A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 235, 2010

Tomasz Żądło[*]

ON PREDICTION OF DOMAIN TOTAL BASED ON BALANCED PANEL DATA

Abstract. In the paper we analyze the accuracy of the best linear unbiased predictor (BLUP) of the domain total (see Royall, 1976) assuming a model for longitudinal data with subject specific (element specific) random components (i.e. Verbeke, Molenberghs (2000), Hedeker, Gibbons (2006)) which is a special case of the general linear model (GLM) and the general linear mixed model (GLMM). To estimate the mean square error (MSE) of the BLUP we use the results obtained by Datta and Lahiri (2000) for the predictor proposed by Henderson (1950) and adopt them for the predictor proposed by Royall (1976) what was shown in some general case by Żądło (2007). Considerations are supported by simulation study including some problem of model misspecification in the case of predicting future domain totals.

Key words: small area estimation, best linear unbiased predictors, balanced longitudinal data

I. BASIC NOTATIONS

We analyze balanced longitudinal data in M periods. Number of periods for which observation are available is denoted by m. Hence, if prediction of future domain totals is considered $m < M$, otherwise (when current or past domain totals are predicted) $m = M$. The finite population Ω (which does not change it time) consists of N units. The population is divided into D domains Ω_d (d=1,..., D), each of size N_d ($d = 1,..., D$) and we assume that they do not change in time. For the domain of interest we add a star to the subscript d, for example the domain of interest is denoted by Ω_{d*} and its size by N_{d*}. Let the set of population elements for which observations are available in period j be denoted by s_j and its size by n_j. The set of domain elements for which observations are available in period j is denoted by s_{dj} and its size by n_{dj}. Let: $\Omega_{rdj} = \Omega_d - s_{dj}$, $N_{rdj} = N_d - n_{dj}$. Because we assume that for m periods balanced panel data are available, hence for these periods: $\forall_{j \in \{1,...,m\}} s_j = s$,

---

[*] PhD, Department of Statistics, University of Economics in Katowice.

$\forall_{j\in\{1,...,m\}} n_j = n$, $\forall_{j\in\{1,...,m\}} s_{dj} = s_d$, $\forall_{j\in\{1,...,m\}} n_{dj} = n_d$, $\forall_{j\in\{1,...,m\}} \Omega_{rdj} = \Omega_{rd}$, $\forall_{j\in\{1,...,m\}} N_{rdj} = N_{rd}$. If t is the number of future period (for which we may predict the domain totals) we have: $s_t = \varnothing$, $n_t = 0$, $s_{dt} = \varnothing$, $n_{dt} = 0,$, $\Omega_{rdt} = \Omega_d$, $N_{rdt} = N_d$.

Values of the variable of interest are realizations of Y_{idj} random variables for the population element i (i=1,…,N) which belong to the domain d (d=1,…,D) in the period j (j=1,…,M). The vector of size $M \times 1$ of random variables Y_{idj} for the ith population element which belong to the dth domain is denoted by $\mathbf{Y_{id}} = \left[Y_{idj} \right]$, where $j = 1,..., M$. Let us consider population elements for which observations are available in the sample. Vector of random variables $Y_{i'd'j'}$ (where i'=1,...,n, j'=1,...,m, d'=1,...,D) of size $m \times 1$ is denoted by $\mathbf{Y_{s\,i'd'}} = \left[Y_{idj'} \right]$, where $j = 1,..., m$.

II. SUPERPOPULATION MODELS

We consider superpopulation models used for longitudinal data (e.g. Verbeke, Molenberghs (2000), Hedeker, Gibbons (2006)). The following two-stage model is assumed. Firstly:

$$\mathbf{Y_{id}} = \mathbf{Z_{id}}\boldsymbol{\beta}_{\mathbf{id}} + \mathbf{e_{id}} \tag{1}$$

