{1\leq i\leq n_{d*}}(\Gamma_{\mathbf{sid*}})\right)^{-1},$$

$$\frac{\partial\mathbf{c_{d*}^T}}{\partial\sigma_u^2} = \gamma_{\mathbf{rd*}}^T\left[\mathbf{Z_{rd*}} - \left(\sigma_u^2\mathbf{Z_{rd*}}\mathbf{H_{d*}}\mathbf{Z_{sd*}^T} + diag_{1\leq i\leq N_{rd*}}(\Gamma_{\mathbf{rsid*}})\right)\mathbf{V_{ssd*}^{-1}}\mathbf{Z_{sd*}}\right]$$
$$\times\mathbf{H_{d*}}\mathbf{Z_{sd*}^T}\mathbf{V_{ssd*}^{-1}},$$

$$\frac{\partial\mathbf{c_{d*}^T}}{\partial\sigma_\varepsilon^2} = \gamma_{\mathbf{rd*}}^T\left(\sigma_\varepsilon^{-2}diag_{1\leq i\leq N_{rd*}}(\Gamma_{\mathbf{rsid*}})\right)\mathbf{V_{ssd*}^{-1}}$$
$$- \gamma_{\mathbf{rd*}}^T\left(\sigma_u^2\mathbf{Z_{rd*}}\mathbf{H_{d*}}\mathbf{Z_{sd*}^T} + diag_{1\leq i\leq N_{rd*}}(\Gamma_{\mathbf{rsid*}})\right)$$
$$\times\mathbf{V_{ssd*}^{-1}}\sigma_\varepsilon^{-2}diag_{1\leq i\leq n_{d*}}(\Gamma_{\mathbf{sid*}})\mathbf{V_{ssd*}^{-1}},$$

$$\frac{\partial \mathbf{c}_{\mathbf{d}*}^T}{\partial \lambda_{(sp)}} = \gamma_{\mathbf{rd}*}^T \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd}*} \mathbf{A}_{\mathbf{d}*}^{\mathbf{MA}} \mathbf{Z}_{\mathbf{sd}*}^T \right) \mathbf{V}_{\mathbf{ssd}*}^{-1}$$

$$- \gamma_{\mathbf{rd}*}^T \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd}*} \mathbf{H}_{\mathbf{d}*} \mathbf{Z}_{\mathbf{sd}*}^T + diag_{1 \leq i \leq N_{rd*}} (\Gamma_{\mathbf{rsid}*}) \right)$$

$$\times \mathbf{V}_{\mathbf{ssd}*}^{-1} \left(\sigma_u^2 \mathbf{Z}_{\mathbf{sd}*} \mathbf{A}_{\mathbf{d}*}^{\mathbf{MA}} \mathbf{Z}_{\mathbf{sd}*}^T \right) \mathbf{V}_{\mathbf{ssd}*}^{-1},$$

$$\frac{\partial \mathbf{c}_{\mathbf{d}*}^T}{\partial \rho_{(t)}} = \gamma_{\mathbf{rd}*}^T \left(diag_{1 \leq i \leq N_{rd*}} (\mathbf{B}_{\mathbf{rsid}*}^{\mathbf{MA}}) \right) \mathbf{V}_{\mathbf{ssd}*}^{-1}$$

$$- \gamma_{\mathbf{rd}*}^T \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd}*} \mathbf{H}_{\mathbf{d}*} \mathbf{Z}_{\mathbf{sd}*}^T + diag_{1 \leq i \leq N_{rd*}} (\Gamma_{\mathbf{rsid}*}) \right)$$

$$\times \mathbf{V}_{\mathbf{ssd}*}^{-1} diag_{1 \leq i \leq n_{d*}} (\mathbf{B}_{\mathbf{sid}*}^{\mathbf{MA}}) \mathbf{V}_{\mathbf{ssd}*}^{-1},$$

klth element of $\mathbf{I}_\delta$ is given by $\mathbf{I}_{kl}(\delta) = \frac{1}{2} tr \left(\mathbf{V}_{\mathbf{ss}}^{-1} \frac{\partial \mathbf{V}_{\mathbf{ss}}}{\partial \delta_k} \mathbf{V}_{\mathbf{ss}}^{-1} \frac{\partial \mathbf{V}_{\mathbf{ss}}}{\partial \delta_l} \right)$, where

