*STATISTICS IN TRANSITION new series, December 2019*
*Vol. 20, No. 4, pp. 113–134, DOI 10.21307/stattrans-2019-037*
*Submitted – 04.03.2019; Paper ready for publication – 25.10.2019*

# ESTIMATION OF INCOME CHARACTERISTICS FOR REGIONS IN POLAND USING SPATIO-TEMPORAL SMALL AREA MODELS

**Alina Jędrzejczak**[1,2], **Jan Kubacki**[2]

## ABSTRACT

The paper presents the comparison of estimation results for spatial and spatiotemporal small area models. The study was carried out for income-related variables drawn from the Polish Household Budget Survey and explanatory variables from the Polish Local Data Bank for the years 2003-2013. The properties of EBLUPs (Empirical Best Linear Unbiased Predictors) based on spatiotemporal models, which utilize spatial correlation between neighbouring areas as well as historical data, were compared and contrasted with EBLUPs based on spatial models obtained separately for each year and with EBLUPs based on the Rao-Yu model. The computations were performed using sae, sae2 and spdep packages for R-project environment. In the case of sae package, the eblupFH, eblupSFH and the eblupSTFH functions were used for point estimation along with the mseFH, mseSFH and the pbmseSTFH functions for the MSE estimation, whereas the eblupRY function was applied for the purposes of sae2 package. The precision of direct estimators was guaranteed by the adoption of the Balanced Repeated Replication method. The results of the analysis demonstrate that a visible reduction of the estimation error was achieved for the implemented spatiotemporal small-area models, especially when significant spatial and time autocorrelations were observed. These results are even more valuable than those achieved by the means of the Rao-Yu model. In the computations, three author-defined functions were adopted, which not only enabled the author to perform the extract of random effects for spatial, spatiotemporal and Rao-Yu models, but also made it possible to obtain their decomposition with respect to spatial and temporal parts, thus creating a novel solution. The comparison was carried out using choropleth maps for spatial effects and distributions of temporal random effects for the considered years.

**Key words:** small area estimation, spatio-temporal models, Rao-Yu model, EBLUP estimation.

[1] Institute of Statistics and Demography, Faculty of Economics and Sociology, University of Łódź. E-mail: alina.jedrzejczak@uni.lodz.pl. ORCID ID: https://orcid.org/0000-0002-5478-9284.
[2] Centre of Mathematical Statistics, Statistical Office in Łódź. E-mail: j.kubacki@stat.gov.pl. ORCID ID: https://orcid.org/0000-0001-8281-0514.

## 1. Introduction

Statistical surveys are often designed to provide data that allow reliable estimation for the whole country and larger administrative units such as regions (in Poland – voivodships). However, for more specific variables the overall sample size is seldom large enough to yield direct estimates of adequate precision for all the domains of interest. In such cases, large estimation errors can make the inference unreliable and useless for decision-makers. The estimation errors can be reduced, however, by means of the model-based approach. Moreover, when an evident correlation exists between survey and administrative data, also the bias of the estimates can be reduced.

Small area estimation (SAE) offers a wide range of methods that can be applied when a sample size is insufficient to obtain high precision by means of conventional direct estimates. The techniques based on small area models - empirical best linear unbiased prediction (EBLUP) as well as empirical and hierarchical Bayes (EB and HB), seem to have a distinct advantage over other methods. One of these techniques is the spatio-temporal EBLUP technique, introduced by Marhuenda, Molina and Morales (2013). It is based on the assumption that the spatial relationships between domains can be modelled by a sum of two components: the simultaneous autoregressive process SAR (see: Pratesi and Salvati (2008), p. 114) and an additional time-related process described by AR(1) scheme (see: Rao, Yu (1992, 1994)). Both these assumptions are involved in the covariance structure of the spatio-temporal model. Related Spatial EBLUP (Spatial Empirical Best Linear Unbiased Prediction), which was introduced by Cressie (1991) and is explained in detail in Saei and Chambers (2003), Pratesi and Salvati (2004, 2005, 2008), Singh et al. (2005), Petrucci and Salvati (2006), can be considered as a reference point for spatio-temporal models. Recently, the Spatial EBLUP technique was used in 'sae' package (Molina, Marhuenda (2013)) for R-project environment published in CRAN resources. Moreover, some spatial econometric models were discussed in Griffith, D.A., Paelinck, J.H.P. (2011) and Kubacki, Jędrzejczak (2016), where MCMC (Markov Chain Monte Carlo) applications for spatial models are presented.

In the paper we compare several approaches to the spatial and spatio-temporal modelling implemented for small area estimation. In our opinion, spatio-temporal estimation can be sometimes useful with respect to the traditional EBLUP approach. This is because of better efficiency of such models, which not only incorporate ordinary spatial relationships (using proximity matrix), but also assume time-related dependencies. It can be useful when visible time-related relationships are observed in a data set. These models, using sample and auxiliary information from other domains as well as other time periods, can yield substantial quality improvements as compared to ordinary small area models, where only explanatory variables from administrative sources or other statistical surveys are used. It is also related to imposing certain constraints that can positively affect the quality of obtained estimates. The models with time-related covariance structure can additionally be helpful in the analysis of the dynamics of the observed phenomena, which can be supplementary related to the econometric models, including the panel models (Jędrzejczak, Kubacki (2016)).

## 2. Estimation for small areas using spatio-temporal Fay-Herriot model

In the paper the primary target results were those related to the Spatio-Temporal Fay-Ferriot model (STFH), being the extension of the classical Fay-Herriot small area model. The methodology for such models was described in detail by Marhuenda, Molina and Morales (2013). Under the spatio-temporal small area model, the area parameter for domain *d* at current time instant *T* is estimated as borrowing strength from other time instants and from the *D* domains. Let $\theta_{dt}$ represent the variable of interest determined for area *d* and time *t* where $d = 1, \dots, D$, and $t = 1, \dots, T$. If the direct estimator of this quantity is denoted by $\hat{\theta}_{dt}^{DIR}$, and the sampling errors are expressed as $e_{dt}$, which are assumed to be independent and normally distributed with known variances $\psi_{dt}$, the spatio-temporal model can be written as below

$$\hat{\theta}_{dt}^{DIR} = \theta_{dt} + e_{dt} \qquad (1)$$

The relationship above is valid for all considered *d* and *t*. The equation (1) can also be expressed by means of the model which incorporates spatio-temporal relationships of the form

$$\theta_{dt} = \mathbf{x}_{dt}^{T}\beta + u_{1d} + u_{2dt} \qquad (2)$$

Here, $\mathbf{x}_{dt}$ are the vectors of *p* auxiliary variables representing regression structure of $\theta_{dt}$, with regression coefficients *β*. The area-time random effects can be expressed by $(u_{2d1}, \dots, u_{2dT})^T$ and are assumed identically and independently distributed for each area. Moreover, they follow the AR(1) process with autocorrelation parameter $\rho_2$, which can be described as

$$u_{2dt} = \rho_2 u_{2d,t-1} + \epsilon_{2dt}, \text{ where } |\rho_2| < 1 \text{ and } \epsilon_{2dt} \overset{iid}{\sim} N(0, \sigma_2^2) \qquad (3)$$

The area-related random effects, expressed by $(u_{11}, \dots, u_{1D})^T$, are subject to the SAR process with variance parameter $\sigma_1^2$, spatial autocorrelation $\rho_1$ and proximity matrix $\mathbf{W} = (w_{d,l})$, which can be obtained from an original proximity matrix $\mathbf{W}^0$ , whose diagonal elements are equal to zero and the remaining entries are equal to 1 (when the two areas corresponding to the row and the column indices are considered as neighbour and zero otherwise). Then, **W** is obtained by row-standardization of $\mathbf{W}^0$, obtained by dividing each entry of $\mathbf{W}^0$ by the sum of elements in the same row. The area level random effects can be described as

$$u_{id} = \rho_1 \sum_{l \neq d} w_{d,l} u_{1l} + \epsilon_{1d} \text{ where } |\rho_1| < 1 \text{ and } \epsilon_{1d} \overset{iid}{\sim} N(0, \sigma_1^2) \qquad (4)$$