where i=1,...,N; d=1,...,D, $\mathbf{Y_{id}}$ is random vector of size $M \times 1$, $\mathbf{Z_{id}}$ is known matrix of size $M \times q$, $\boldsymbol{\beta}_{\mathbf{id}}$ is vector of unknown parameters of size $q \times 1$, $\mathbf{e_{id}}$ is random component vector of size $M \times 1$. Vectors $\mathbf{e_{id}}$ (i=1,...,N; d=1,...,D) are independent with $\mathbf{0}$ vector of expected values and variance-covariance matrix $\mathbf{R_{id}}$. Although $\mathbf{R_{id}}$ may depend on i it is often assumed that $\mathbf{R_{id}} = \sigma_e^2 \mathbf{I_M}$, where $\mathbf{I_M}$ is identity matrix of rank M. Secondly, we assume that:

$$\boldsymbol{\beta}_{\mathbf{id}} = \mathbf{K_{id}}\boldsymbol{\beta} + \mathbf{v_{id}} \tag{2}$$

where i=1,...,N; d=1,...,D, $\mathbf{K_{id}}$ is known matrix of size $q \times p$, $\boldsymbol{\beta}$ is vector of unknown parameters of size $p \times 1$, $\mathbf{v_{id}}$ is vector of random components of size $q \times 1$. It is assumed, that vectors $\mathbf{v_{id}}$ (i=1,...,N; d=1,...,D) are independent with **0** vector of expected values and variance-covariance matrix, $\mathbf{G_{id}}$ = **H**, what means that $\mathbf{G_{id}}$ does not depend on i.

Similar assumption to (1) and (2) are presented by Verbeke, Molenberghs (2000) p. 20 but there are 2 differences. Firstly, in the book the domain subscripts are omitted but the sense of (1) and (2) is the same. Secondly, in the book assumptions are made only for sampled elements (i.e. i=1,...,n), in this paper they are made for all of population elements (i=1,...,N). Based on (1) and (2) it is obtained that:

$$\mathbf{Y_{id}} = \mathbf{X_{id}}\boldsymbol{\beta} + \mathbf{Z_{id}}\mathbf{v_{id}} + \mathbf{e_{id}} \tag{3}$$

where i=1,...,N; d=1,...,D, $\mathbf{X_{id}} = \mathbf{Z_{id}}\mathbf{K_{id}}$ is known matrix of size $M \times p$. Let $\mathbf{V_{id}} = D^2_{\xi}(\mathbf{Y_{id}})$. Hence, $\mathbf{V_{id}} = D^2_{\xi}(\mathbf{Y_{id}}) = \mathbf{Z_{id}}\mathbf{H}\mathbf{Z}^T_{\mathbf{id}} + \mathbf{R_{id}}$.

Let $\mathbf{A}_d$ be a column vector and $\mathbf{B}_d$ be a square matrix. Then, $col_{1\le d\le D}(\mathbf{A}_d) = [\mathbf{A}_1 \ \ ... \ \ \mathbf{A}_d \ \ ... \ \ \mathbf{A}_D]^T$ is a column vector obtained by stacking $\mathbf{A}_d$ vectors and $diag_{1\le d\le D}(\mathbf{B}_d) = \begin{bmatrix} \mathbf{B}_1 & ... & \mathbf{0} \\ ... & ... & ... \\ \mathbf{0} & ... & \mathbf{B}_D \end{bmatrix}$ is a block-diagonal matrix. Note that by stacking $\mathbf{Y_{id}}$ vectors (i.e. $\mathbf{Y} = col_{1\le d\le D}\left(col_{1\le i\le N_d}\left(\mathbf{Y_{id}}\right)\right)$ from (3) we obtain the formula of the GLMM. Let us introduce additional notations. Let the subscript s be used for elements observed in the sample and subscript r for elements unobserved in the sample. If the population elements are rearranged so that the first nm elements of **Y** and the first nm rows of **X** are for observations in the sample, we may use inter alia following notations: $\mathbf{Y} = \begin{bmatrix} \mathbf{Y_s} \\ \mathbf{Y_r} \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} \mathbf{X_s} \\ \mathbf{X_r} \end{bmatrix}$, $diag_{1\le d\le D}diag_{1\le i\le N_d}(\mathbf{V_{id}}) = \mathbf{V} = \begin{bmatrix} \mathbf{V_{ss}} & \mathbf{V_{sr}} \\ \mathbf{V_{rs}} & \mathbf{V_{rr}} \end{bmatrix}$.