$$\frac{\partial \mathbf{V}_{\mathbf{ss}}}{\partial \sigma_u^2} = diag_{1 \leq d \leq D} \left(\mathbf{Z}_{\mathbf{sd}} \mathbf{H}_{\mathbf{d}} \mathbf{Z}_{\mathbf{sd}}^T \right),$$

$$\frac{\partial \mathbf{V}_{\mathbf{ss}}}{\partial \sigma_\varepsilon^2} = diag_{1 \leq d \leq D} \left(\sigma_\varepsilon^{-2} diag_{1 \leq i \leq n_d} (\Gamma_{\mathbf{sid}}) \right),$$

$$\frac{\partial \mathbf{V}_{\mathbf{ss}}}{\partial \lambda_{(t)}} = diag_{1 \leq d \leq D} \left(diag_{1 \leq i \leq n_d} (\mathbf{B}_{\mathbf{sid}}^{\mathbf{MA}}) \right),$$

$$\frac{\partial \mathbf{V}_{\mathbf{ss}}}{\partial \lambda_{(sp)}} = diag_{1 \leq d \leq D} \left(\sigma_u^2 \mathbf{Z}_{\mathbf{sd}} \mathbf{A}_{\mathbf{d}}^{\mathbf{MA}} \mathbf{Z}_{\mathbf{sd}}^T \right).$$

Biases of estimators of δ for maximum likelihood method (ML) and restricted maximum likelihood method (REML) are given respectively by general equations presented in Datta and Lahiri (2000):

$$\mathbf{B}_{\hat{\delta}^{ML}}(\delta) = \frac{1}{2} \mathbf{I}_\delta^{-1}(\delta) \, col_{1 \leq k \leq q} tr \left[\mathbf{I}_\beta^{-1}(\delta) \frac{\partial}{\partial \delta_k} \mathbf{I}_\beta(\delta) \right] + o(D^{-1}), \qquad (14)$$

$$\mathbf{B}_{\hat{\delta}^{REML}}(\delta) = o(D^{-1}), \qquad (15)$$

where for the proposed model (1):

$$\mathbf{I}_\beta^{-1}(\delta) = diag_{1 \leq d \leq D} \left(\mathbf{X}_{\mathbf{sd}}^T \mathbf{V}_{\mathbf{ssd}}^{-1} \mathbf{X}_{\mathbf{sd}} \right)^{-1},$$

$$\frac{\partial}{\partial \sigma_u^2} \mathbf{I}_\beta(\delta) = diag_{1 \leq d \leq D} \left(\mathbf{X}_{\mathbf{sd}}^T \mathbf{V}_{\mathbf{ssd}}^{-1} \mathbf{Z}_{\mathbf{sd}} \mathbf{H}_{\mathbf{d}} \mathbf{Z}_{\mathbf{sd}}^T \mathbf{V}_{\mathbf{ssd}}^{-1} \mathbf{X}_{\mathbf{sd}} \right),$$

$$\frac{\partial}{\partial \sigma_\varepsilon^2} \mathbf{I}_\beta(\delta) = diag_{1 \leq d \leq D} \left(\mathbf{X}_{\mathbf{sd}}^T \mathbf{V}_{\mathbf{ssd}}^{-1} \sigma_\varepsilon^{-2} diag_{1 \leq i \leq n_d} (\Gamma_{\mathbf{sid}}) \mathbf{V}_{\mathbf{ssd}}^{-1} \mathbf{X}_{\mathbf{sd}} \right),$$

$$\frac{\partial}{\partial \lambda_{(t)}} \mathbf{I}_\beta(\delta) = diag_{1 \leq d \leq D}\left(\mathbf{X}_{\mathbf{sd}}^T \mathbf{V}_{\mathbf{ssd}}^{-1} diag_{1 \leq i \leq n_d}(\mathbf{B}_{\mathbf{sid}}^{\mathbf{MA}}) \mathbf{V}_{\mathbf{ssd}}^{-1} \mathbf{X}_{\mathbf{sd}}\right),$$