Using the stacking notations for vectors and matrices one can present the following relationships for the considered model:

$$\mathbf{y} = \underset{1 \le d \le D}{col}\left(\underset{1 \le t \le T}{col}(\hat{\theta}_{dt}^{dir})\right), \quad \mathbf{X} = \underset{1 \le d \le D}{col}\left(\underset{1 \le t \le T}{col}(x_{dt}^{T})\right)$$

$$\mathbf{e} = \underset{1 \le d \le D}{col}\left(\underset{1 \le t \le T}{col}(e_{dt})\right), \mathbf{u}_1 = \underset{1 \le d \le D}{col}(u_{1d}), \mathbf{u}_2 = \underset{1 \le d \le D}{col}\left(\underset{1 \le t \le T}{col}(u_{2dt})\right)$$

Also, one can define additionally $\mathbf{Z}_1 = \mathbf{I}_D \otimes \mathbf{1}_T$ , where $\mathbf{I}_D$, is the D x D identity matrix, $\mathbf{1}_T$ is the vector of 1's and has length T, and ⊗ is the Kronecker product, $\mathbf{Z}_2 = \mathbf{I}_n$, where *n=DT* is the total number of observations, $\mathbf{u} = (\mathbf{u}_1^T, \mathbf{u}_2^T)^T$ and $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$.

Using the notation presented above we can describe the spatio-temporal model in terms of the general linear mixed model, which has the following form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

Let $\boldsymbol{\delta} = (\sigma_1^2, \rho_1, \sigma_2^2, \rho_2)$ denote the vector of covariance structure parameters involved in the model above. We can use the following relationships for the vector **e** related to direct estimation error: $\mathbf{e} \sim N(\mathbf{0}_n, \boldsymbol{\psi})$, where $\mathbf{0}_n$ denotes a vector of zeroes that has the length *n* and $\boldsymbol{\psi}$ is the diagonal matrix $\psi = diag_{1 \le d \le D}(diag_{1 \le t \le T}(\psi_{dt}))$.

The random effects **u** are also normally distributed $\mathbf{u} \sim N\{\mathbf{0}_n, \mathbf{G}(\boldsymbol{\delta})\}$, where the covariance matrix **G** can be expressed as the block diagonal matrix of the following form: $\mathbf{G}(\boldsymbol{\delta}) = diag\{\sigma_1^2 \boldsymbol{\Omega}_1(\rho_1), \sigma_2^2 \boldsymbol{\Omega}_2(\rho_2)\}$. The matrices $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ are defined as

$$\boldsymbol{\Omega}_1(\rho_1) = \{(\mathbf{I}_D - \rho_1 \mathbf{W})^T (\mathbf{I}_D - \rho_1 \mathbf{W})\}^{-1} \tag{5}$$

$$\boldsymbol{\Omega}_2(\rho_2) = diag_{1 \le d \le D}\{\boldsymbol{\Omega}_{2d}(\rho_2)\}$$

$$\boldsymbol{\Omega}_{2d}(\rho_2) = \frac{1}{1-\rho_2^2} \begin{pmatrix} 1 & \rho_2 & 0 \dots & \rho_2^{T-2} & \rho_2^{T-1} \\ \rho_2 & 1 & \ddots & 1 & \rho_2^{T-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho_2^{T-2} & & \ddots & 1 & \rho_2 \\ \rho_2^{T-1} & \rho_2^{T-2} & \dots & \rho_2 & 1 \end{pmatrix} \tag{6}$$

The covariance matrix for the full model (including the sampling error) can be expressed as

$$\mathbf{V}(\boldsymbol{\delta}) = \mathbf{Z}\mathbf{G}(\boldsymbol{\delta})\mathbf{Z}^T + \boldsymbol{\Psi}$$

The vector **β** and the random effects **u** can be obtained using BLUP estimator $\widetilde{\boldsymbol{\beta}}(\boldsymbol{\delta})$ by means of the following equations, utilizing **X**, **G**, **V** and **Z** matrices:

$$\widetilde{\boldsymbol{\beta}}(\boldsymbol{\delta}) = \{\mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\delta})\mathbf{X}\}^{-1} \mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\delta})\mathbf{y}$$

$$\widetilde{\mathbf{u}}(\boldsymbol{\delta}) = \mathbf{G}(\boldsymbol{\delta})\mathbf{Z}^T \mathbf{V}^{-1}(\boldsymbol{\delta})\{\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\delta})\}$$

Because $\mathbf{u} = (\mathbf{u}_1^T, \mathbf{u}_2^T)^T$, the second equation given above can be decomposed as follows

$$\widetilde{\boldsymbol{u}}_1(\boldsymbol{\delta}) = \sigma_1^2 \boldsymbol{\Omega}_1(\rho_1)\mathbf{Z}_1^T \mathbf{V}^{-1}(\boldsymbol{\delta})\{\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\delta})\} \tag{7}$$

$$\widetilde{\mathbf{u}}_2(\boldsymbol{\delta}) = \sigma_2^2 \boldsymbol{\Omega}_2(\rho_2)\mathbf{V}^{-1}(\boldsymbol{\delta})\{\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\delta})\}$$

## 3. REML estimation method for spatio-temporal model

The method of Restricted Maximum Likelihood (REML) uses maximization for the likelihood function, which corresponds to the joint probability density function as a vector of *n-p* linearly independent contrasts expressed as $\mathbf{F^T y}$ where **F** is the $n \times (n-p)$ full column rank satisfying the relationships $\mathbf{F}^T\mathbf{F} = \mathbf{I}_{n-p}$ and $\mathbf{F}^T\mathbf{X} = \mathbf{0}_{n-p}$. From the previous conditions, the probability density function of the contrast vectors can be expressed as

$$f_R(\boldsymbol{\delta};\mathbf{y}) = (2\pi)^{-(n-p)/2}|\mathbf{X}^T\mathbf{X}|^{1/2}|\mathbf{V}(\boldsymbol{\delta})|^{-1/2}|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{y}^T\mathbf{P}(\boldsymbol{\delta})\mathbf{y}\right\}$$

where **P** matrix satisfies the condition

$$\mathbf{P}(\boldsymbol{\delta}) = \mathbf{V}^{-1}(\boldsymbol{\delta}) - \mathbf{V}^{-1}(\boldsymbol{\delta})\mathbf{X}\{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\delta})\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\delta})$$

The matrix **P** satisfies the following relationships $\mathbf{P}(\boldsymbol{\delta})\mathbf{V}(\boldsymbol{\delta})\mathbf{P}(\boldsymbol{\delta}) = \mathbf{P}(\boldsymbol{\delta})$ and $\mathbf{P}(\boldsymbol{\delta})\mathbf{X} = \mathbf{0}_n$. The REML estimator maximizes the log likelihood function $\ell_R(\boldsymbol{\delta};\mathbf{y}) = \log f_R(\boldsymbol{\delta};\mathbf{y})$ using Fisher scoring algorithm. This algorithm utilizes scoring vectors of the form $S_R(\boldsymbol{\delta}) = \partial\ell_R(\boldsymbol{\delta};\mathbf{y})/\partial\boldsymbol{\delta}$ as well as the Fisher information matrix which is $\Im_R(\boldsymbol{\delta}) = -E\left\{\frac{\partial^2\ell_R(\delta;y)}{\partial\delta\partial\delta'}\right\} = (\Im_{rs}^R(\boldsymbol{\delta}))$.