Let us consider the following special case of (3):

$$Y_{idj} = (\beta_d + v_{id})j + e_{idj} = \beta_d j + v_{id} j + e_{idj} \tag{4}$$

where i=1,...,N; d=1,...,D, j=1,...,M. Similar model is considered by Verbeke and Molenberghs (2000) p. 25 but instead of domains they consider groups, they include random constant in the model and they make assumptions only for available observations (i=1,...,n).

In the considered model we assume (Verbeke and Molenberghs (2000)) that $\mathbf{R_{id}} = \sigma_e^2 \mathbf{I_M}$. What is more $\mathbf{H} = \sigma_v^2$. Hence,

$$Cov_\xi(Y_{idj}, Y_{i'd'j'}) = \begin{cases} 0 & if \quad i \neq i' \\ \sigma_e^2 + j^2\sigma_v^2 & if \quad i = i' \wedge j = j' \\ jj'\sigma_v^2 & if \quad i = i' \wedge j \neq j' \end{cases} \tag{5}$$

It is very important for further considerations to note that for balanced data under (5) we obtain $\mathbf{V_{rs}} = \mathbf{0}$.

All of the presented models – what is important for further considerations – are special cases of the GLM and the GLMM. Let us introduce assumptions of the GLM which will be studied in the next section. We assume that:

$$\begin{cases} E_\xi(\mathbf{Y}) = \mathbf{X\beta} \\ D_\xi^2(\mathbf{Y}) = \mathbf{V} \end{cases}, \tag{6}$$

where $\mathbf{X}$ is a $N \times p$ matrix of values of p auxiliary variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters and $\mathbf{V}$ is a variance-covariance matrix depending on some $\boldsymbol{\delta}$ vector of unknown in practice parameters .

III. BLUP, ITS MSE AND MSE ESTIMATOR

In the paper the BLUP proposed by Royall (1976) is studied but some comments on Henderson's (1950) BLUP will be presented. Under the GLMM Henderson studied the BLUP of the form $\hat{\theta}^s = \mathbf{a}^T \mathbf{Y_s} + b$ to predict $\theta^s = \mathbf{l}^T \boldsymbol{\beta} + \mathbf{m}^T \mathbf{v}$. He obtained the BLUP: $\hat{\theta}^s_{BLU} = \mathbf{l}^T \hat{\boldsymbol{\beta}} + \mathbf{m}^T \hat{\mathbf{v}}$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{v}}$ are some predictors of $\boldsymbol{\beta}$ and $\mathbf{v}$ (its precise formulae are not essential for the considerations). For prediction of the total in domain d* in the (past, current or future) period *t* we have $\mathbf{l}^T = \boldsymbol{\gamma}^T\mathbf{X}$ and $\mathbf{m}^T = \boldsymbol{\gamma}^T\mathbf{Z}$, where the element *k* of $\boldsymbol{\gamma}$ vector is given by $\gamma_k = \begin{cases} 1 & if \quad k \in \Omega_{d^*t} \\ 0 & if \quad k \notin \Omega_{d^*t} \end{cases}$.

For the considered model $\mathbf{v}$ is $Nq \times 1$, hence $\hat{\mathbf{v}}$ is also $Nq \times 1$ what means that it impossible to estimate $\mathbf{v}$ based only on observations from the sampled *n* elements. Hence, in practice Henderson's BLUP and EBLUP cannot be used under the discussed model.

Royall (1976) proposed the BLUP and presented its MSE for the GLM.

Theorem 1. (Royall (1976)). Assume that the population data obey the general linear model. Among the linear, model-unbiased predictors $\hat{\theta} = \mathbf{g}_\mathbf{s}^\mathbf{T}\mathbf{Y}_\mathbf{s}$ of linear combination of random variables $\theta = \boldsymbol{\gamma}^T\mathbf{Y}$ (where $\boldsymbol{\gamma} = \left[\boldsymbol{\gamma}_\mathbf{s}^\mathbf{T}, \boldsymbol{\gamma}_\mathbf{r}^\mathbf{T}\right]^T$) the MSE is minimized by:

