STATISTICS IN TRANSITION new series, September 2016
Vol. 17, No. 3, pp. 365–390

SMALL AREA ESTIMATION OF INCOME UNDER SPATIAL SAR MODEL

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

ABSTRACT

The paper presents the method of hierarchical Bayes (HB) estimation under small area models with spatially correlated random effects and a spatial structure implied by the Simultaneous Autoregressive (SAR) process. The idea was to improve the spatial EBLUP by incorporating the HB approach into the estimation algorithm. The computation procedure applied in the paper uses the concept of sampling from a posterior distribution under generalized linear mixed models implemented in WinBUGS software and adapts the idea of parameter estimation for small areas by means of the HB method in the case of known model hyperparameters. The illustration of the approach mentioned above was based on a real-world example concerning household income data. The precision of the direct estimators was determined using own three-stage procedure which employs Balanced Repeated Replication, bootstrap and Generalized Variance Function. Additional simulations were conducted to show the influence of the spatial autoregression coefficient on the estimation error reduction. The computations performed by 'sae' package for R project and a special procedure for WinBUGS reveal that the method provides reliable estimates of small area means. For high spatial correlation between domains, noticeable MSE reduction was observed, which seems more evident for HB-SAR method as compared with the traditional spatial EBLUP. In our opinion, the Gibbs sampler, revealing the simultaneous nature of processes, especially for random effects, can be a good starting point for the simulations based on stochastic SAR processes.

Key words: small area estimation (SAE), SAR model, hierarchical Bayes estimation, spatial empirical best linear unbiased predictor.

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

[1] Centre of Mathematical Statistics, Statistical Office in Łódź. E-mail: j.kubacki@stat.gov.pl.
[2] Institute of Statistics and Demography, University of Łódź; Centre of Mathematical Statistics, Statistical Office in Łódź. E-mail: jedrzej@uni.lodz.pl.

(in Poland – voivodships). Smaller areas are usually not included into sampling designs mainly because of financial and organizational limitations, and the overall sample size is seldom large enough to yield direct estimates of adequate precision for all the domains of interest. In such cases the inferences are connected with large estimation errors which make them 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 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 distinct advantage over other methods. The model-based approach treats the population values as random and the associated inferences are based on the probability distribution induced by the assumed superpopulation model. One of these techniques is the Spatial EBLUP (Spatial Empirical Best Linear Unbiased Prediction). It is usually based on the assumption that the spatial relationships between domains can be modelled by the simultaneous autoregressive process SAR (see Pratesi and Salvati (2008), p. 114 for better explanation of this term). The method was introduced by Cressie (1991) and is explained in detail in the publications of Saei and Chambers (2003), Pratesi and Salvati (2004, 2005, 2008), Singh et al. (2005), Petrucci and Salvati (2006). The spatial SAR estimation was also applied in SAMPLE project (2010) for the purpose of bootstrap estimation of the MSE for the populations having various spatial autocorrelation levels. 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), where MCMC (Markov Chain Monte Carlo) applications for spatial models are presented.

In the paper we compare two approaches to the spatial SAR modelling implemented for small area estimation. Besides the above-mentioned ordinary Spatial EBLUP, we develop a HB model, which is based on the spatial autoregressive structure of random effects incorporated into Bayesian inference. The model will be called the SAR HB model.

In our opinion, HB estimation can be practically appealing with respect to the traditional EBLUP approach. First, the most common method used to fit EBLUP models was the ML method, although maximum likelihood estimators are asymptotic in nature and little is known about their behaviour in small samples. Moreover, when using the HB approach it is possible not only to obtain the point estimates of the parameters, but also approximate their distributions (including the distributions of model variance and random effects). For SAR process, one can also obtain the approximation of spatial autoregression coefficient distribution. It may be helpful in obtaining the model diagnostics, which is a non-trivial problem in the case of linear mixed models.

2. Small area model for spatially correlated random effects based on SAR process.

In the paper a special case of the area level model (type-A model) is discussed, where the parameter of interest is a vector $\boldsymbol{\theta}$ of size m (where m is the number of small areas), which is related to the direct estimator $\hat{\theta}$ of this quantity by means of the following relationship

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta} + \mathbf{e} \tag{1}$$

where $\mathbf{e}$ is a vector of independent sampling errors having mean 0 and diagonal variance matrix $\boldsymbol{\Psi}$. The parameter $\boldsymbol{\theta}$ also satisfies the common relationship connected with linear mixed models, which incorporates the spatial correlation between areas. This relationship is as follows

$$\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} \tag{2}$$

where $\mathbf{X}$ is the matrix of area-dependent auxiliary variables of size $m \times p$, $\boldsymbol{\beta}$ is the vector of regression parameters of size $p \times 1$, $\mathbf{Z}$ is the matrix ($m \times m$) of known positive constants and $\mathbf{v}$ is the $m \times 1$ vector of the second order variation. Within the scope of the study it is assumed that the random effects are described by the SAR process. In such a case the vector $\mathbf{v}$ can be described as

$$\mathbf{v} = \rho\mathbf{W}\mathbf{v} + \mathbf{u} \;\Rightarrow \mathbf{v} = (\mathbf{I}_{\boldsymbol{m}} - \rho\mathbf{W})^{-1}\mathbf{u} \tag{3}$$

where ρ is the parameter of the spatial autoregression and $\mathbf{W}$ is the spatial weight matrix (of size $m \times m$), which can be defined in many different ways. In the paper the entries of the spatial weight matrix take values in the interval <0,1> and indicate whether the row and column domains are neighbours or not. The additional restriction imposed on $\mathbf{W}$ is that row elements add up to 1, $\mathbf{u}$ is the vector of independent error term with zero mean and constant variance σ_u^2 and $\mathbf{I_m}$ is the identity matrix of size $m \times m$. The random effects have the following covariance matrix $\mathbf{G}$ (also called SAR dispersion matrix)

$$\mathbf{G} = \sigma_u^2[(\mathbf{I}_{\boldsymbol{m}} - \rho\mathbf{W})^T(\mathbf{I}_{\boldsymbol{m}} - \rho\mathbf{W})]^{-1} \tag{4}$$

and the sampling error $\mathbf{e}$ has the following covariance matrix

$$\mathbf{R} = \boldsymbol{\Psi} = diag(\psi_i) \tag{5}$$

Further, we will assume that the matrix $\mathbf{Z}$ is equal to $\mathbf{I_m}$. Thus, using (1), (2) and (3) the model can be described as follows

$$\widehat{\boldsymbol{\theta}} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_{\boldsymbol{m}} - \rho\mathbf{W})^{-1}\mathbf{u} + \mathbf{e} \tag{6}$$

The covariance matrix for $\hat{\theta}$ is equal to

$$\mathbf{V} = \mathbf{G} + \boldsymbol{\Psi} \tag{7}$$

Under the model (8) the Spatial EBLUP estimator is equal to (see for example formula (8) in Pratesi and Salvati (2008))

$$\widetilde{\boldsymbol{\theta}}_i = \mathbf{x}_\mathbf{i}^\mathbf{T}\widetilde{\boldsymbol{\beta}} + \boldsymbol{b}_i^T \boldsymbol{G}\boldsymbol{V}^{-1}(\widehat{\boldsymbol{\theta}} - \mathbf{X}\widetilde{\boldsymbol{\beta}}) \tag{8}$$

where $\widetilde{\boldsymbol{\beta}} = [\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{V}^{-1}\widehat{\boldsymbol{\theta}}$ is the generalized least squares estimator of the regression parameter and $\boldsymbol{b}_i^T$ is the $1\times m$ vector (0,…,0,1,0,…0) with 1 in the i-th position. This estimator is dependent on σ_u^2 and ρ. These parameters can be obtained by means of the Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML) method (where a log likelihood function is used), both applying the Fisher scoring algorithm. The method was implemented, for example, in 'sae' package for R-project. More details on this procedure can be found in the SAMPLE project deliverable 22 (2011) in part 2.1.2.