For the spatio-temporal model with four variance components we have the following relationship

$$\frac{\partial\mathbf{P}(\boldsymbol{\delta})}{\partial\delta_r} = -\mathbf{P}(\boldsymbol{\delta})\frac{\partial\mathbf{V}(\boldsymbol{\delta})}{\partial\delta_r}\mathbf{P}(\boldsymbol{\delta})$$

for *r=1,…,4*. The first order derivative of $\ell_R(\boldsymbol{\delta};\mathbf{y})$, with respect to $\delta_r$ can be given as below

$$S_r^R(\boldsymbol{\delta}) = -\frac{1}{2}tr\left\{\mathbf{P}(\boldsymbol{\delta})\frac{\partial\mathbf{V}(\boldsymbol{\delta})}{\partial\delta_r}\right\} + \frac{1}{2}\mathbf{y}^T\mathbf{P}(\boldsymbol{\delta})\frac{\partial\mathbf{V}(\boldsymbol{\delta})}{\partial\delta_r}\mathbf{P}(\boldsymbol{\delta})\mathbf{y}$$

so the element indexed by (*r*,*s*) in the Fisher information matrix can be expressed as

$$\Im_{rs}^R(\boldsymbol{\delta}) = \frac{1}{2}tr\left\{\mathbf{P}(\boldsymbol{\delta})\frac{\partial\mathbf{V}(\boldsymbol{\delta})}{\partial\delta_r}\mathbf{P}(\boldsymbol{\delta})\frac{\partial\mathbf{V}(\boldsymbol{\delta})}{\partial\delta_s}\right\}$$

The detailed expressions for the partial derivatives of **V** with respect to the variance components used in the expression for scoring vectors and the Fisher information matrix have the following form:

$$\frac{\partial V(\boldsymbol{\delta})}{\partial\sigma_1^2} = \mathbf{Z}_1\Omega_1(\rho_1)\mathbf{Z}_1^T, \qquad \frac{\partial\mathbf{V}(\boldsymbol{\delta})}{\partial\rho_1} = -\sigma_1^2\mathbf{Z}_1\boldsymbol{\Omega}_1(\rho_1)\frac{\partial\boldsymbol{\Omega}_1^{-1}(\rho_1)}{\partial\rho_1}\boldsymbol{\Omega}_1(\rho_1)\mathbf{Z}_1^T$$

$$\frac{\partial V(\boldsymbol{\delta})}{\partial\sigma_2^2} = \underset{1\le d\le D}{diag}\{\boldsymbol{\Omega}_{2d}(\rho_2)\} \qquad \frac{\partial V(\boldsymbol{\delta})}{\delta\rho_2} = \sigma_2^2\underset{1\le d\le D}{diag}\{\frac{\partial\boldsymbol{\Omega}_{2d}(\rho_2)}{\partial\rho_2}\}$$

where

$$\frac{\partial \mathbf{\Omega}_1^{-1}(\rho_1)}{\partial \rho_1} = -\mathbf{W} - \mathbf{W}^T + 2\rho_1 \mathbf{W}^T \mathbf{W},$$

$$\frac{\partial \mathbf{\Omega}_{2d}(\rho_2)}{\partial \rho_2} = \frac{1}{1-\rho_2^2}\begin{pmatrix} 0 & 1 & \cdots & \cdots & (T-1)\rho_2^{T-2} \\ 1 & 0 & \ddots & & (T-1)\rho_2^{T-3} \\ \vdots & \ddots & \ddots & & \vdots \\ (T-2)\rho_2^{T-3} & & \ddots & 0 & 1 \\ (T-1)\rho_2^{T-2} & \dots & \dots & 1 & 0 \end{pmatrix} + \frac{2\rho_2 \mathbf{\Omega}_{2d}(\rho_2)}{1-\rho_2^2}$$

Finally, the scoring algorithm assumes that the variance component vector converges to the common value, using the following iterative procedure

$$\boldsymbol{\delta}^{(k+1)} = \boldsymbol{\delta}^{(k)} + \mathfrak{I}_{rs}^{R}(\boldsymbol{\delta}^{(k)}) S_R(\boldsymbol{\delta}^{(k)})$$

## 4. Determining the MSE of spatio-temporal estimates using parametric bootstrap method.

The estimation of MSE of spatio-temporal estimator was determined using the parametric bootstrap method implemented in sae package. This method can be summarized as follows:

1. Using direct income estimates and available auxiliary data $\{(\hat{\theta}_{dt}^{DIR}, x_{dt}),$ *t=1,..,T, d=1,…,D*}, obtain the estimates of the STFH model described by the equations (1) - (4) together with the estimates of the model parameter **β** and **δ.**
2. Generate bootstrap area effects $\{u_{1d}^{*(b)}$, *d=1,…,D* } from the SAR(1) process given in (4), assuming $(\hat{\sigma}_1^2, \hat{\rho}_1)$ as true values of $(\sigma_1^2, \rho_1)$
3. Independently of $\{u_{1d}^{*(b)}\}$ and independently for each d, generate bootstrap time effects $\{u_{2dt}^{*(b)}$, *t=1,…,T* } from the AR(1) process given in (3), with $(\hat{\sigma}_2^2, \hat{\rho}_2)$ acting as true values of the parameters $(\sigma_2^2, \rho_2)$
4. Calculate true bootstrap quantities, using the formula
$$\theta_{dt}^{*(b)} = \mathbf{x}_{dt}^T \beta + u_{1d}^{*(b)} + u_{2dt}^{*(b)}$$
5. Generate errors $e_{dt}^{*(b)} \overset{ind}{\sim} N(0, \psi_{dt})$ and obtain bootstrap data from the sampling model
$$\hat{\theta}_{dt}^{DIR*(b)} = \theta_{dt}^{*(b)} + e_{dt}^{*(b)}$$
6. Using the new bootstrap data $\{(\hat{\theta}_{dt}^{DIR*(b)}, x_{dt}),$ *t=1,..,T, d=1,…,D*} determine the estimates of STFH model described with equations from (1) to (4) and obtain the bootstrap EBLUPs
$$\hat{\theta}_{dt}^{*(b)} = \mathbf{x}_{dt}^T \hat{\beta}^{*(b)} + \hat{u}_{1d}^{*(b)} + \hat{u}_{2dt}^{*(b)}$$
7. Repeat steps (1)-(6) for *b = 1, … ,B* , where B is a large number.

8. The parametric bootstrap MSE estimates are given by

$$mse\left(\hat{\theta}_{dt}\right) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}_{dt}^{*(b)} - \theta_{dt}^{*(b)}\right)^2$$

## 5. Results and discussion

In the application, we were interested in the estimation of various per capita income components (in particular *income from social-security benefits*) by region NUTS2, based on the micro data coming from the Polish Household Budget Survey (HBS) and regional data obtained from the GUS Local Data Bank. Spatial and spatio-temporal models can fit to this kind of situations as they account for the correlation related to neighbourhood of areas and time-dependency, which both determine the random effects. They are based on cross-sectional and time-series data involving spatial autocorrelation. The model-based approach generally improves the estimation quality due to the use of explanatory variables coming from administrative registers and area random effects, which additionally account for the variability between domains. In current approach we can have extra gains coming from spatial and time dependencies between areas. To obtain better estimates for the year 2013, we decided to utilize historical data coming from ten years before, which enabled "borrowing strength" not only across areas but also over time and space. The results obtained on the basis of these models were compared to the ones obtained from the classical Fay-Herriot model and to direct estimates.

At the first stage, direct estimates of both parameters of interest for 16 regions were calculated from the HBS sample together with their standard errors obtained by means of the Balanced Repeated Replication (BRR) technique (see Westat (2007) for details). In the computations conducted in R-project environment, the packages sae, sae2 and spdep were applied.