$$\hat{\theta}_{BLU} = \boldsymbol{\gamma}_\mathbf{s}^\mathbf{T}\mathbf{Y}_\mathbf{s} + \boldsymbol{\gamma}_\mathbf{r}^\mathbf{T}\left[\mathbf{X}_\mathbf{r}\hat{\boldsymbol{\beta}} + \mathbf{V}_\mathbf{rs}\mathbf{V}_\mathbf{ss}^{-1}\left(\mathbf{Y}_\mathbf{s} - \mathbf{X}_\mathbf{s}\hat{\boldsymbol{\beta}}\right)\right], \tag{7}$$

where $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}_\mathbf{s}^\mathbf{T}\mathbf{V}_\mathbf{ss}^{-1}\mathbf{X}_\mathbf{s}\right)^{-1}\mathbf{X}_\mathbf{s}^\mathbf{T}\mathbf{V}_\mathbf{ss}^{-1}\mathbf{Y}_\mathbf{s}$

The MSE of $\hat{\theta}_{BLU}$ is given by

$$MSE_\xi(\hat{\theta}_{BLU}) = Var_\xi(\hat{\theta}_{BLU} - \theta) = g_1(\boldsymbol{\delta}) + g_2(\boldsymbol{\delta}), \tag{8}$$

where

$$g_1(\boldsymbol{\delta}) = \boldsymbol{\gamma}_\mathbf{r}^\mathbf{T}\left(\mathbf{V}_\mathbf{rr} - \mathbf{V}_\mathbf{rs}\mathbf{V}_\mathbf{ss}^{-1}\mathbf{V}_\mathbf{sr}\right)\boldsymbol{\gamma}_\mathbf{r}, \tag{9}$$

$$g_2(\boldsymbol{\delta}) = \boldsymbol{\gamma}_\mathbf{r}^\mathbf{T}\left(\mathbf{X}_\mathbf{r} - \mathbf{V}_\mathbf{rs}\mathbf{V}_\mathbf{ss}^{-1}\mathbf{X}_\mathbf{s}\right)\left(\mathbf{X}_\mathbf{s}^\mathbf{T}\mathbf{V}_\mathbf{ss}^{-1}\mathbf{X}_\mathbf{s}\right)^{-1}\left(\mathbf{X}_\mathbf{r} - \mathbf{V}_\mathbf{rs}\mathbf{V}_\mathbf{ss}^{-1}\mathbf{X}_\mathbf{s}\right)^\mathbf{T}\boldsymbol{\gamma}_\mathbf{r} \tag{10}$$

In the paper we derive the formula of the BLUP and its MSE for the model with assumption (4) and (5) based on the Royall's results. In the unbalanced case (which is not studied in this paper), the formula of the BLUP will depend on some unknown in practice variance parameters. Replacing them by the appropriate estimators gives the formula of the EBLUP which remains unbiased under some assumptions (Żądło (2004)). Then based on results presented by Żądło (2007) for the GLMM with block-diagonal variance-covariance matrix the formulae of MSE of the EBLUP and estimators of MSE may be derived assuming the considered model with assumptions (4) and (5).

In the considered case (i.e. for balanced data) where $\mathbf{V_{rs}} = \mathbf{0}$, the BLUP will not depend on unknown parameters. Let $s = \sum_{j=1}^{m} j^2 = \frac{m(m+1)(2m+1)}{6}$. Hence the BLUP and its MSE are as follows:

$$\hat{\theta}_{BLU} = \sum_{i \in s_{d^*t}} Y_i + tN_{rd^*t}\hat{\beta}_{d^*} \tag{11}$$

where $\hat{\beta}_d = (n_d s)^{-1} \sum_{i=1}^{n_d} \sum_{j=1}^{m} Y_{idj} j$

$$MSE(\hat{\theta}_{BLU}) = Var_{\xi}(\hat{\theta}_{BLU} - \theta) = g_1(\boldsymbol{\delta}) + g_2(\boldsymbol{\delta}), \tag{12}$$

where

$$g_1(\boldsymbol{\delta}) = N_{rd^*t}(\sigma_e^2 + t^2\sigma_v^2) \tag{13}$$

$$g_2(\boldsymbol{\delta}) = \left(\sigma_e^2 + \sigma_v^2 s\right) t^2 N_{rd^*t}^2 n_d^{-1} s^{-1} \tag{14}$$