The mean square error for the Spatial EBLUP estimator (8) can be expressed as the sum of four components, which can be given by (see for example formula (2.17) in Singh et al. (2005) or formula (43) in Molina and Marhuenda (2015))

$$mse\left[\tilde{\theta}_i\right] = g_{1i} + g_{2i} + 2g_{3i} - g_{4i} \tag{9}$$

where g_1 is connected with uncertainty about the small area estimate and is of order $O(1)$, g_2 is connected with uncertainty about $\widetilde{\boldsymbol{\beta}}$ and is of order $O(m^{-1})$ for large m, g_3 is connected with uncertainty about σ_u^2 (or variance components) and g_4 is connected with uncertainty of spatial autocorrelation parameter ρ. The first two components of MSE are given by

$$g_{1i} = \mathbf{b}_\mathbf{i}^\mathbf{T}[\mathbf{G} - \mathbf{G}\mathbf{V}^{-1}\mathbf{G}]\mathbf{b}_\mathbf{i} \tag{10}$$

$$g_{2i} = \mathbf{b}_\mathbf{i}^\mathbf{T}\left[\mathbf{I}_\mathbf{m} - \mathbf{G}\mathbf{V}^{-1}\right]\mathbf{X}(\mathbf{X}^\mathbf{T}\mathbf{V}^{-1}\mathbf{X})^{-1}\,\mathbf{X}^\mathbf{T}[\mathbf{I}_\mathbf{m} - \mathbf{V}^{-1}\mathbf{G}]\mathbf{b}_\mathbf{i} \tag{11}$$

The third element has a more complicated form and for the spatial EBLUP estimator can be expressed by the following equation

$$g_{3i} = trace\{\mathbf{L}_\mathbf{i}\mathbf{V}\mathbf{L}_\mathbf{i}^\mathbf{T} I^{-1}\} \tag{12}$$

where

$$\mathbf{L}_i = \begin{pmatrix} \mathbf{b}_\mathbf{i}^\mathbf{T}[\mathbf{C}^{-1}\mathbf{V}^{-1} - \sigma_u^2\mathbf{C}^{-1}\mathbf{V}^{-1}\mathbf{C}^{-1}\mathbf{V}^{-1}] \\ \mathbf{b}_\mathbf{i}^\mathbf{T}[\mathbf{A}\mathbf{V}^{-1} - \sigma_u^2\mathbf{C}^{-1}\mathbf{V}^{-1}\mathbf{A}\mathbf{V}^{-1}] \end{pmatrix} \tag{13}$$

$$\mathbf{C} = (\mathbf{I}_\mathbf{m} - \rho\mathbf{W})^T(\mathbf{I}_\mathbf{m} - \rho\mathbf{W})$$

$$\mathbf{A} = -\sigma_u^2\mathbf{C}^{-1}\frac{\partial\mathbf{C}}{\partial\rho}\mathbf{C}^{-1} = -\sigma_u^2\mathbf{C}^{-1}(-\mathbf{W} - \mathbf{W}^\mathbf{T} + 2\rho\mathbf{W}^\mathbf{T}\mathbf{W})\mathbf{C}^{-1}$$

and I^{-1} is the Fisher information matrix inverse. It depends on σ_u^2 and ρ and its elements can be expressed as

$$I(\sigma_u^2, \rho) = \begin{pmatrix} I_{\sigma_u^2\sigma_u^2} & I_{\sigma_u^2\rho} \\ I_{\rho\sigma_u^2} & I_{\rho\rho} \end{pmatrix} \tag{14}$$

and their elements are given by

$$I_{\sigma_u^2\sigma_u^2} = \tfrac{1}{2} trace\{\mathbf{V^{-1}C^{-1}V^{-1}C^{-1}}\} \tag{15}$$

$$I_{\sigma_u^2\rho} = I_{\rho\sigma_u^2} = \tfrac{1}{2} trace\{\mathbf{V^{-1}AV^{-1}C^{-1}}\} \tag{16}$$

$$I_{\rho\rho} = \tfrac{1}{2} trace\{\mathbf{V^{-1}AV^{-1}A}\} \tag{17}$$

The last term g_4 can be expressed by

$$g_{4i} = \tfrac{1}{2}\sum_{k=1}^{2}\sum_{l=1}^{2} \mathbf{b_i^T \Psi V^{-1}} \frac{\partial^2 V(\omega)}{\partial\omega_k \partial\omega_l} \mathbf{V^{-1}\Psi} I_{\mathrm{kl}}^{-1}(\omega)\mathbf{b_i} \tag{18}$$

where $\omega_1 = \sigma_u^2$, $\omega_2 = \rho$ and the second derivatives can be expressed as

$$\frac{\partial^2 \mathbf{V}(\omega)}{\partial(\sigma_u^2)^2} = 0_{m\times m}$$

$$\frac{\partial^2 \mathbf{V}(\omega)}{\partial\sigma_u^2\partial\rho} = \frac{\partial^2 \mathbf{V}(\omega)}{\partial\rho\partial\sigma_u^2} = -\mathbf{C^{-1}}\frac{\partial \mathbf{C}}{\partial\rho}\mathbf{C^{-1}}$$

$$\frac{\partial^2 \mathbf{V}(\omega)}{\partial\rho^2} = 2\sigma_u^2\mathbf{C^{-1}}\frac{\partial \mathbf{C}}{\partial\rho}\mathbf{C^{-1}}\frac{\partial \mathbf{C}}{\partial\rho}\mathbf{C^{-1}} - 2\sigma_u^2\mathbf{C^{-1}W^T WC^{-1}}$$

The above relationships were obtained under the assumption that the variance components can be described by the spatial SAR process. The spatial hierarchical Bayes model can be formulated in the manner analogous to the model (10.3.1) from Rao (2003), but in the model definition the spatial dependence between domains, determining the structure of the SAR process, should be specified (via ρ and spatial weight matrix **W**). Contrary to the other parameters, for the parameter of spatial autoregression ρ it is difficult to elicit an informative prior, either subjectively or from previous data. A uniform prior which assigns equal weight to all values of the spatial parameter seems unreasonable, as most of the SAR models based on real data sets reported in the literature have yielded moderate or large (positive) ρ estimates. When the values of ρ coefficients are treated as constants (they can be obtained from the previous Spatial EBLUP estimation), the model can be expressed as follows:

(i) $\quad \widehat{\boldsymbol{\theta}}|\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma_u^2 \sim N(\boldsymbol{\theta}, \boldsymbol{\Psi})$

(ii) $\quad \boldsymbol{\theta}|\boldsymbol{\beta}, \sigma_u^2, \rho \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma_u^2[(\mathbf{I_m} - \rho\mathbf{W})^{\boldsymbol{T}}(\mathbf{I_m} - \rho\mathbf{W})]^{-\mathbf{1}})$

(iii) $\quad f(\boldsymbol{\beta}) \propto 1$

(iv) $\quad \sigma_u^2 \sim G^{-1}(a, b)$ (19)

It has also been assumed that the initial parameters a and b of the Gamma prior describing the σ_u^2 distribution are both known. In the study they were obtained on the basis of the EBLUP models estimated for 16 regions and two time periods, which may be a good approximation of σ_u^2 variability. A similar approach was used in the previous work by Kubacki (2012) in the context of the traditional Fay-Heriott model, where the distribution of σ_u^2 was obtained from the ordinary regression. The model (19) applied in the paper is somewhat similar to that presented in Gharde, Rai and Jaggi (2013), but is an area-level (not unit-level) model and includes additional assumptions on σ_u^2 prior distribution as well as on direct relationships between the model estimates and the values $\mathbf{W}$, ρ and σ_u^2 (to be discussed further).

3. Illustration

The application of the proposed procedures to the Polish income data consisted of the following steps:

1. Direct estimation of average per capita income for counties.
2. Estimation of standard errors of the direct estimates.
3. Specification of small area with spatially correlated random effects for counties *(powiats)* [formulas (2).(3)].
4. Model-based estimation for counties based on EBLUP and Spatial EBLUP procedures [formula (9)].
5. Formulation of hierarchical Bayes model incorporating spatially correlated random effects for counties [formula (19)].
6. Implementation of computations for HB spatial model (to be described as a separate paragraph).

The variable of interest was household available income. We were particularly interested in the estimation of its average per capita value for counties, i.e. NUTS-4 areas according to the Eurostat classification. The basis of the direct estimation was the individual data coming from the Polish Household Budget Survey (HBS).

The precision of direct estimates is usually computed by means of the Balanced Repeated Replication (BRR) technique. This method is valid when the sample for each county is composed of two subsamples that allow constructing the replications called half-samples. However, in the case of extremely small samples there might not be two sub-samples for each county, so the simple bootstrap method must be used instead. Another difficulty arises when for a particular county there is no information available about the variable of interest. In such a case the Generalized Variance Function (GVF), traditionally used to smooth out the uncertainty of the design-based variance estimates, can be helpful. It is worth mentioning that the previous investigations of the authors revealed no underestimation in the bootstrap and GVF-based estimates of precision

(see Kubacki and Jędrzejczak (2012). Therefore, using such an approximation may properly reflect the precision for all counties.

In the applied small area models, two auxiliary variables coming from the Polish tax register POLTAX were specified as covariates. They include: the average salary and the average universal health insurance premium contribution, both determined by dividing the sums of the respective totals by de facto population sizes for particular NUTS-4 units.

To provide the entries of the spatial weight matrix W, necessary for the spatial model specification, the digital maps for Polish counties were used. During the computations (using 'spdep' package for R-project environment) sub-maps for regions were automatically generated, which simplified the visualization of the results.