At the second stage, the appropriate models were formulated and estimated from the data, and finally, EBLUP estimates were obtained as well as their MSE estimates. In order to evaluate the possible advantages of the estimators obtained by means of **Spatio-Temporal model (STFH)** we also estimated the parameters of simpler small area models. In particular, we additionally estimated the parameters of the following small-area models:

- **Rao-Yu model (RY)**, ("borrowing strength" from areas and over time),
- **Spatial Fay-Herriot model (SFH),** ("borrowing strength" only from other areas with proximity matrices, separately for each of them),
- **Fay-Herriot model (FH)**, ("borrowing strength" only from other areas),

and additionally, for comparison purposes, we estimated the unknown parameters using:

- classical spatial econometrics models based on SAR process, including **spatial lag model (lagsar)** and **spatial error model (errorsar)**.

At the third stage, using the model parameters which were estimated at the second stage, we obtained the predictors of per capita income for regions in Poland. In particular, for the STFH, FH, SFH, RY models, the appropriate EBLUPs were obtained, while for the spatial econometric models the appropriate linear predictors were evaluated.

The sae package made it possible to obtain estimation for spatial related model and spatio-temporal model. The sae2 package includes the implementation of the estimation procedure for the Rao-Yu model (see Fay R.E., Li J. (2012), Li J., Diallo M.S., Fay R.E. (2012), Fay, R. E., Diallo, M., (2015)), which provides an extension of the basic type A model to handle time series and cross-sectional data (Rao, Molina (2015)). The spdep package (Bivand, R., Lewin-Koh, N., (2013), Bivand, R., Piras, G., (2015)) was used for estimation of classical spatial econometrics models and also Moran I statistics for the considered variables. Our own R macro was developed, performing calculations for ordinary EBLUP models, spatial EBLUP models, Rao-Yu models, classical econometric models and spatio-temporal models, model diagnostics, as well as the maps for regions.

**Table 1.** Diagnostics of Rao-Yu and Spatio-Temporal models of income per capita from *social security benefits* based on sample and administrative data

| Model/ explanatory variable | Coefficient estimates | Standard error | t-statistics | p-value |
|---|---|---|---|---|
| 1. Rao-Yu model | $\sigma_2^2$ = 111.410 | $\sigma_1^2$ =124.710 | $\rho_2$=0.540 | |
| Intercept | 51.1520 | 16.7460 | 3.0546 | 0.0023 |
| Average salary in nat.economy | 0.0168 | 0.0123 | 1.3680 | 0.1713 |
| Average retirements pay | 0.1424 | 0.0236 | 6.0462 | 1.483E-09 |
| GDP per capita (Poland 100%) | -0.3314 | 0.1833 | -1.8081 | 0.0706 |
| 2. Spatio-temporal model | $\sigma_2^2$ = 110.640 | $\sigma_1^2$ =88.691 | $\rho_1$ =0.856 | $\rho_2$=0.501 |
| Intercept | 64.8670 | 21.3040 | 3.0448 | 0.0023 |
| Average salary in nat.economy | 0.0261 | 0.0124 | 2.1052 | 0.0353 |
| Average retirements pay | 0.1245 | 0.0237 | 5.2651 | 1.402E-07 |
| GDP per capita (Poland 100%) | -0.4945 | 0.1653 | -2.9913 | 0.0028 |

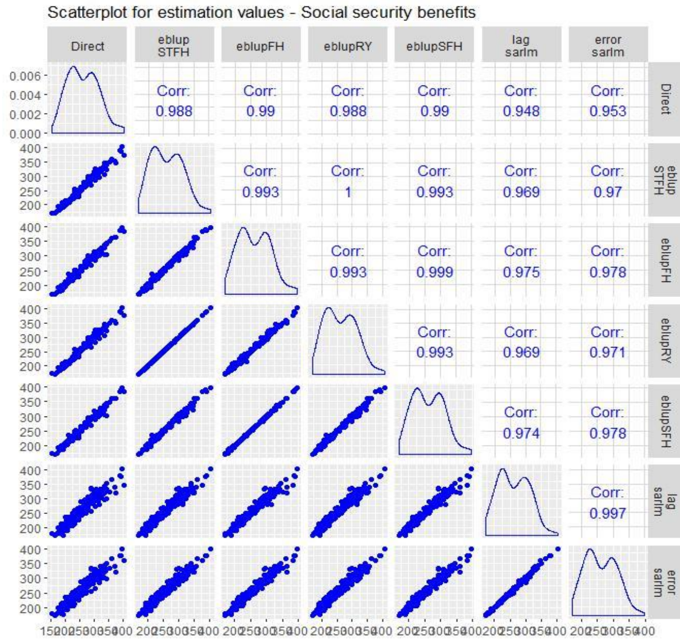
*Source: Authors' calculations based on the Polish Household Budget Survey and data from the Local Data Bank.*

**Table 2.** Estimation results for per capita income from *social security benefits* by region in Poland for the year 2013

| Region | Direct estimate | | 1. Rao-Yu estimate | | | 2. Spatio-temporal estimate | | |
|---|---|---|---|---|---|---|---|---|
| | Value | REE | Value | REE | Time-related random effect | Value | REE | Time-related random effect |
| | zł. | % | zł. | % | zł. | zł. | % | zł. |
| Dolnośląskie | 401.58 | 4.28 | 375.83 | 2.75 | 16.936 | 375.66 | 2.55 | 15.951 |
| Kujaw. Pomor. | 329.55 | 3.69 | 327.55 | 2.64 | 9.699 | 326.43 | 2.34 | 10.640 |
| Lubelskie | 282.68 | 3.50 | 298.15 | 2.65 | -20.880 | 298.20 | 2.78 | -20.569 |
| Lubuskie | 333.49 | 6.70 | 346.87 | 3.24 | 11.211 | 346.04 | 3.17 | 9.664 |
| Łódzkie | 354.50 | 2.97 | 350.90 | 2.31 | 9.263 | 350.10 | 2.14 | 9.794 |
| Małopolskie | 305.97 | 3.74 | 320.03 | 2.63 | -13.364 | 319.93 | 2.43 | -13.158 |
| Mazowieckie | 335.47 | 3.23 | 339.16 | 2.41 | -7.202 | 338.44 | 2.44 | -5.301 |
| Opolskie | 342.17 | 5.89 | 343.94 | 3.30 | -4.882 | 345.93 | 2.82 | -6.008 |
| Podkarpackie | 292.47 | 3.41 | 299.82 | 2.57 | -9.169 | 299.47 | 2.63 | -8.394 |
| Podlaskie | 325.53 | 5.03 | 323.70 | 3.00 | 3.633 | 322.50 | 2.95 | 4.609 |
| Pomorskie | 320.82 | 5.62 | 326.00 | 3.25 | -5.475 | 325.55 | 3.43 | -4.576 |
| Śląskie | 395.95 | 2.86 | 404.36 | 2.16 | -7.007 | 403.70 | 2.30 | -6.054 |
| Świętokrzyskie | 329.61 | 4.49 | 329.70 | 2.89 | 4.471 | 329.32 | 3.07 | 4.145 |
| Warm.-Mazur. | 297.09 | 5.83 | 306.51 | 3.45 | -5.081 | 304.95 | 3.88 | -3.562 |
| Wielkopolskie | 290.97 | 2.48 | 297.73 | 2.10 | -21.435 | 298.20 | 2.23 | -22.928 |
| Zach.-pomor. | 333.86 | 7.52 | 335.91 | 3.39 | -5.610 | 336.01 | 3.49 | -5.285 |

*Source: Authors' calculations based on the Polish Household Budget Survey and data from the Local Data Bank.*

**Figure 1.** Pair scatterplots of direct and model-based estimates for per capita income from *social security benefits* by region in Poland for the years 2003–2013