To estimate MSE (12) unknown parameters should be replaced by some estimates. In this paper we consider widely used in practice: biased maximum likelihood (ML) estimators under normality assumption, and approximately unbiased restricted maximum likelihood (REML) estimators under normality assumption. In this case (under some more general assumptions see Żądło (2007)):

$$E_{\xi}(g_1(\hat{\boldsymbol{\delta}})) = g_1(\boldsymbol{\delta}) + \mathbf{B}_{\hat{\boldsymbol{\delta}}}^T(\boldsymbol{\delta}) \frac{\partial g_1(\boldsymbol{\delta})}{\delta \boldsymbol{\delta}} + o(D^{-1}) \tag{15}$$

$$E_{\xi}(g_2(\hat{\boldsymbol{\delta}})) = g_2(\boldsymbol{\delta}) + o(D^{-1}) \tag{16}$$

$$E_{\xi}\left(\mathbf{B}_{\hat{\boldsymbol{\delta}}}^T(\hat{\boldsymbol{\delta}}) \frac{\partial g_1(\hat{\boldsymbol{\delta}})}{\delta \boldsymbol{\delta}} \right) = \mathbf{B}_{\hat{\boldsymbol{\delta}}}^T(\boldsymbol{\delta}) \frac{\partial g_1(\boldsymbol{\delta})}{\delta \boldsymbol{\delta}} + o(D^{-1}) \tag{17}$$

where $\mathbf{B}_{\hat{\boldsymbol{\delta}}}(\boldsymbol{\delta})$ is vector of biases of $\hat{\boldsymbol{\delta}}$.

Hence (based on the results presented by Żądło (2007)) we obtain the following, approximately unbiased MSE estimators. For REML variance estimators:

$$M\hat{S}E_{\xi}(\hat{\theta}_{BLU}(\hat{\boldsymbol{\delta}})) = g_1(\hat{\boldsymbol{\delta}}) + g_2(\hat{\boldsymbol{\delta}}) \tag{18}$$

and for ML variance estimators:

$$M\hat{S}E_{\xi}(\hat{\theta}_{BLU}(\hat{\boldsymbol{\delta}})) = g_1(\hat{\boldsymbol{\delta}}) + g_2(\hat{\boldsymbol{\delta}}) - \mathbf{B}_{\hat{\boldsymbol{\delta}}}^T(\hat{\boldsymbol{\delta}})\frac{\partial g_1(\hat{\boldsymbol{\delta}})}{\delta \boldsymbol{\delta}} \tag{19}$$

where $g_1(\hat{\boldsymbol{\delta}})$ and $g_2(\hat{\boldsymbol{\delta}})$ are given by (13) and (14) respectively where $\boldsymbol{\delta}$ is replaced by $\hat{\boldsymbol{\delta}}$; $-\mathbf{B}_{\hat{\boldsymbol{\delta}}}^T(\hat{\boldsymbol{\delta}})\frac{\partial g_1(\hat{\boldsymbol{\delta}})}{\delta \boldsymbol{\delta}}$ is given by

$$-\mathbf{B}_{\hat{\boldsymbol{\delta}}}^T(\boldsymbol{\delta})\frac{\partial g_1(\boldsymbol{\delta})}{\delta \boldsymbol{\delta}} = \frac{1}{2} D N_{rd*t}\left(I_{vv}^{(-1)} + I_{ve}^{(-1)}(s+t^2) + st^2 I_{ee}^{(-1)}\right),$$

where $I_{vv}^{(-1)} = 2b_{\#}^{-1}a^{-2}ns^2$, $I_{ve}^{(-1)} = -2b_{\#}^{-1}a^{-2}ns$, $I_{ee}^{(-1)} = 2b_{\#}^{-1}n\left((m-1)\sigma_e^{-4} + a^{-2}\right)$, $b_{\#} = n^2 s^2 a^{-2}(m-1)\sigma_e^{-4}$ and $\boldsymbol{\delta}$ is replaced by $\hat{\boldsymbol{\delta}}$.