When formulating the HB model for counties one should determine the prior distribution for the model variance σ_u^2. In the paper, ordinary EBLUP and Spatial EBLUP estimation results were used to obtain the empirical distribution of this variance. The results of these computations were summarized in Figure 1.

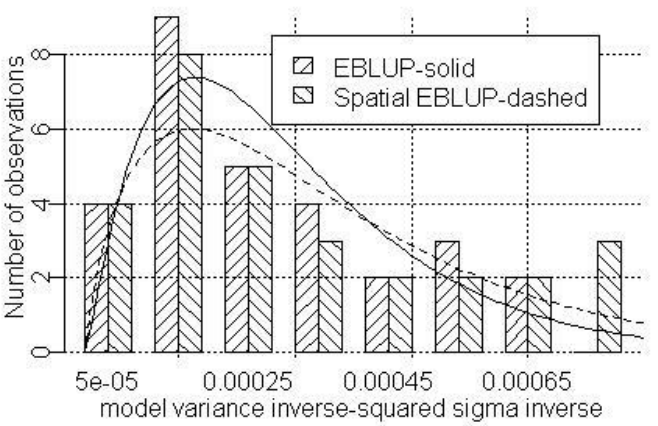

Figure 1. Empirical distributions of model variance inverse σ_u^{-2}obtained for small area models (EBLUP and Spatial EBLUP) of household per capita available income by NUTS-4 - counties in Poland.

The distributions of model errors were found similar for both EBLUP and Spatial EBLUP models. The distributions of inverse model variances (Figure 1) are both positively skew and can be approximated by gamma priors as it was assumed in the hierarchical model (19). Slight differences, which can be observed in Figure 1, seem not to have significant influence on the variability of the estimates which were obtained using EBLUP and Spatial EBLUP techniques. On the other hand, one should be careful while dealing with more specific types of income and some further simulations should be made for them.

4. Results and discussion

The results presented for the Silesian region and for the year 2004 (Tables 1, 2 and 3) show that in the case of ρ values significantly different from zero (in the case presented here it is equal to 0.681), estimation based on spatial models may significantly reduce the estimation error. This effect is more evident for HB estimator. For non-spatial models some consistency between relative estimation errors and random effect values is observed (see Figure 3). Please note that the random effects arise from the residuals obtained from generalized regression models, which can be easily derived from the second component of equation (8).

Examining the diagnostic graphs obtained for Gibbs sampler simulations, one can easily notice that in the case where ordinary HB scheme is used normality for the model estimates (denoted mu – see Figure 4), no autocorrelation (see Figure 5) and relative stability of simulations run are observed. Similar results were also presented in the authors' previous works (see: Kubacki (2012)). It can be noted that autocorrelation plots show the correlation between the values coming from the simulation obtained for the iteration k and the iteration $k+t$, where t indicates the lag between k and $k+t$ value. The absence of autocorrelation on the plot indicates that for $t=0$ the correlation is equal to 1, and for further lags it is close to zero.

However, for the spatial version of HB estimator, some autocorrelation (see Figure 9), and sometimes lack of normality (see Figure 8) in model estimates is observed. In the case considered here, this is partially due to serious direct estimation errors (as it was observed for Mikołowski, Pszczyński and Bieruńsko-Lędziński counties). The consistency between Spatial EBLUP and HB SAR-based estimates (see Figure 6), between the estimation error and the obtained random effects (see Figure 7), is relatively weaker, but it is achieved for the considered case. Random effects, obtained for Spatial EBLUP estimator, are presented on the map below (see Figure 10). Here, some regularity between the absolute values of random effects and the geographic location of the county is observed. This regularity is connected with their central or peripheral location, which means that the central part of the considered region dominates over the rest of the region, and no isolated counties (islands) in the considered region are observed.

The comparison of relative estimation error (REE) distribution and the distribution of reduction of this error shows that all the considered model-based techniques are significantly more efficient than the corresponding direct ones (Figure 11). In fact, all the considered techniques present similar efficiency and have similar REE reduction structure (Figure 12) - only HB-SAR performs slightly better, as compared to the other model-based techniques. This regularity can also be observed, when a comparison between a spatial and a respective non-spatial model is made (Figure 13).

Table 1. Estimation results for per capita available income by counties in the Silesian region obtained using direct estimation method, EBLUP method (REML technique) and HB method (Gibbs sampler).

County (NUTS-4 unit)	Available income							
	Direct estimation		EBLUP estimation – REML			HB estimation		
	Parameter estimate	REE %	Parameter estimate	REE %	Random effects	Parameter estimate	REE %	Random effects
będziński	821.59	5.82	784.71	4.28	23.81	789.52	4.40	27.52
bielski	781.94	1.22	778.63	1.20	53.92	779.40	1.19	54.18
cieszyński	762.78	4.99	734.21	3.92	29.12	738.61	3.97	33.71
częstochowski	570.65	6.54	590.44	5.00	-21.00	586.37	5.08	-21.73
gliwicki	693.20	11.14	706.85	5.19	-3.38	705.97	5.61	-3.95
kłobucki	539.04	8.00	574.29	5.59	-28.02	568.43	5.77	-30.27
lubliniecki	596.73	4.14	610.82	3.63	-34.12	608.03	3.63	-34.85
mikołowski	796.64	15.83	793.96	5.01	0.25	796.39	5.41	-0.06
myszkowski	613.46	7.74	622.49	5.35	-5.92	620.01	5.42	-5.43
pszczyński	629.75	12.14	724.15	5.41	-23.85	719.27	5.79	-30.21
raciborski	758.34	4.50	716.66	3.88	52.76	722.32	3.94	59.80
rybnicki	783.98	8.18	764.15	4.69	7.12	766.10	5.08	7.97
tarnogórski	671.46	5.07	686.73	3.92	-19.45	684.87	3.99	-21.04
bieruńsko-lędziński	625.85	11.69	764.96	4.93	-38.35	757.22	5.32	-49.19
wodzisławski	855.72	4.24	812.73	3.53	48.34	819.80	3.66	54.13
zawierciański	671.04	7.36	674.02	4.89	-1.80	672.20	5.06	-2.18
żywiecki	730.75	1.80	725.43	1.75	45.63	726.56	1.75	47.68
Bielsko-Biała city	792.14	7.38	771.68	4.42	8.84	774.27	4.66	9.78
Bytom city	705.56	1.29	705.28	1.27	4.93	705.23	1.26	5.59
Chorzów city	656.28	2.50	666.95	2.34	-58.78	664.70	2.37	-61.08
Częstochowa city	771.36	10.35	715.50	5.12	12.94	719.17	5.57	16.93
Dąbrowa Górnicza city	777.52	4.38	782.61	3.47	-6.48	783.03	3.47	-8.58
Gliwice city	745.60	5.50	761.17	3.91	-13.66	760.25	3.96	-16.57
Jastrzębie-Zdrój city	748.66	5.52	774.86	3.90	-22.65	771.66	4.01	-28.39
Jaworzno city	748.49	6.85	780.11	4.22	-17.74	778.39	4.42	-22.11
Katowice city	859.53	1.13	859.17	1.12	5.66	859.38	1.12	1.51
Mysłowice city	813.19	2.41	810.39	2.23	10.79	811.09	2.20	8.73
Piekary Śląskie city	744.14	13.14	741.89	5.31	0.35	741.68	5.88	-0.41
Ruda Śląska city	671.92	15.46	772.90	5.09	-13.82	768.99	5.75	-19.80
Rybnik city	763.03	1.45	764.52	1.41	-17.86	764.20	1.41	-20.39
Siemianowice Śląskie city	915.69	9.15	774.56	4.93	29.68	783.96	5.63	38.39
Sosnowiec city	818.88	7.44	803.93	4.39	5.95	806.06	4.65	5.59
Świętochłowice city	686.82	9.03	708.28	4.90	-8.24	706.42	5.28	-10.05
Tychy city	832.03	0.07	832.03	0.07	-4.80	832.02	0.07	-9.04
Zabrze city	703.72	0.21	703.74	0.21	-15.66	703.73	0.21	-15.83
Żory city	830.68	8.20	781.97	4.53	15.50	786.79	4.88	18.58

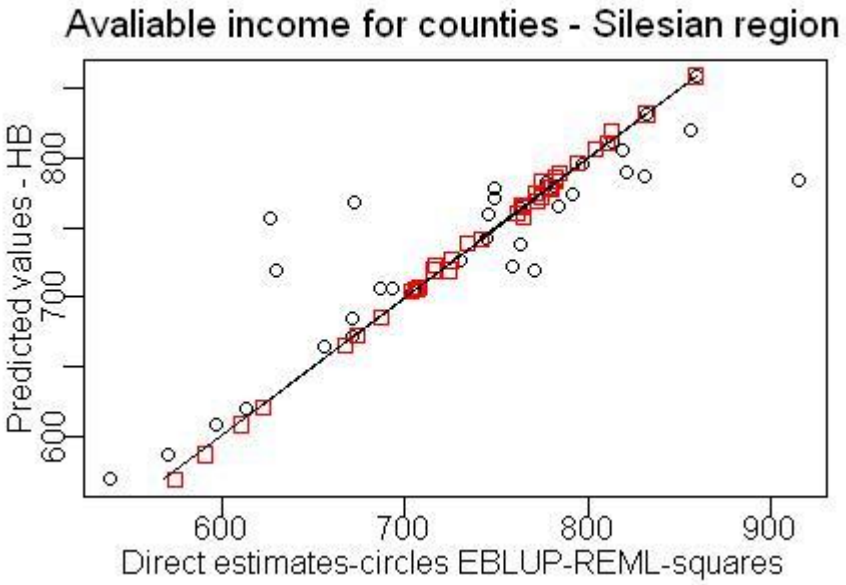

Figure 2. Observed per capita available income (direct estimates - black circles, EBLUP estimates - red squares) vs. predicted values estimated under hierarchical Bayes model for counties in the Silesian region.