$$\frac{\partial}{\partial \lambda_{(sp)}} \mathbf{I}_\beta(\delta) = diag_{1 \leq d \leq D}\left(\mathbf{X}_{\mathbf{sd}}^T \mathbf{V}_{\mathbf{ssd}}^{-1} \sigma_u^2 \mathbf{Z}_{\mathbf{sd}} \mathbf{A}_{\mathbf{d}}^{\mathbf{MA}} \mathbf{Z}_{\mathbf{sd}}^T \mathbf{V}_{\mathbf{ssd}}^{-1} \mathbf{X}_{\mathbf{sd}}\right).$$

Elements of $\frac{\partial g_1(\delta)}{\partial \delta}$ are given by

$$\begin{aligned}
\frac{\partial g_1(\delta)}{\partial \sigma_u^2} &= \gamma_{\mathbf{rd*}}^T \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{rd*}}^T \gamma_{\mathbf{rd*}} - 2\gamma_{\mathbf{rd*}}^T \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{sd*}}^T \mathbf{V}_{\mathbf{ssd*}}^{-1} \\
&\quad \times \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{sd*}}^T + diag_{1 \leq i \leq N_{rd*}}(\Gamma_{\mathbf{rsid*}})\right)^T \gamma_{\mathbf{rd*}} \\
&\quad + \gamma_{\mathbf{rd*}}^T \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{sd*}}^T + diag_{1 \leq i \leq N_{rd*}}(\Gamma_{\mathbf{rsid*}})\right) \\
&\quad \times \mathbf{V}_{\mathbf{ssd*}}^{-1} \mathbf{Z}_{\mathbf{sd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{sd*}}^T \mathbf{V}_{\mathbf{ssd*}}^{-1} \\
&\quad \times \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{sd*}}^T + diag_{1 \leq i \leq N_{rd*}}(\Gamma_{\mathbf{rsid*}})\right)^T \gamma_{\mathbf{rd*}}, \quad (16)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial g_1(\delta)}{\partial \sigma_\varepsilon^2} &= \gamma_{\mathbf{rd*}}^T \sigma_\varepsilon^{-2} diag_{1 \leq i \leq N_{rd*}}(\Gamma_{\mathbf{rid*}}) \gamma_{\mathbf{rd*}} \\
&\quad - 2\gamma_{\mathbf{rd*}}^T \left(\sigma_\varepsilon^{-2} diag_{1 \leq i \leq N_{rd*}}(\Gamma_{\mathbf{rsid*}})\right) \mathbf{V}_{\mathbf{ssd*}}^{-1} \\
&\quad \times \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{sd*}}^T + diag_{1 \leq i \leq N_{rd*}}(\Gamma_{\mathbf{rsid*}})\right)^T \gamma_{\mathbf{rd*}} \\
&\quad + \gamma_{\mathbf{rd*}}^T \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{sd*}}^T + diag_{1 \leq i \leq N_{rd*}}(\Gamma_{\mathbf{rsid*}})\right) \\
&\quad \times \mathbf{V}_{\mathbf{ssd*}}^{-1} \sigma_\varepsilon^{-2} diag_{1 \leq i \leq n_{d*}}(\Gamma_{\mathbf{sid*}}) \mathbf{V}_{\mathbf{ssd*}}^{-1} \\
&\quad \times \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{sd*}}^T + diag_{1 \leq i \leq N_{rd*}}(\Gamma_{\mathbf{rsid*}})\right)^T \gamma_{\mathbf{rd*}}, \quad (17)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial g_1(\delta)}{\partial \lambda_{(sp)}} &= \gamma_{\mathbf{rd*}}^T \sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{A}_{\mathbf{d*}}^{\mathbf{MA}} \mathbf{Z}_{\mathbf{rd*}}^T \gamma_{\mathbf{rd*}} - 2\gamma_{\mathbf{rd*}}^T \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{A}_{\mathbf{d*}}^{\mathbf{MA}} \mathbf{Z}_{\mathbf{sd*}}^T\right) \mathbf{V}_{\mathbf{ssd*}}^{-1} \\
&\quad \times \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{sd*}}^T + diag_{1 \leq i \leq N_{rd*}}(\Gamma_{\mathbf{rsid*}})\right)^T \gamma_{\mathbf{rd*}} \\
&\quad + \gamma_{\mathbf{rd*}}^T \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{sd*}}^T + diag_{1 \leq i \leq N_{rd*}}(\Gamma_{\mathbf{rsid*}})\right) \\
&\quad \times \mathbf{V}_{\mathbf{ssd*}}^{-1} \sigma_u^2 \mathbf{Z}_{\mathbf{sd*}} \mathbf{A}_{\mathbf{d*}}^{\mathbf{MA}} \mathbf{Z}_{\mathbf{sd*}}^T \mathbf{V}_{\mathbf{ssd*}}^{-1} \\
&\quad \times \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd*}} \mathbf{H}_{\mathbf{d*}} \mathbf{Z}_{\mathbf{sd*}}^T + diag_{1 \leq i \leq N_{rd*}}(\Gamma_{\mathbf{rsid*}})\right)^T \gamma_{\mathbf{rd*}}, \quad (18)
\end{aligned}$$