IV. SIMULATION STUDY

In the Monte Carlo simulation study 50 000 iterations were made using R language (R Development Core Team (2008)) based on the following (arbitrarily chosen) values of parameters: $\beta = 10$, $\sigma_e^2 = 1$, $\sigma_v^2 = 3$. In the simulation we study special case of the considered model (assuming $\forall_j j = 1$, $\forall_d \beta_d = \beta$) and hence special cases of the presented equations:

$$\hat{\theta}_{BLU} = \sum_{i \in s_{d*t}} Y_i + N_{rd*t}\hat{\mu} \tag{20}$$

and

$$\hat{\mu} = n^{-1}m^{-1}\sum_{d=1}^{D}\sum_{i=1}^{n_d}\sum_{j=1}^{m} Y_{idj} , \quad g_1(\boldsymbol{\delta}) = N_{rd*t}(\sigma_e^2 + \sigma_v^2) ,$$

$$g_2(\boldsymbol{\delta}) = \left(\sigma_e^2 + \sigma_v^2 m\right) N_{rd*t}^2 m^{-1} n^{-1},$$

$$-\mathbf{B}_{\hat{\boldsymbol{\delta}}}^T(\boldsymbol{\delta}) \frac{\partial g_1(\boldsymbol{\delta})}{\delta \boldsymbol{\delta}} = \frac{1}{2} a^{-1} N_{rd*t} \left(I_{vv}^{(-1)} + I_{ve}^{(-1)}(m+1) + m I_{ee}^{(-1)} \right),$$

where $a = \sigma_e^2 + \sigma_v^2 m$, $I_{vv}^{(-1)} = 2b_{\#}^{-1} a^{-2} n m^2$, $I_{ve}^{(-1)} = -2b_{\#}^{-1} a^{-2} nm$,

$$I_{ee}^{(-1)} = 2b_{\#}^{-1} n\left((m-1)\sigma_e^{-4} + a^{-2}\right),\ b_{\#} = n^2 m^2 a^{-2}(m-1)\sigma_e^{-4}.$$

We consider 9 cases of different distributions of random components (with values of variances presented above):

NN – normal distribution of v_{id}, normal distribution of e_{idj}

NU – normal distribution of v_{id}, uniform distribution of e_{idj}

NE – normal distribution of v_{id}, shifted exponential distribution of e_{idj}

UN – uniform distribution of v_{id}, normal distribution of e_{idj}

EN – shifted exponential distribution of v_{id}, normal distribution of e_{idj}

UU – uniform distribution of v_{id}, uniform distribution of e_{idj}

EE – shifted exponential distribution of v_{id}, shifted exponential distribution of e_{idj}

UE – uniform distribution of v_{id}, shifted exponential distribution of e_{idj}

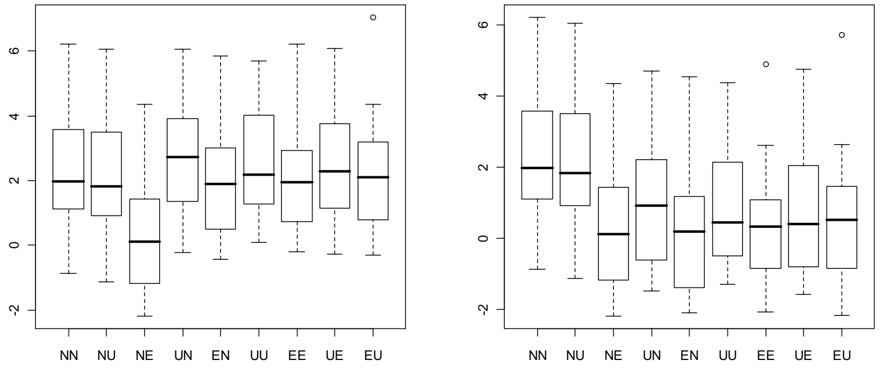
EU – shifted exponential distribution of v_{id}, uniform distribution of e_{idj}

In the simulation we study the BLUP predictor (20) with MSE estimated using ML variance estimators (given by the general equation (19)) and using REML variance estimators (given by the general equation (18)). We study population of size 2500 divided into 20 domains, balanced sample in 9 periods and we predict domain totals in period 10 (see table 1).

The considered predictors are model unbiased what does not depend on distribution of random components. Relative RMSEs are from 3,127% to 3,972% in 20 domains (for all of considered distributions of random components). The relative biases of approximately unbiased MSE estimators are not large – from about –2% to 6% for all of domains and different distributions of random components. The relative biases of MSE estimators are presented on the graph 1.