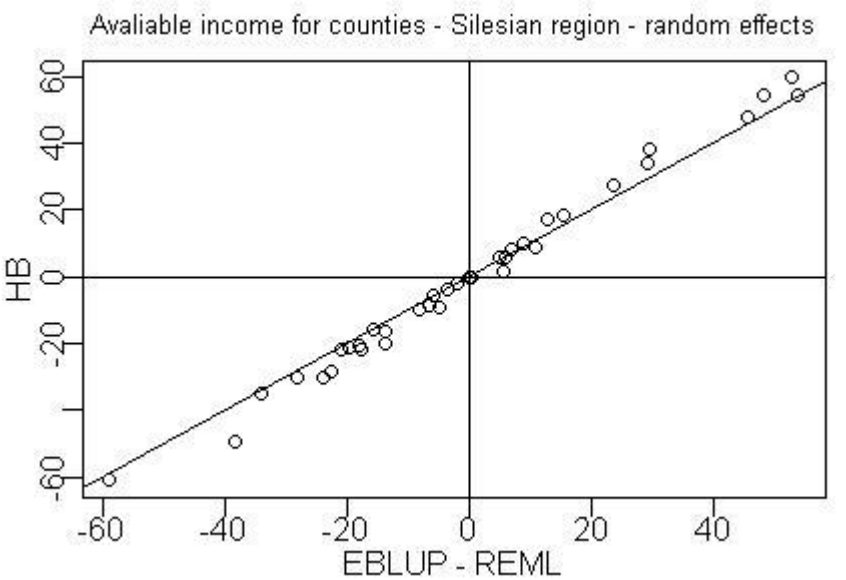

Figure 3. EBLUP vs. HB estimates of random effects for small area models of per capita available income, obtained for counties in the Silesian region.

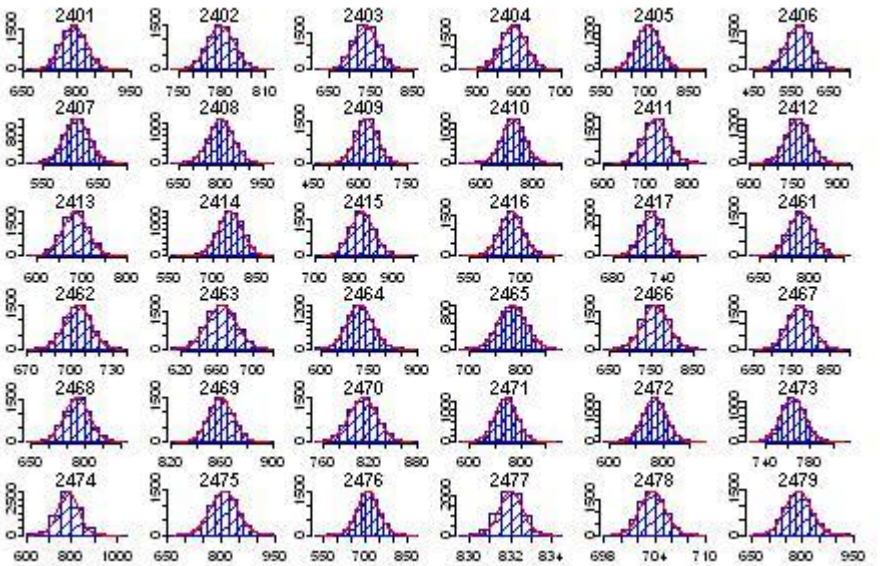

Figure 4. A posteriori distributions of per capita available income for counties in the Silesian region obtained by MCMC simulation (Gibbs sampler) under conventional HB model.

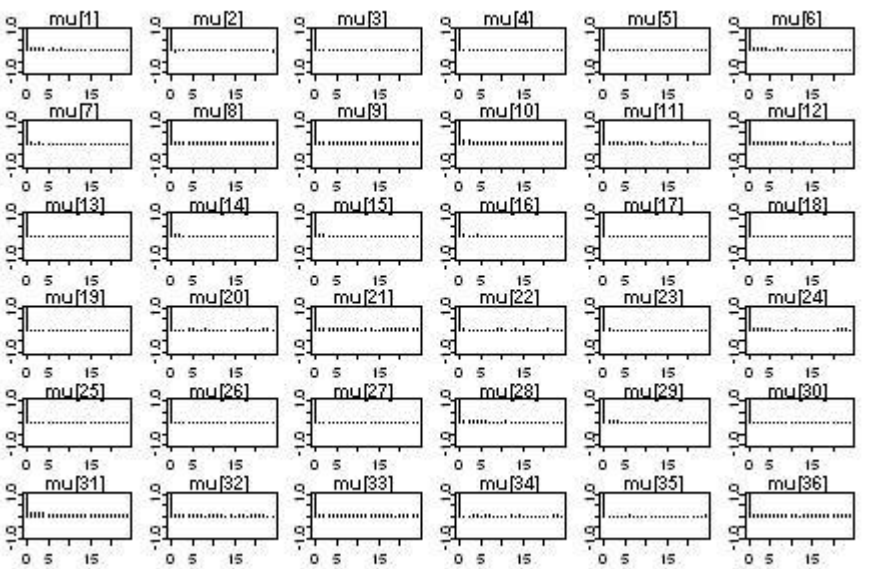

Figure 5. Autocorrelations of model estimates for per capita available income obtained for counties in the Silesian region by MCMC simulation using Gibbs sampler.

Table 2. Estimation results for per capita available income by counties in the Silesian region obtained using direct estimation method and Spatial EBLUP method (REML technique) and HB-SAR method (Gibbs sampler).

County (NUTS-4 unit)	Available income							
	Direct estimation		EBLUP estimation – Spatial REML			HB-SAR estimation		
	Parameter estimate	REE %	Parameter estimate	REE %	Random effects	Parameter estimate	REE %	Random effects
będziński	821.59	5.82	781.77	3.95	11.06	785.61	3.87	22.85
bielski	781.94	1.22	779.45	1.20	59.60	780.12	1.16	67.81
cieszyński	762.78	4.99	746.23	3.51	42.66	748.93	3.51	52.75
częstochowski	570.65	6.54	584.70	4.74	-36.01	582.12	4.68	-31.99
gliwicki	693.20	11.14	711.72	4.55	-5.22	714.97	3.88	5.50
kłobucki	539.04	8.00	569.85	5.37	-41.62	566.06	5.40	-38.91
lubliniecki	596.73	4.14	603.87	3.58	-41.71	601.72	3.54	-37.00
mikołowski	796.64	15.83	788.05	4.24	4.11	798.48	5.15	22.67
myszkowski	613.46	7.74	618.68	4.95	-22.87	617.30	4.89	-17.47
pszczyński	629.75	12.14	734.98	4.50	-0.23	747.92	4.78	20.42
raciborski	758.34	4.50	710.69	4.00	49.21	718.45	3.90	63.98
rybnicki	783.98	8.18	780.23	4.20	17.16	782.12	4.34	26.94
tarnogórski	671.46	5.07	683.65	3.66	-24.50	684.51	3.56	-16.23
bieruńsko-lędziński	625.85	11.69	770.56	4.07	-20.52	776.08	4.08	-6.79
wodzisławski	855.72	4.24	813.85	3.47	42.01	819.97	3.56	56.09
zawierciański	671.04	7.36	673.66	4.56	-12.07	671.67	4.64	-6.88
żywiecki	730.75	1.80	729.35	1.75	51.79	729.74	1.69	59.34
Bielsko-Biała city	792.14	7.38	799.78	3.96	41.89	797.44	4.03	47.44
Bytom city	705.56	1.29	703.46	1.27	-6.94	703.80	1.26	0.81
Chorzów city	656.28	2.50	672.64	2.31	-62.83	668.96	2.28	-58.88
Częstochowa city	771.36	10.35	682.88	5.31	-19.93	683.66	5.42	-11.78
Dąbrowa Górnicza city	777.52	4.38	781.40	3.48	-0.41	780.42	3.35	6.72
Gliwice city	745.60	5.50	753.29	3.78	-17.26	753.28	3.57	-9.28
Jastrzębie-Zdrój city	748.66	5.52	795.11	3.67	-1.68	789.51	3.61	0.94
Jaworzno city	748.49	6.85	784.13	3.88	-10.23	782.63	3.92	-3.53
Katowice city	859.53	1.13	857.99	1.12	9.63	859.12	1.07	19.47
Mysłowice city	813.19	2.41	806.30	2.22	11.49	806.86	2.15	20.27
Piekary Śląskie city	744.14	13.14	738.50	4.50	-12.01	738.26	4.86	-4.48
Ruda Śląska city	671.92	15.46	765.34	4.29	-25.23	758.58	4.60	-23.85
Rybnik city	763.03	1.45	768.11	1.41	-10.11	767.02	1.39	-3.12
Siemianowice Śląskie city	915.69	9.15	760.25	4.35	7.50	768.19	5.00	23.23
Sosnowiec city	818.88	7.44	807.21	3.85	8.00	805.53	4.00	14.55
Świętochłowice city	686.82	9.03	692.04	4.51	-29.30	689.37	4.76	-24.46
Tychy city	832.03	0.07	832.03	0.07	10.70	832.02	0.07	19.18
Zabrze city	703.72	0.21	703.73	0.21	-16.69	703.71	0.21	-9.18
Żory city	830.68	8.20	770.40	4.19	8.08	773.90	4.45	19.50