$$\frac{\partial g_1(\delta)}{\partial \lambda_{(t)}} = \gamma_{\mathbf{rd*}}^T \left(diag_{1 \leq i \leq N_{rd*}}(\mathbf{B}_{\mathbf{rid*}}^{\mathbf{MA}})\right) \gamma_{\mathbf{rd*}} - 2\gamma_{\mathbf{rd*}}^T \left(diag_{1 \leq i \leq N_{rd*}}(\mathbf{B}_{\mathbf{rsid*}}^{\mathbf{MA}})\right) \mathbf{V}_{\mathbf{ssd*}}^{-1}$$

$$\times \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd}*} \mathbf{H}_{\mathbf{d}*} \mathbf{Z}_{\mathbf{sd}*}^T + diag_{1 \leq i \leq N_{rd*}} (\Gamma_{\mathbf{rsid}*}) \right)^T \gamma_{\mathbf{rd}*}$$

$$+ \gamma_{\mathbf{rd}*}^T \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd}*} \mathbf{H}_{\mathbf{d}*} \mathbf{Z}_{\mathbf{sd}*}^T + diag_{1 \leq i \leq N_{rd*}} (\Gamma_{\mathbf{rsid}*}) \right)$$

$$\times \mathbf{V}_{\mathbf{ssd}*}^{-1} diag_{1 \leq i \leq n_{d*}} (\mathbf{B}_{\mathbf{sid}*}^{\mathbf{MA}}) \mathbf{V}_{\mathbf{ssd}*}^{-1}$$

$$\times \left(\sigma_u^2 \mathbf{Z}_{\mathbf{rd}*} \mathbf{H}_{\mathbf{d}*} \mathbf{Z}_{\mathbf{sd}*}^T + diag_{1 \leq i \leq N_{rd*}} (\Gamma_{\mathbf{rsid}*}) \right)^T \gamma_{\mathbf{rd}*}. \tag{19}$$

References

Białek J (2014) Simulation study of an original price index formula. Commun Stat Simul Comput 43:285–297

Bernardo R (1996) Best linear unbiased prediction of maize single-cross performance given erroneous inbred relationships. Crop Sci 36:862–866

Chandra H, Salvati N, Chambers R (2007) Small area estimation for spatially correlated populations: a comparison of direct and indirect model-based methods. Stat Transit 8:331–350

Chatterjee S, Lahiri P, Li H (2008) Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models. Ann Stat 36:1221–1245

Chen S, Lahiri P (2002) A weighted jackknife MSPE estimator in small-area estimation. In: Proceedings of the section on survey research methods. American Statistical Association, pp 473–477

Das K, Jiang J, Rao JNK (2004) Mean squared error of empirical predictor. Ann Stat 32:818–840

Datta GS, Lahiri P (2000) A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. Stat Sin 10:613–627

Demidenko E (2004) Mixed models. Theory and application. Wiley, New Jersey

Esteban MD, Morales D, Perez A, Santamaria L (2012) Small area estimation of poverty proportions under area-level time models. Comput Stat Data Anal 56:2840–2855