*Source: Authors' calculations based on the Polish HBS and Local Data Bank.*

**Table 3.** Variance structure parameters of selected small area and econometric models together with their log likelihoods for per capita income from *social security benefits* by region in Poland for the years 2003–2013

| Year | Fay-Herriot model | | Spatial Fay-Herriot model | | | Spatial SAR error model | | |
|---|---|---|---|---|---|---|---|---|
| | $\sigma_1^2$ | Log likeli-hood | $\sigma_1^2$ | $\rho$ | Log likeli-hood | $\sigma^2$ | λ | Log likeli-hood |
| 2003 | 93.82 | -61.95 | 43.34 | 0.961 | -62.56 | 92.38 | 0.572 | -59.69 |
| 2004 | 223.00 | -66.34 | 146.24 | 0.829 | -66.02 | 184.78 | 0.523 | -65.09 |
| 2005 | 118.48 | -62.76 | 116.01 | 0.371 | -62.93 | 153.29 | 0.058 | -62.97 |
| 2006 | 38.29 | -59.62 | 42.75 | 0.342 | -59.88 | 117.54 | 0.309 | -61.04 |
| 2007 | 275.76 | -68.37 | 245.69 | 0.625 | -68.24 | 215.60 | 0.600 | -66.56 |
| 2008 | 261.88 | -69.35 | 205.51 | 0.826 | -69.98 | 355.35 | 0.401 | -70.03 |
| 2009 | 318.67 | -69.02 | 273.93 | -0.411 | -68.40 | 215.58 | -0.953 | -67.39 |
| 2010 | 202.88 | -68.20 | 196.06 | -0.130 | -68.07 | 247.33 | -0.646 | -67.54 |
| 2011 | 133.95 | -66.56 | 113.34 | 0.702 | -66.78 | 302.40 | 0.302 | -68.59 |
| 2012 | 558.55 | -73.16 | 551.95 | -0.047 | -73.09 | 497.60 | -0.341 | -72.59 |
| 2013 | 615.95 | -74.11 | 421.61 | -0.693 | -72.86 | 369.39 | -0.903 | -71.51 |

*Source: Authors' calculations based on the Polish HBS and Local Data Bank.*

In Table 1 we show estimation results obtained for the Rao-Yu and spatio-temporal STFH model (eq.:(1)-(4)). For each dependent variable the estimates of fixed effects **β** and the parameters of variance-covariance structure of the models denoted as **δ** are presented. The model diagnostics indicate that both parameters related to the variability of random effects (σ) have visible contribution to the variability of the model. This is in contrast with previously presented models (see Jędrzejczak, Kubacki (2016)), where for the model of available income this variability was mostly determined by time-related component. It should be noted that a similar comparison of small area-models of available income also revealed such relationships for the Rao-Yu and spatio-temporal models, which may mean that both these approaches are complementary. Figures 5 and 6 additionally show decompositions of time-related random effects of the Rao-Yu and STFH models and make it possible to observe the impact and the distribution of these effects over time. Figure 1 summarizes dependencies between all pairs of estimates obtained in the study over the analysed period by means of the Pearson correlation coefficients and scatterplots.

In Table 2 there are income estimates, their corresponding relative estimation errors (REE) and time-related random effects $u_2$. Covariance structure parameters together with log likelihoods for all the years and selected models are given in Table 3. In the table, for comparison purposes, also the spatial SAR error model was provided due to identical assumptions about spatial random effects.

Precision of different estimation methods can be analysed on the basis of the detailed results given in Tables 4-6 and in Figures 2-4. The tables show Consistency Coefficients (CC), Relative Estimation Errors (REE) and REE reduction, respectively. The consistency coefficients presented in Table 4 and Figure 2 can be defined as follows

$$CC_{dt} = (\theta_{dt}^{model} - \hat{\theta}_{dt}^{DIR})/\hat{\theta}_{dt}^{DIR}$$

This measure can be used as a simple approximation of bias of model-based estimates. The results given in Table 4 and Figure 2 indicate that simpler estimation techniques may be less biased than the more complicated ones (Rao-Yu and STFH models). It seems that introducing more assumptions about the random-effects not always reflects the real-world relationships. Special attention should be paid to the values of CC obtained for econometric spatial error (SAR) model, which confirm that the classical spatial econometric models may be insufficient for small area estimation (Figure 2).

In Table 5, REE values for different estimation techniques are summarized, while in Table 6 REE reduction is presented with respect to both: direct and ordinary EBLUP estimates, corresponding to the first and the second column for each model. This approach can be helpful to recognise efficiency gains coming from model-based estimation and additional gains coming from temporal (and spatial) correlation incorporated in more advanced small area models.

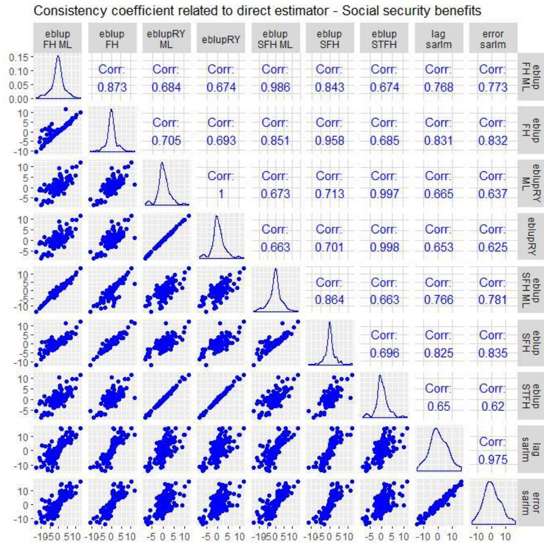
Comparisons of the distributions of REE and REE reduction (Figure 3 and Figure 4) show that all the considered model-based techniques are significantly more efficient than the corresponding direct ones. The Rao-Yu model and spatio-temporal model STFH perform similarly, as compared to the other model-based techniques which have been considered in the study. This regularity can also be observed when a comparison between a spatial, spatio-temporal and Rao-Yu models in terms of REE Reduction related to the ordinary EBLUPs is made.

**Table 4.** Consistency Coefficients [in %] for model-based estimates related to direct estimates for per capita income from *social security benefits* by region in Poland for the year 2013

| Region | Consistency coefficient [in %] | | | |
|---|---|---|---|---|
| | Fay-Heriot EBLUP | Spatial Fay-Herriot EBLUP | Rao-Yu EBLUP | Spatio-temporal EBLUP |
| Dolnośląskie | -4.759 | -4.620 | -6.413 | -6.455 |
| Kujaw. Pomor. | -0.549 | -0.049 | -0.606 | -0.948 |
| Lubelskie | 1.052 | 0.544 | 5.474 | 5.492 |
| Lubuskie | -2.020 | -3.089 | 4.013 | 3.762 |
| Łódzkie | -1.384 | -1.501 | -1.017 | -1.241 |
| Małopolskie | 1.520 | 1.803 | 4.594 | 4.563 |
| Mazowieckie | 0.263 | 0.150 | 1.101 | 0.885 |
| Opolskie | -0.844 | -2.366 | 0.518 | 1.098 |
| Podkarpackie | 0.397 | 0.688 | 2.511 | 2.395 |
| Podlaskie | -1.370 | -1.182 | -0.562 | -0.931 |
| Pomorskie | 1.287 | 2.550 | 1.615 | 1.474 |
| Śląskie | 0.251 | 0.185 | 2.123 | 1.958 |
| Świętokrzyskie | -1.437 | -1.473 | 0.028 | -0.089 |
| Warm.-Mazur. | 1.673 | 1.149 | 3.171 | 2.646 |
| Wielkopolskie | 1.115 | 1.065 | 2.324 | 2.485 |
| Zach.-pomor. | -0.752 | 0.979 | 0.61 | 0.645 |