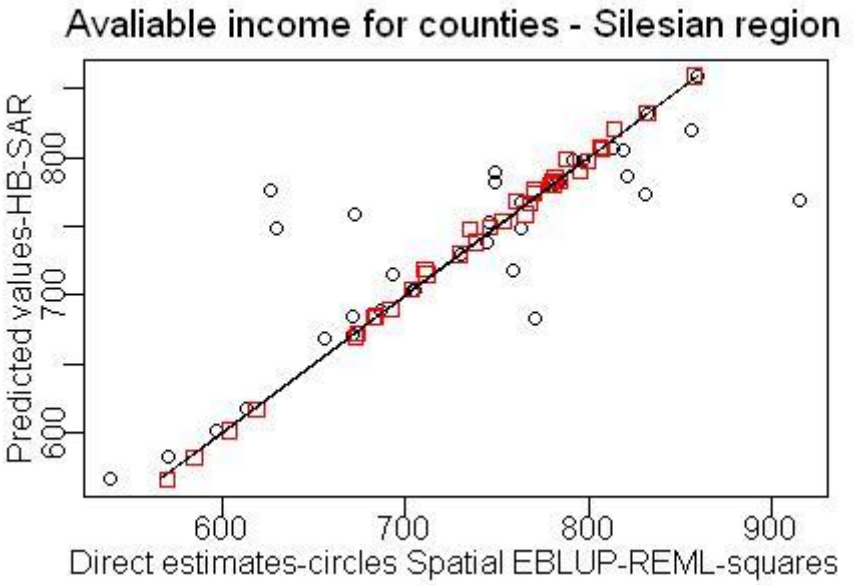

Figure 6. Observed per capita available income (direct estimates - black circles, EBLUP estimates - red squares) vs. predicted values estimated under hierarchical Bayes with SAR relationships between areas, for counties in the Silesian region.

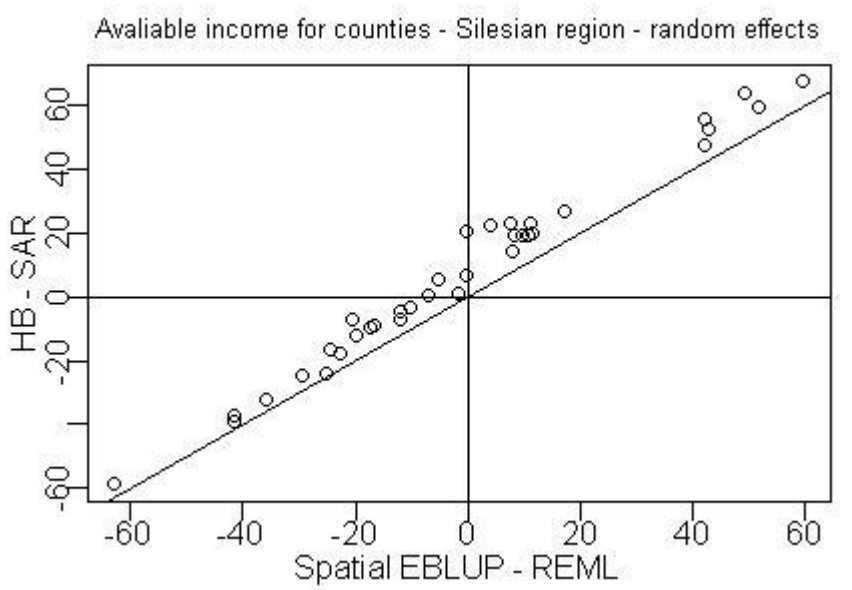

Figure 7. EBLUP vs. HB-SAR estimates of random effects for small area models of per capita available income, obtained for counties in the Silesian region.

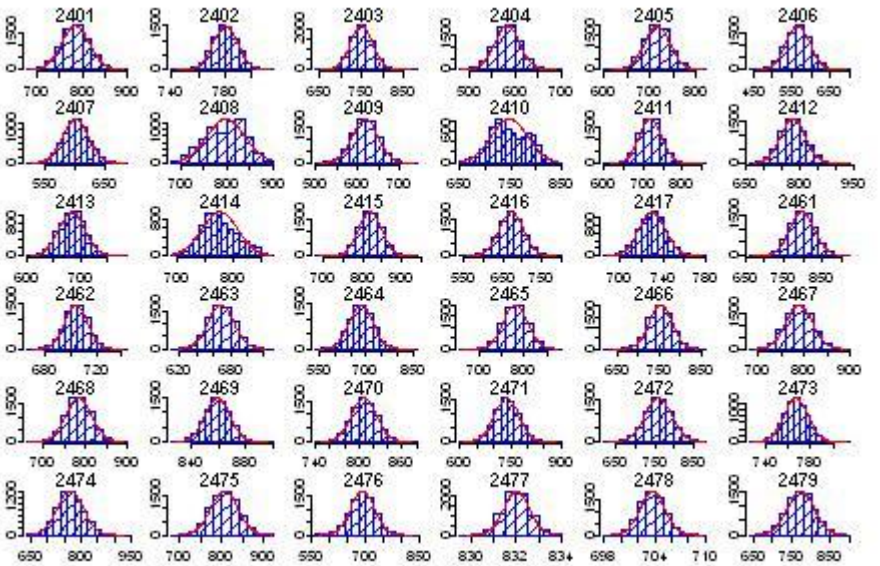

Figure 8. A posteriori distributions of per capita available income for counties in the Silesian region obtained by MCMC simulation (Gibbs sampler) under HB-SAR model.

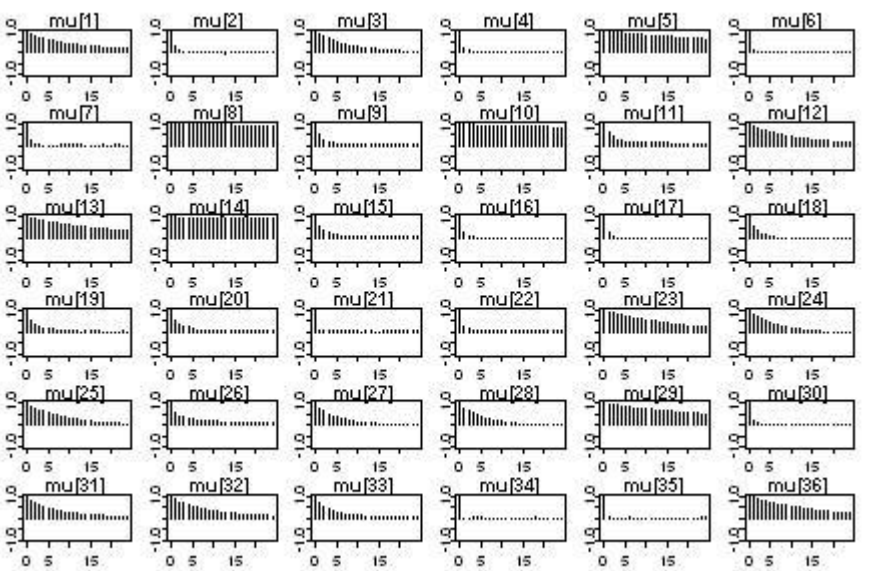

Figure 9. Autocorrelations of model estimates for per capita available income obtained for counties in the Silesian region by MCMC with HB-SAR simulation using Gibbs sampler.