Fay RE, Herriot RA (1979) Estimates of income for small places: an application of James-Stein procedures to census data. J Am Stat Assoc 74:269–277

González-Manteiga W, Lombardía MJ, Molina I, Morales D, Santamaría L (2007) Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model. Comput Stat Data Anal 51:2720–2733

González-Manteiga W, Lombardía MJ, Molina I, Morales D, Santamaría L (2008) Bootstrap mean squared error of small-area EBLUP. J Stat Comput Simul 78:443–462

Hedeker D, Gibbons RD (2006) Longitudinal data analysis. Wiley, New Jersey

Henderson CR (1950) Estimation of genetic parameters (Abstract). Ann Math Stat 21:309–310

Kackar RN, Harville DA (1981) Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. Commun Stat Ser A 10:1249–1261

Kackar RN, Harville DA (1984) Approximations for standard errors of estimators of fixed and random effects in mixed linear models. J Am Stat Assoc 79:853–862

Marhuenda Y, Molina I, Morales D (2013) Small area estimation with spatio-temporal Fay-Herriot models. Comput Stat Data Anal 58:308–325

Midzuno H (1952) On the sampling system with probability proportional to sum of size. Ann Inst Stat Math 3:99–107

Molina I, Morales D, Pratesi M, Tzavidis N (eds) (2010a) Final small area estimation developments and simulation results, SAMPLE project: small area methods for poverty and living conditions estimates

Molina I, Morales D, Pratesi M, Tzavidis N (eds) (2010b) Software on small area estimation. SAMPLE project: small area methods for poverty and living conditions estimates

Molina I, Salvati N, Pratesi M (2009) Bootstrap for estimating the MSE of the spatial EBLUP. Comput Stat 24:441–458

Pesarin F, Salmaso L (2010) Permutation tests for complex data. Theory, applications and software. Wiley, Chichester

Petrucci A, Pratesi M, Salvati N (2005) Geographic information in small area estimation: small area models and spatially correlated random area effects. Stat Transit 7:609–623

Petrucci A, Salvati N (2006) Small area estimation for spatial correlation in watershed erosion assessment. J Agric Biol Environ Stat 11:169–182

Pinheiro JC, Bates DM (2000) Mixed-effects models in S and S-plus. Springer, New York

Prasad NGN, Rao JNK (1990) The estimation of mean the mean squared error of small area estimators. J Am Stat Assoc 85:163–171

Pratesi M, Salvati N (2008) Small area estimation: the EBLUP estimator based on spatially correlated random area effects. Stat Methods Appl 17:113–141

Rao JNK (2003) Small area estimation. Wiley, New Jersey

Rao JNK, You M (1994) Small-area estimation by combining time-series and cross-sectional data. Can J Stat 22:511–528

R Development Core Team (2013) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Royall RM (1976) The linear least squares prediction approach to two-stage sampling. J Am Stat Assoc 71:657–664

Saei A, Chambers R (2003) Small area estimation under linear and generalized linear mixed models with time an area effects. S3RI methodology working paper M03/15r, University of Southampton

Salvati N, Pratesi M, Tzavidis N, Chambers R (2009) Spatial M-quantile models for small area estimation. Stat Transit 10:251–261

Singh BB, Shukla GK, Kundu D (2005) Spatio-temporal models in small area estimation. Surv Methodol 31:183–195

Ugarte MD, Goicoa T, Militino AF, Durban M (2009) Spline smoothing in small area trend estimation and forecasting. Comput Stat Data Anal 53:3616–3629

Verbeke G, Molenberghs G (2000) Linear mixed models for longitudinal data. Springer, New York

Wolny-Dominiak A (2009) The multi-level factors in insurance rating technique. In: Brozova H (ed) In: Proceedings of 27th international conference mathematical methods in economics 2009. Czech University of Life Sciences, pp 346–351

Żądło T (2004) On unbiasedness of some EBLU predictor. In: Antoch J (ed) In: Proceedings in computational statistics 2004. Physica Verlag, Heidelberg-New York, pp 2019–2026

Żądło T (2009) On MSE of EBLUP. Statistical papers, vol 50, pp 101–118