*Source: Authors' calculations based on the Polish HBS and Local Data Bank.*

**Figure 2.** Pair scatterplots for CC obtained for different model-based estimates



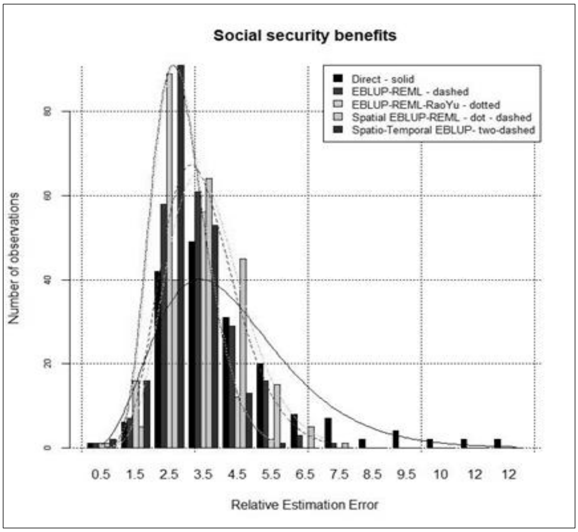

*Source: Authors' calculations based on the Polish HBS and Local Data Bank.*

**Table 5.** Estimates of REE [in %] for different estimates of income per capita from *social security benefits* by voivodships for the year 2013

| Region | Relative estimation error [in %] | | | | |
|---|---|---|---|---|---|
| | Direct | Fay-Heriot EBLUP | Spatial Fay-Herriot EBLUP | Rao-Yu EBLUP | Spatio-temporal EBLUP |
| Dolnośląskie | 4.28 | 3.96 | 4.10 | 2.75 | 2.55 |
| Kujaw. Pomor. | 3.69 | 3.50 | 3.67 | 2.64 | 2.34 |
| Lubelskie | 3.50 | 3.39 | 3.63 | 2.65 | 2.78 |
| Lubuskie | 6.70 | 5.61 | 5.91 | 3.24 | 3.17 |
| Łódzkie | 2.97 | 2.88 | 3.04 | 2.31 | 2.14 |
| Małopolskie | 3.74 | 3.48 | 3.68 | 2.63 | 2.43 |
| Mazowieckie | 3.23 | 3.27 | 3.46 | 2.41 | 2.44 |
| Opolskie | 5.89 | 4.96 | 5.39 | 3.30 | 2.82 |
| Podkarpackie | 3.41 | 3.28 | 3.48 | 2.57 | 2.63 |
| Podlaskie | 5.03 | 4.58 | 4.86 | 3.00 | 2.95 |
| Pomorskie | 5.62 | 4.85 | 4.87 | 3.25 | 3.43 |
| Śląskie | 2.86 | 2.90 | 3.07 | 2.16 | 2.30 |
| Świętokrzyskie | 4.49 | 4.15 | 4.48 | 2.89 | 3.07 |
| Warm.-Mazur. | 5.83 | 5.04 | 5.24 | 3.45 | 3.88 |
| Wielkopolskie | 2.48 | 2.44 | 2.52 | 2.10 | 2.23 |
| Zach.-pomor. | 7.52 | 5.75 | 5.72 | 3.39 | 3.49 |

*Source: Authors' calculations based on the Polish HBS and Local Data Bank.*

**Figure 3.** Distribution of estimated relative estimation errors (REE) for different estimation methods (direct and model-based)



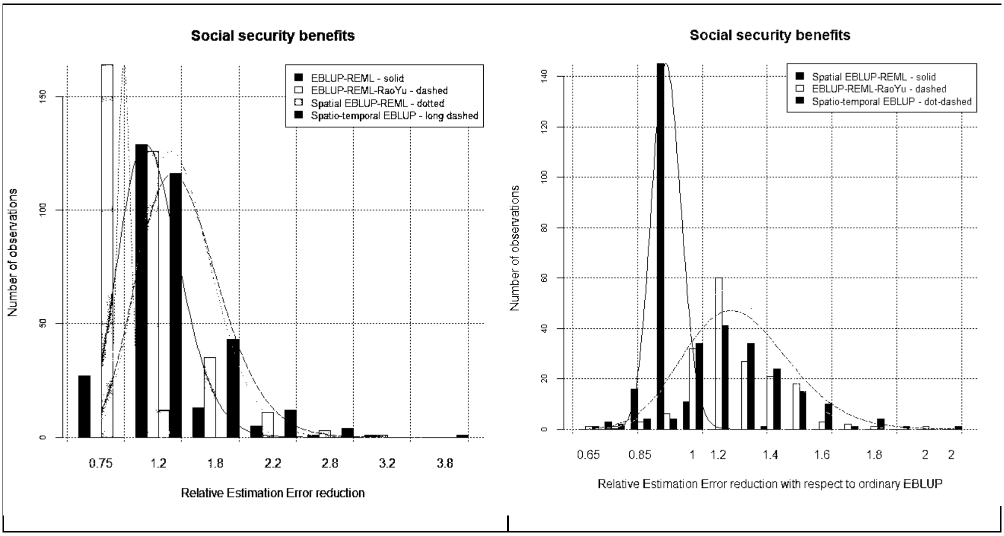

*Source: Authors' calculations based on the Polish HBS and Local Data Bank.*

**Table 6.** REE Reduction for income per capita from *social security benefits* related to the direct and FH estimates by voivodships for the year 2013

| Region | FH EBLUP | SFH EBLUP | | RY EBLUP | | STFH EBLUP | |
|---|---|---|---|---|---|---|---|
| | REE reduction | REE reduction | Spatial REE reduction | REE reduction | Spatio-temporal REE reduction | REE reduction | Spatio-temporal REE reduction |
| Dolnośląskie | 1.0796 | 1.0431 | 0.9662 | 1.5575 | 1.4427 | 1.6802 | 1.5563 |
| Kujaw. Pomor. | 1.0549 | 1.0066 | 0.9542 | 1.4010 | 1.3281 | 1.5798 | 1.4976 |
| Lubelskie | 1.0342 | 0.9663 | 0.9344 | 1.3231 | 1.2794 | 1.2593 | 1.2177 |
| Lubuskie | 1.1946 | 1.1339 | 0.9491 | 2.0684 | 1.7314 | 2.1127 | 1.7685 |
| Łódzkie | 1.0312 | 0.9775 | 0.9479 | 1.2860 | 1.2471 | 1.3863 | 1.3444 |
| Małopolskie | 1.0756 | 1.0154 | 0.9440 | 1.4191 | 1.3193 | 1.5359 | 1.4279 |
| Mazowieckie | 0.9874 | 0.9330 | 0.9449 | 1.3415 | 1.3587 | 1.3272 | 1.3442 |
| Opolskie | 1.1858 | 1.0916 | 0.9206 | 1.7817 | 1.5026 | 2.0906 | 1.7631 |
| Podkarpackie | 1.0403 | 0.9813 | 0.9432 | 1.3257 | 1.2743 | 1.2957 | 1.2455 |
| Podlaskie | 1.0974 | 1.0335 | 0.9418 | 1.6770 | 1.5281 | 1.7060 | 1.5546 |
| Pomorskie | 1.1588 | 1.1536 | 0.9955 | 1.7287 | 1.4918 | 1.6369 | 1.4126 |
| Śląskie | 0.9882 | 0.9335 | 0.9446 | 1.3267 | 1.3425 | 1.2450 | 1.2599 |
| Świętokrzyskie | 1.0812 | 1.0026 | 0.9273 | 1.5520 | 1.4354 | 1.4598 | 1.3501 |
| Warm.-Mazur. | 1.1570 | 1.1122 | 0.9613 | 1.6911 | 1.4617 | 1.5039 | 1.2999 |
| Wielkopolskie | 1.0156 | 0.9858 | 0.9706 | 1.1819 | 1.1637 | 1.1118 | 1.0947 |
| Zach.-pomor. | 1.3075 | 1.3141 | 1.0051 | 2.2186 | 1.6969 | 2.1542 | 1.6476 |