Table 3. Relative estimation error reduction and spatial gain for estimation errors calculated for available income estimates using EBLUP (both ordinary and spatial) method using REML technique and HB (both ordinary and SAR version) using Gibbs sampler.

County (NUTS-4 unit)	Relative estimation error reduction				Spatial estimation error reduction	
	EBLUP	HB	Spatial EBLUP	HB-SAR	Spatial EBLUP	HB-SAR
będziński	1.361	1.322	1.476	1.504	1.084	1.137
bielski	1.013	1.026	1.015	1.049	1.002	1.023
cieszyński	1.273	1.257	1.422	1.420	1.117	1.130
częstochowski	1.307	1.286	1.379	1.396	1.055	1.086
gliwicki	2.145	1.987	2.448	2.869	1.141	1.444
kłobucki	1.430	1.386	1.488	1.482	1.041	1.069
lubliniecki	1.140	1.138	1.156	1.167	1.013	1.025
mikołowski	3.159	2.924	3.737	3.074	1.183	1.051
myszkowski	1.446	1.429	1.563	1.583	1.081	1.108
pszczyński	2.242	2.097	2.698	2.541	1.203	1.211
raciborski	1.161	1.144	1.125	1.155	0.969	1.010
rybnicki	1.746	1.610	1.950	1.886	1.116	1.172
tarnogórski	1.294	1.272	1.385	1.423	1.071	1.118
bieruńsko-lędziński	2.374	2.199	2.871	2.867	1.209	1.303
wodzisławski	1.199	1.156	1.221	1.190	1.019	1.029
zawierciański	1.506	1.454	1.614	1.588	1.072	1.092
żywiecki	1.024	1.024	1.028	1.065	1.004	1.040
Bielsko-Biała city	1.669	1.582	1.864	1.829	1.117	1.156
Bytom city	1.014	1.021	1.011	1.019	0.997	0.998
Chorzów city	1.066	1.051	1.081	1.093	1.014	1.040
Częstochowa city	2.022	1.858	1.949	1.908	0.964	1.027
Dąbrowa Górnicza city	1.263	1.262	1.261	1.306	0.999	1.035
Gliwice city	1.407	1.388	1.456	1.543	1.035	1.111
Jastrzębie-Zdrój city	1.417	1.375	1.504	1.531	1.062	1.113
Jaworzno city	1.624	1.550	1.767	1.749	1.088	1.128
Katowice city	1.014	1.014	1.010	1.053	0.996	1.039
Mysłowice city	1.077	1.095	1.085	1.117	1.007	1.020
Piekary Śląskie city	2.473	2.234	2.921	2.706	1.181	1.211
RudaŚląska city	3.035	2.689	3.602	3.359	1.187	1.250
Rybnik city	1.027	1.029	1.032	1.046	1.005	1.017
Siemianowice Śląskie city	1.857	1.626	2.103	1.831	1.132	1.126
Sosnowiec city	1.693	1.598	1.932	1.858	1.141	1.162
Świętochłowice city	1.844	1.710	2.000	1.897	1.085	1.109
Tychy city	1.000	1.012	1.000	0.994	1.000	0.983
Zabrze city	1.000	0.991	1.000	0.991	1.000	1.001
Żory city	1.811	1.682	1.959	1.845	1.082	1.097

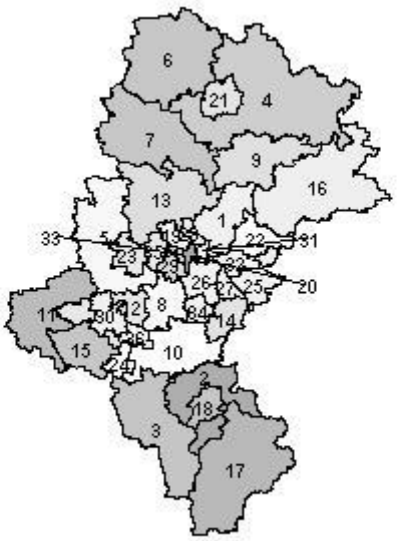

Figure 10. Choropleth map of counties in the Silesian region presenting the absolute values of random effects obtained for per capita available income estimated by Spatial EBLUP estimator (more intense colour means higher absolute random effect).

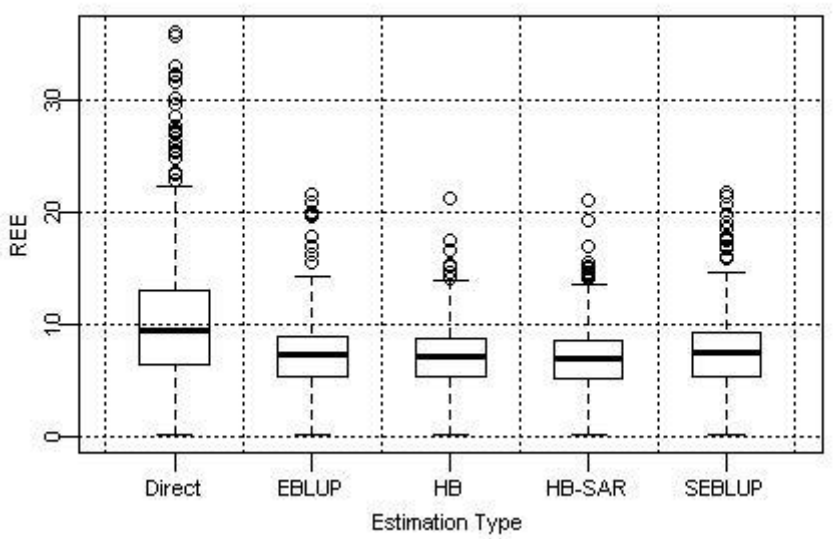

Figure 11. Distribution of relative estimation error for direct estimator, EBLUP (both ordinary and spatial) and HB estimator (ordinary and using SAR relationships) for counties in Poland.

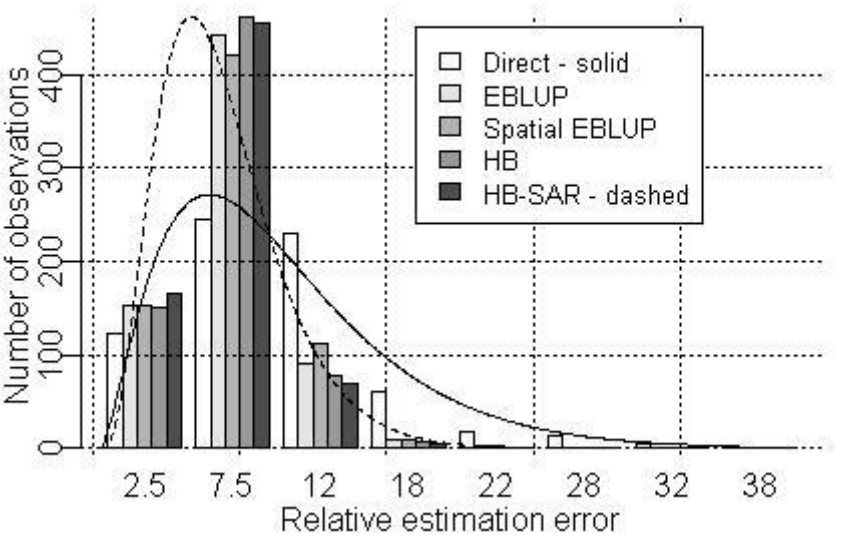

Figure 12. Distribution of relative estimation error for direct estimator, EBLUP (both ordinary and spatial) and for HB estimator (ordinary and using SAR relationships) of per capita available income by NUTS4 in Poland.