Table 1. Description of the population, domains and samples

d	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**sum**
N_d	**50**	**50**	**50**	**50**	**50**	**100**	**100**	**100**	**100**	**100**	**150**	**150**	**150**	**150**	**150**	**200**	**200**	**200**	**200**	**200**	**2500**
n_{d1}	0	3	2	0	1	3	2	3	2	2	3	5	0	1	1	2	4	2	2	2	**40**
n_{d2}	0	3	2	0	1	3	2	3	2	2	3	5	0	1	1	2	4	2	2	2	**40**
n_{d3}	0	3	2	0	1	3	2	3	2	2	3	5	0	1	1	2	4	2	2	2	**40**
n_{d4}	0	3	2	0	1	3	2	3	2	2	3	5	0	1	1	2	4	2	2	2	**40**
n_{d5}	0	3	2	0	1	3	2	3	2	2	3	5	0	1	1	2	4	2	2	2	**40**
n_{d6}	0	3	2	0	1	3	2	3	2	2	3	5	0	1	1	2	4	2	2	2	**40**
n_{d7}	0	3	2	0	1	3	2	3	2	2	3	5	0	1	1	2	4	2	2	2	**40**
n_{d8}	0	3	2	0	1	3	2	3	2	2	3	5	0	1	1	2	4	2	2	2	**40**
n_{d9}	0	3	2	0	1	3	2	3	2	2	3	5	0	1	1	2	4	2	2	2	**40**
sum	0	27	18	0	9	27	18	27	18	18	27	45	0	9	9	18	36	18	18	18	**360**



Graph 1. Biases of MSE estimators for REML (on the left) and ML (on the right) method of estimation for 20 domains

V. CONCLUSION

In the paper the BLUP of domain total in the case of some longitudinal data has been proposed with approximately unbiased MSE estimator. It has been shown in the simulation study that the predictor and MSE estimator may be used even in some cases of model misspecification.

BIBLIOGRAPHY

Datta G. S. and Lahiri P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, 10, 613-627.

Hedeker D., Gibbons R.D. (2006), *Longitudinal Data Analysis*, John Wiley & Sons, Hoboken, New Jersey.

Henderson C.R. (1950). Estimation of genetic parameters (Abstract). *Annals of Mathematical Statistics*, 21, 309-310.

Royall R.M. (1976), The linear least squares prediction approach to two-stage sampling. *Journal of the American Statistical Association*, 71, 657-473.

R Development Core Team (2008), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Verbeke G., Molenberghs G. (2000), *Linear Mixed Models for Longitudinal Data*, Springer-Verlag, New York.

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

Żądło (2007). On MSE of EBLUP, *Statistical papers,* DOI 10.1007/s00362-007-0066-3, pozytywnie zrecenzowany, nieopublikowany, dostępny w Springer Online First.

Tomasz Żądło

O PREDYKCJI WARTOŚCI GLOBALNEJ W DOMENIE NA PODSTAWIE ZBALANSOWANYCH DANYCH PANELOWYCH

W pracy rozważa się predyktory wartości globalnej wykorzystujące do predykcji wartości globalnej w pewnym (obecnym, przeszłym a nawet przyszłym) okresie dane z innych okresów. Model nadpopulacji jest szczególnym przypadkiem ogólnego liniowego modelu mieszanego – jest to model z specyficznymi dla elementów populacji składnikami losowymi możliwy do zastosowania dla danych wielookresowych. Postać predyktora typu BLU wyprowadzono w oparciu o twierdzenie Royalla (1976) – szeroko stosowane w literaturze podejście Hendersona (1950) nie jest możliwe do zastosowania dla rozważanego modelu nadpopulacji. W rozważanym przypadku wyprowadzono postać MSE w oparciu o twierdzenie Royalla i zaproponowano jego estymatory wykorzystując rezultaty uzyskane przez Żądło (2007). W przypadku estymacji MSE rozważano dwie metody estymacji wariancji składników losowych – metodę największej wiarygodności oraz metodę największej wiarygodności z ograniczeniami. W analizie symulacyjnej uwzględniono problem dokładności predyktora oraz obciążeń estymatorów MSE dla różnych rozkładów składników losowych.