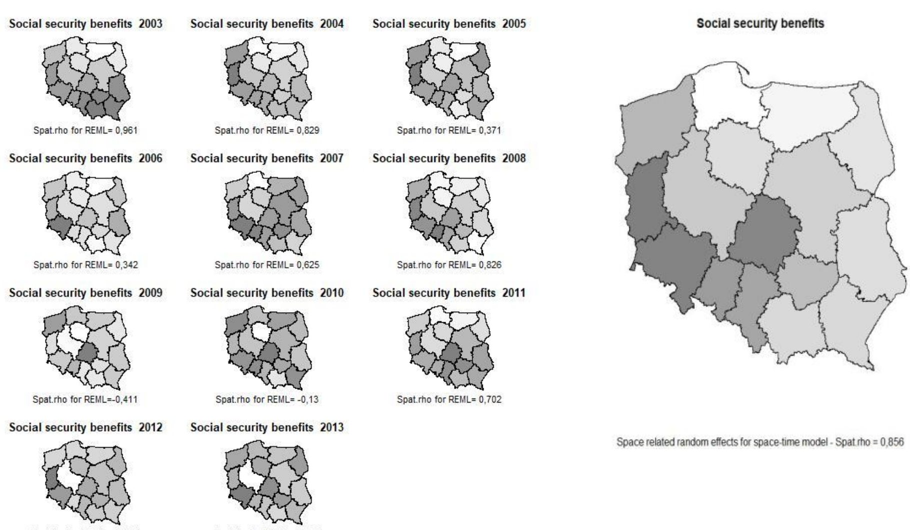
*Source: Authors' calculations based on the Polish HBS and Local Data Bank.*

**Figure 4.** Distribution of REE Reduction (left) and REE Reduction due to spatio-temporal effects (right) for different model-based estimation methods



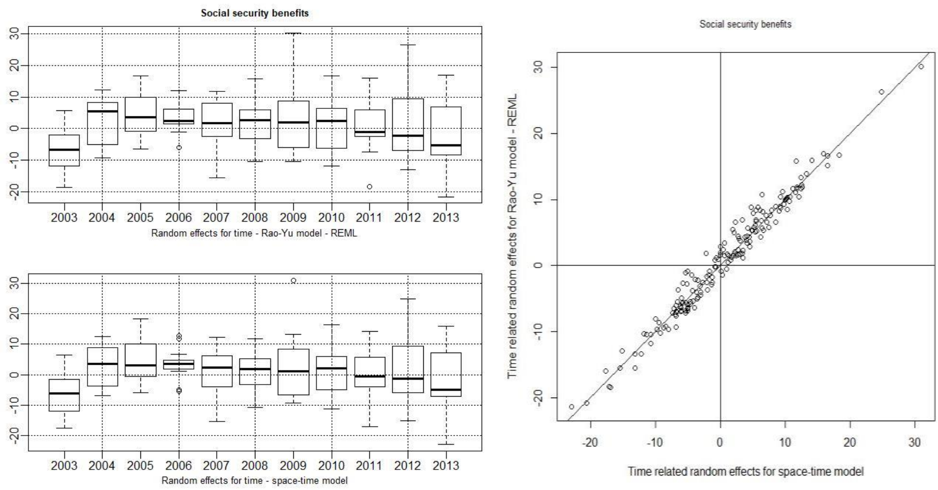

*Source: Authors' calculations based on the Polish HBS and Local Data Bank.*

**Figure 5.** Choropleth maps for **spatial random effects** of Spatial Fay-Herriot model (left) and Spatio-Temporal model (right) for *per capita income from social security benefits* by region




*Source: Authors' calculations based on the Polish HBS and Local Data Bank.*

**Figure 6.** Distributions of random **time-related effects** for RY model (top-left), and STFH model (bottom-left) and the scatterplot (right) for time-related effects ($u_2$) for STFH and Rao-Yu model.




*Source: Authors' calculations based on the Polish HBS and Local Data Bank.*

The maps presented in Figure 5 show the spatial structure of space-related random effects. It can be noticed that for the spatial models under consideration (but not for all years) significant spatial relationships are obtained. Note that such a behaviour becomes evident for the regions where higher spatial autoregression coefficients are observed, i.e. for *north-western* and *south-eastern* regions. This confirms the well-known relationships of regional differences in Poland and is obviously connected with higher industrialization of these voivodships (which may cause larger income from social security benefits). Similar conclusions were made in Kubacki and Jędrzejczak (2016), where spatial relationships for small area models in Polish counties were presented.

Interesting relationships can be observed when comparing the distributions of time-related random effects obtained for the Rao-Yu model and for the spatio-temporal model, as illustrated in Figure 6. Here some consistency between time-related random effects is noticeable. It results from the fact that in the STFH model as well as in the Rao-Yu model they follow the autoregressive process of the first order, AR(1). This regularity can also be observed in the distributions of random effects presented for each year and in the scatterplot obtained for all the years. Note that the values of random effects related to time were determined by different estimators, and calculated using different software (in particular sae and sae2 packages).

Some consistency can also be observed for REE and REE reduction distributions (Fig. 3, Fig. 4), obtained for the Rao-Yu and spatio-temporal models. It should be noted, however, that the methods used to obtain the REE values for these models were also different. In the case of the Rao-Yu model, the method applied for MSE estimation was based on extensions of the Prasad and Rao

(1990) approach, while in the case of spatio-temporal model it was based on the parametric bootstrap technique. Of course, these methods have different implementations, which may indicate that they are convergent. Such relationships were also obtained for other income-related variables.

## 6. Conclusions

The paper shows a procedure of efficient estimation for small areas based on the application of a spatio-temporal model, i.e. the general linear mixed model with spatially correlated random effects and significant correlation over time. In particular, spatial Simultaneous Autoregressive Process, using spatial neighbourhood as auxiliary information, and AR(1) process for time-related random effects, were incorporated into the estimation.

The presented spatio-temporal model improves the precision of small-area estimates not only in relation to direct estimates, which is easy to obtain, but also in comparison with other indirect techniques based on small-area models, also spatial small area models and sometimes the Rao-Yu model.

The efficiency of the proposed method was proven based on real-world examples prepared for the Polish data coming from the Household Budget Survey and the administrative data. The detailed comparison of relative estimation errors and REE reductions shows that all the considered model-based techniques are significantly more efficient than the direct estimation one, yet the spatio-temporal and the Rao-Yu models provide greater REE reduction than the others. The calculations, where some additional assumptions on the spatial relationships were made, also confirm efficiency gains of the estimators. However, such a correspondence does not always occur for all the years, so one should be conscious that for lower $\rho_2$ values the benefit of using the spatial method may be ambiguous.

It is worth pointing out that the number of observations used to fit the area-level models was small so the model parameters were estimated with less efficiency and therefore the efficiency gains with respect to direct estimators were obviously smaller than under the unit level models. What is more, the applied models require normality of random effects for MSE estimation and violating this assumption can seriously affect the results.

A more detailed analysis also reveals some correspondence between the Rao-Yu model and spatio-temporal models induced by identical assumptions about the stochastic process for time-related random effects. Further benefits can be expected when time-dependent nonlinear relationships are taken into account, for example nonlinear dependence on explanatory variables. The previously performed analysis of nonlinear models (see Jędrzejczak, Kubacki (2016), Jędrzejczak, Kubacki (2017)) may be a starting point for more detailed comparisons between the Rao-Yu method, nonlinear models and econometric panel models.

## Acknowledgement

We thank Mr. Adam Graboś from the Statistics Poland Computing Centre for providing us with the digital maps of borders of the Polish Regions, which were used for visualizations presented in this article.

## REFERENCES

BIVAND, R., LEWIN-KOH, N., (2013). maptools: Tools for reading and handling spatial objects, R package version 0.8-25, http://CRAN.R-project.org/package=maptools.