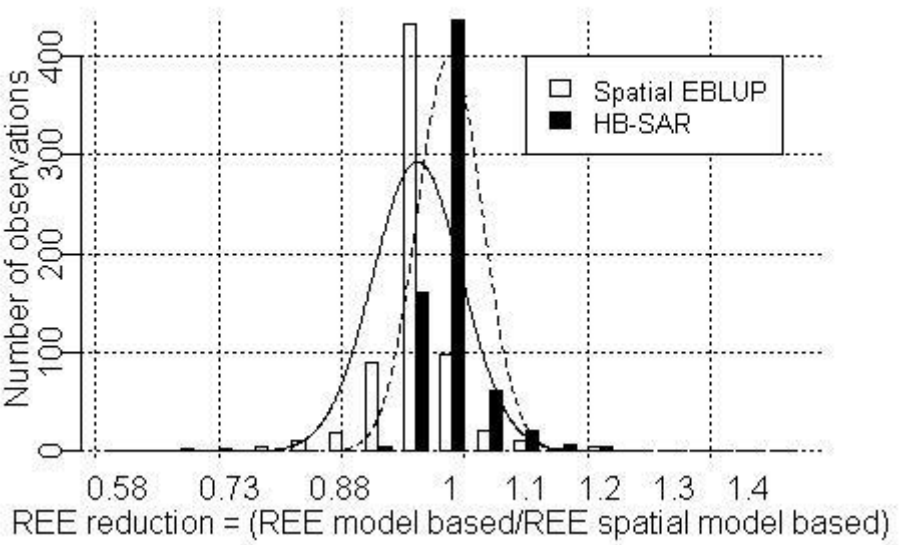

Figure 13. Distribution of relative estimation error reduction due to spatial relationships for EBLUP, Spatial EBLUP and HB (ordinary and using SAR relationships) obtained for per capita available income by NUTS4 in Poland.

This comparison reveals that HB-SAR technique has slightly better performance than its corresponding Spatial EBLUP estimator. However, it should be noted here that only some of the considered models have large enough parameter of spatial autoregression ρ. For most of the regions this measure is below 0.5 and sometimes the ρ coefficient is negative, which may mean that the REE reduction due to the spatial relationships may be not very significant. In order to verify this assumption, we conduct the simulation study, which has to resolve how the ρ value affects the efficiency of spatial estimation. The simulation was prepared in such a way that only MSE values were changed according to the a priori ρ value. The rest of the estimation process, i.e. point estimation procedure for Spatial EBLUP, spatial weight matrix, direct estimates and explanatory variables, remains unchanged. This is in contrast with the experiment conducted in SAMPLE project, where some arbitrary assumptions concerning direct estimates (with fixed values of direct estimation precision), the fixed value of σ_u^2 and explanatory variables were made. In our opinion this may slightly change the real-world conditions and may affect such simulation results.
It is assumed in our experiment that ρ value has four a priori values equal to 0.95, 0.75, 0.25 and -0.50. In our case we also observe that setting the ρ parameter can improve the performance of the estimates. However, that was clear only for SAR-based HB estimator. For spatial EBLUP technique the influence of ρ value is ambiguous (see Figure 14). This is evident when the analysis of REE reduction due to the spatial relationships is made (see Figure 15). Here, for higher ρ values some reduction of REE values obtained for spatial version of HB estimator is observed. For Spatial EBLUP this reduction is rather not the rule (the average spatial REE reduction is slightly below 1). Similar results were also obtained in the work published recently by Gharde, Rai, Jaggi (2013). The authors reach the conclusion that "there is % gain in efficiency in Spatial HB (SHB) approach with respect to SEBLUP approach". It should also be noted that for lower ρ values this gain obtained in our simulation is not significant, even for HB-SAR method.
The simulation results conducted for HB-SAR method reveals also some interesting properties of the obtained stochastic processes generated by Gibbs sampler. It is related to the level of ρ values. When the ρ value is high, the obtaining process for random effect v reveals a characteristic trace, which reveals the simultaneous nature of this process for all v random effects. Their nature has a typical autoregressive run, which becomes evident when Hurst exponent is determined for such a process (using aggvarFit function from 'fArma' package for R-project environment). When ρ value is equal to 0.95, this is practically the rule that the process of v has an autoregressive nature (with Hurst exponent higher than 0.9), which is in contrast to the trace of the process for random effects u in ordinary HB simulations (where Hurst exponents for most cases are considerably lower – see Fig. 16).

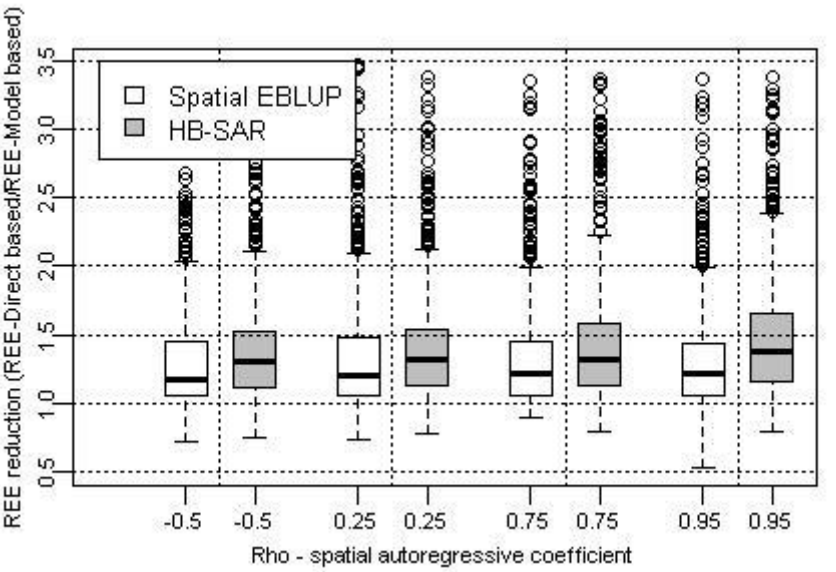

Figure 14. Relative estimation error reduction for Spatial EBLUP and HB-SAR estimators of per capita available income in Poland by counties, for different a priori spatial autoregressive coefficients.

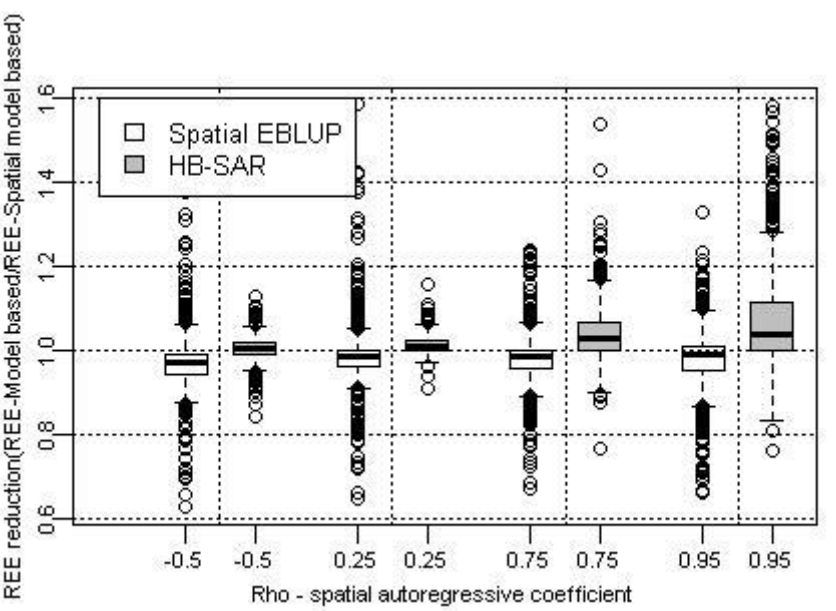

Figure 15. Relative estimation error reduction due to spatial relationships for Spatial EBLUP and HB-SAR estimators of per capita available income in counties, for different a priori spatial autoregressive coefficients.

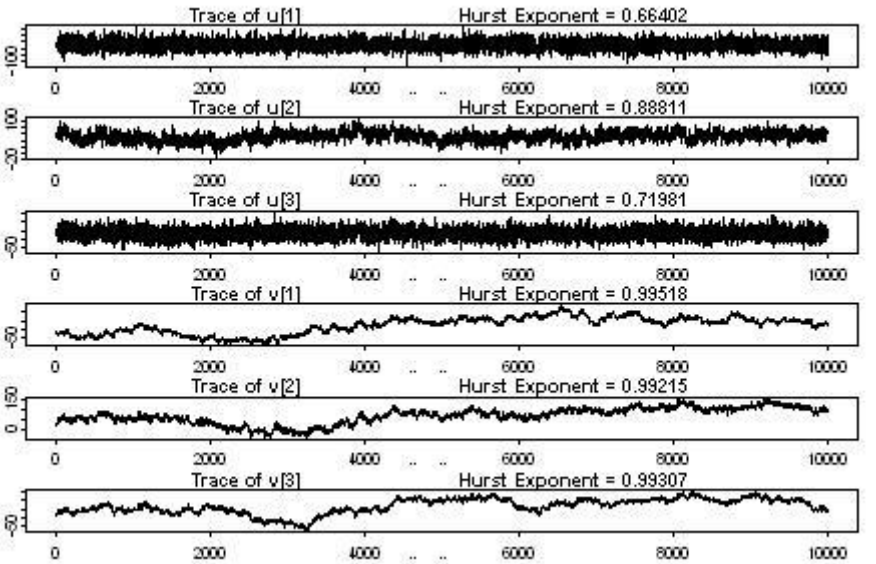

Figure 16. Trace of results for Gibbs sampler simulation obtained for first 3 values of random effects (u description) obtained for ordinary HB method and first 3 values of random effects (v description) obtained for HB method (using SAR relationships) assuming (for v values) that the spatial autoregressive coefficient is equal to 0.95.