BIVAND, R., PIRAS, G., (2015). Comparing Implementations of Estimation Methods for Spatial Econometrics, Journal of Statistical Software, 63, No. 18, pp. 1–36, http://www.jstatsoft.org/v63/i18/.

CRESSIE, N. A .C., (1991). Small-area prediction of undercount using the general linear model, Proceedings of Statistic Symposium 90: Measurement and Improvement of Data Quality, Ottawa, Statistics Canada, pp. 93–105.

FAY, R. E., DIALLO, M., (2015). sae2: Small Area Estimation: Time-series Models, package version 0.1-1, https://cran.r-project.org/web/packages/sae2/index.html.

FAY, R. E., HERRIOT, R. A., (1979). Estimation of Income for Small Places: An Application of James-Stein Procedures to Census Data, "Journal of the American Statistical Association", 74, pp. 269–277, http://www.jstor.org/stable/2286322.

FAY, R. E., LI, J., (2012). Rethinking the NCVS: Subnational Goals through Direct Estimation, presented at the 2012 Federal Committee on Statistical Methodology Conference, Washington, DC, Jan. 10–12.

GRIFFITH, D. A., PAELINCK, J. H. P., (2011). Non-standard Spatial Statistics and Spatial Econometrics, Advances in Geographic Information Science, Springer Berlin Heidelberg.

JĘDRZEJCZAK, A., KUBACKI, J., (2016). Estimation of Mean Income for Small Areas in Poland Using Rao-Yu Model, Acta Universitatis Lodziensis, Folia Oeconomica, 3 (322), pp. 37–53, https://czasopisma.uni.lodz.pl/foe/article/view/755/660.

JĘDRZEJCZAK, A., KUBACKI, J., (2017). Analiza rozkładów dochodu rozporządzalnego według województw z uwzględnieniem czasu, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Nr 469, Taksonomia 29, Klasyfikacja i analiza danych – teoria i zastosowania, pp. 69–81, https://www.dbc.wroc.pl/dlibra/publication/41063/edition/37066/content?ref=desc.

KUBACKI, J., JĘDRZEJCZAK, A., (2016) Small Area Estimation of Income Under Spatial SAR Model, Statistics in Transition new series, September 2016, Vol. 17, No. 3, pp. 365–390, https://www.exeley.com/exeley/journals/statistics_in_transition/18/4/pdf/10.21307_stattrans-2017-009.pdf.

LI, J., DIALLO, M. S., FAY, R. E., (2012). Rethinking the NCVS: Small Area Approaches to Estimating Crime, presented at the Federal Committee on Statistical Methodology Conference, Washington, DC, Jan. 10–12.

MARHUENDA, Y. MOLINA, I., MORALES, D., (2013). Small area estimation with spatio-temporal Fay–Herriot models", Computational Statistics & Data Analysis, Vol. 58(C), pp. 308–325.

MOLINA, I., MARHUENDA, Y., (2013). sae: Small Area Estimation, R package version 1.0-2, http://CRAN.R-project.org/package=sae.

MOLINA, I., MARHUENDA, Y., (2015). R package sae: Methodology – sae package vignette: https://cran.r-project.org/web/packages/sae/vignettes/sae_methodology.pdf

MOLINA, I., MARHUENDA, Y., (2015). sae: An R Package for Small Area Estimation, The R Journal, Vol. 7, No. 1, pp. 81–98, http://journal.r-project.org/archive/2015-1/molina-marhuenda.pdf.

PETRUCCI, A., SALVATI, N., (2006). Small Area Estimation for Spatial Correlation in Watershed Erosion Assessment, Journal of Agricultural, Biological & Environmental Statistics, 11, No. 2, pp. 169–182, http://dx.doi.org/10.1198/108571106x110531.

PRASAD, N. G. N., RAO, J. N. K., (1990). The estimation of mean squared error of small area estimators. Journal of the American Statistical Association, 85, pp. 163–171, https://www.jstor.org/stable/2289539.

PRATESI, M., SALVATI, N., (2004). Spatial EBLUP in agricultural survey. An application based on census data, Working paper no. 256, Universitá di Pisa, Dipartimento di statistica e matematica applicata all'economia.

PRATESI, M., SALVATI, N., (2005). Small Area Estimation: The EBLUP Estimator with Autoregressive Random Area Effects, Working paper n. 261 Pubblicazioni del Dipartimento di statistica e matematica applicata all'economia.

PRATESI, M., SALVATI, N., (2008). Small area estimation: the EBLUP estimator based on spatially correlated random area effects, Statistical Methods and Applications, 17, No. 1, pp. 113–141, http://dx.doi.org/10.1007/s10260-007-0061-9.

RAO, J. N. K., MOLINA, I., (2015). Small Area Estimation (2nd edition), John Wiley & Sons, Inc., Hoboken, New Jersey.

RAO, J. N. K., YU, M. (1992). Small area estimation combining time series and cross-sectional data. “Proc. Survey Research Methods Section. Amer. Statist. Assoc.”, pp. 1–9, http://www.asasrms.org/Proceedings/papers/1992_001.pdf.

RAO, J. N. K., YU, M., (1994). Small-Area Estimation by Combining Time-Series and Cross-Sectional Data, “The Canadian Journal of Statistics”, Vol. 22, No. 4, pp. 511–528, http://www.jstor.org/stable/3315407.

R CORE TEAM, (2016). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.

SAEI, A., CHAMBERS, R., (2003). Small Area Estimation Under Linear and Generalized Linear Mixed Models With Time and Area Effects, M03/15, Southampton Statistical Sciences Research Institute, http://eprints.soton.ac.uk/8165/.

SINGH, B. B., SHUKLA, G. K., KUNDU, D., (2005). Spatio-Temporal Models in Small Area Estimation, Survey Methodology, 31, No. 2, pp. 183–195.

WESTAT, (2007). WesVar® 4.3 User’s Guide.

# APPENDIX

## Simple R code illustrating the computations

```
# reading the libraries
library(RODBC)
library(sae2)
library(sae)
library(maptools)
library(spdep)

# obtaining the proximity matrix
region.poly <- readShapePoly("Polish_regions")
region.nb   <- poly2nb(region.poly)
W <- nb2mat(wojew.nb, style = "W")
# reading the data from Excel spreadsheet
channel <- odbcConnectExcel2007("Input.xlsx",sep="")
command <- paste("select * from [Sheet1$] order by region,year", sep="")
base <- sqlQuery(channel, command)
d <- cbind(base,desvar=(base[,3])^2)
# please note that position of variance (or standard error) variable in Input file
may be different for particular case
# variable for number of domains
D <- 16
# variable for number of time periods
T <- 11
# formula for particular model - see for example Table 1
formula <- "D905_AVG ~ PRZECGOSP + PRZECEMER + PKB_PC"

# obtaining the Rao-Yu estimates
resultT.RY <- eblupRY(as.formula(formula), D, T, vardir =
diag((base[,3])^2),data=base, ids=base$region, method="REML")
# obtaining the decomposition of random effects for Rao-Yu model
```

```
resultT.RY_RE <- eblupRY_randeff_1d(formula, D, T, vardir = diag((base[,3])^2),
data=base, delta = resultT.RY$delta)
# obtaining the spatio-temporal estimates
resultST <- eblupSTFH(as.formula(formula), D, T, desvar, W, data=d)
# obtaining the MSE valus for spatio-temporal model using the parametric
bootstrap procedures
resultPBST <- pbmseSTFH(as.formula(formula), D, T, desvar, W, data=d)
# obtaining the decomposition of random effects for spatio-temporal model
resultST_RE <- eblupSTFH_randeff(as.formula(formula), D, T, vardir =
(base[,3])^2, W, data=d, rho1 = resultST$fit$estvarcomp[2,1], rho2 =
resultST$fit$estvarcomp[4,1], sigma21 = resultST$fit$estvarcomp[1,1], sigma22 =
resultST$fit$estvarcomp[3,1])
.
```