From these results, it can be concluded that the Gibbs sampler reproduces the simultaneous nature of the SAR process in a proper way. However, it is still questionable whether the Markov Chain inference can be done in such situations (mainly because of autocorrelation and lack of stability). To overcome this difficulty, some other simulation techniques, like that shown in De Oliveira, V., Jin Song, J., (2008), would be advisable. A non-iterative Monte Carlo algorithm based on factoring the posterior distribution and the adaptive rejection Metropolis sampling (ARMS) proposed by Gilks, Best and Tan (1995) can also be a reasonable choice. The comparison of these two techniques may reveal whether the MCMC approach is valid for the SAR-based simulation conditions. It seems that the Gibbs sampler can be a good starting point for obtaining such simulations. Moreover, for lower ρ values, the autoregressive nature of the process is rather small, and because of this it can be useful in practice for moderate ρ values, as it was shown for the Silesian region in our paper.

5. Conclusions

The paper shows a procedure of efficient estimation for small areas based on the application of the hierarchical Bayes approach to the general linear mixed model with spatially correlated random effects. In particular, the spatial Simultaneous Autoregressive Process, using spatial neighbourhood as auxiliary information, was incorporated into the estimation process. The efficiency of the proposed method was proven on the basis of real-world examples prepared for the Polish data coming from the Household Budget Survey and the tax register. The comparison of relative estimation error distribution and REE reduction shows that all the considered model-based techniques are significantly more efficient than the direct estimation one, however HB-SAR technique shows slightly more REE reduction than the other model techniques. The simulation-based calculations, where some additional assumptions on the spatial autoregressive coefficient were made, also confirm efficiency gains for spatial-based estimators, especially for higher values of this coefficient. However, such a correspondence does not always occur for all the regions, so one should be conscious that for lower ρ values the benefit of using the spatial method may be ambiguous. However, this effect is more evident for Spatial HB method than for Spatial EBLUP technique.

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.

GHARDE, Y., RAI, A., JAGGI, S., (2013). Bayesian Prediction in Spatial Small Area Models, Journal of the Indian Society of Agricultural Statistics, 67, pp. 355–362.

GILKS, W. R., BEST, N. G., TAN, K. K. C., (1995). Adaptive Rejection Metropolis Sampling within Gibbs Sampling, Applied Statistics, 4, pp. 455–472.

GOMEZ-RUBIO, V., (2008). Tutorial: Small Area Estimation with R The R User Conference 2008, August 12-14, Technische Universitat Dortmund, Germany.

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

KUBACKI, J. (2012). Estimation of parameters for small areas using hierarchical Bayes method in the case of known model hyperparameters, Statistics in Transition-new series, 13, No. 2, pp. 261–278.

KUBACKI, J., JĘDRZEJCZAK, A., (2012). The Comparison of Generalized Variance Function with Other Methods of Precision Estimation for Polish Household Budget Survey, Studia Ekonomiczne, 120, pp. 58–69.

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

DE OLIVEIRA, V., JOON, JIN SONG, (2008). Bayesian Analysis of Simultaneous Autoregressive Models, Sankhya, 70-B, No. 2, pp. 323–350.

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.

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., (2003). Small Area Estimation, John Wiley & Sons, http://books.google.pl/books?id=f8NY6M-5EEwC.

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/.

SALVATI, N., GIUSTI, C., MARCHETTI, S., PRATESI, M., TZAVIDIS, N., MOLINA, I., MORALES, D., ESTEBAN, M. D., SANTAMARIA, L., MARHUENDA, Y., PEREZ, A., PAGLIARELLA, M., CHAMBERS, R., RAO, J. N. K., FERRETTI, C., (2011). Software on small area estimation, Deliverable 22, http://sample-project.eu/images/stories/deliverables/d22.pdf.

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

SPIEGELHALTER, D. J., THOMAS, A., BEST, N., LUNN, D., (2003). WinBUGS User Manual, Version 1.4, http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/manual14.pdf.

STURTZ, S., LIGGES, U., GELMAN, A., (2005). R2WinBUGS: A Package for Running WinBUGS from R, Journal of Statistical Software, 12, No. 3.

VOGT, M., (2010). Bayesian Spatial Modeling: Propriety and Applications to Small Area Estimation with Focus on the German Census 2011, Universitat Trier.

APPENDIX

Implementations of computations for EBLUP and hierarchical models

At the computation stage, the WinBUGS (Spiegelhalter et.al. (2003)) and R-project software were used, including the modules 'sae' (Molina and Marhuenda (2013)), 'R2WinBUGS' (Sturtz, Ligges and Gelman (2005)), 'coda', 'maptools' (Bivand and Lewin-Koh (2013)), 'spdep' (Bivand and Piras (2015)) and 'MASS'.

The computation scheme applied to obtain the normal and Spatial HB estimates for counties in Poland is the following:

```
model
{for(p in 1 : N)
  {Y[p] ~ dnorm(mu[p], tau[p])
   mu[p] <- alpha[1] + alpha[2] * A[p] + alpha[3] * B[p] + u[p]
   u[p] ~ dnorm(0, precu)   }
 precu ~ dgamma (a0,b0)
 alpha[1] ~ dflat()
 alpha[2] ~ dflat()
 alpha[3] ~ dflat()
 sigmau<-1/precu }
```

For SAR version of hierarchical models the following scheme is used:

```
model
{for(p in 1 : N)
  {Y[p] ~ dnorm(mu[p], tau[p])
   mu[p] <- alpha[1] + alpha[2] * A[p] + alpha[3] * B[p] + v[p]
   u[p] ~ dnorm(0, precu)
   v[p] <- inprod(rho_w[p,1:N],u[1:N])   }
 precu ~ dgamma (a0,b0)
 alpha[1] ~ dflat()
 alpha[2] ~ dflat()
 alpha[3] ~ dflat()
 sigmau<-1/precu }
```

In the notation presented above, the symbol Y[p] stands for the direct estimates while tau[p] is their estimation error; the values of A[p] to B[p] are assumed as observed explanatory variables specified for the regression model. The parameters a0 and b0 come from the empirical distribution of model errors for EBLUP (ordinary or spatial), while alphas denote the linear regression coefficients. It should be stressed that the random effects v[p] are linked with the spatial weight matrix W and the values of ρ by the inprod WinBUGS function. It uses rho_w matrix passed to the WinBUGS by a special macro prepared in R-

project environment. Moreover, a special authors' macro for R-project was prepared, which was used as a connector with data input, performing necessary computations and automatic visualization of results (by means of 'coda' module). This macro has a (simplified) form given by the code presented below:

```
# fitting the Spatial EBLUP model
resultSREML <- eblupSFH(Y ~ 1 + A + B, desvar, W, method="REML", data=d, MAXITER=1500)
mseSREML <- mseSFH(Y ~ 1 + A + B, desvar, W, method="REML", data=d, MAXITER=1500)
sigmas_reml <- resultSREML$fit$refvar
rho_REML <- resultSREML$fit$spatialcorr
# determining the model parameters
I <-diag(1,N)
for (j in 1:N) {
  W_row <- W[j,]
  for (k in 1:lpow) {
    W_mat[j,k] <- W_row[k]
  }}
rho_W <- solve(I-rho_REML*W_mat)
a0 <- dochg_shape_Sp
b0 <- dochg_rate_Sp
infile <- "coda1.txt"
indfile <- "codaindex.txt"
data <- list(N=N, Y=Y, tau=tau, A=A, B=B, a0=a0, b0=b0, rho_w=rho_W)
model <- lm(Y ~ 1 + A + B)
mod_smry <- summary(model)
alpha <- as.vector(mod_smry$coefficients[,1])
sigma_2 <- (mod_smry$sigma)*(mod_smry$sigma)
precu <- 1/sigma_2
v <- vector(mode = "numeric", length = N)
u <- vector(mode = "numeric", length = N)
inits <- list(list(alpha=alpha, precu=precu, u=u))
parameters <- c("mu", "alpha", "precu", "v", "u")
working.directory <- getwd()
# simulations - WinBUGS call and collecting the data
sim_HB <- bugs(data, inits, parameters, model_HB,n.chains=1, n.burnin = 1, n.iter=10000, n.thin = 1, codaPkg=TRUE, working.directory = working.directory)
 results <- read.coda(infile, indfile, 2, 10000, 1)
```

The code includes (for clarity of expression) only the sections that present how the model parameters were determined and how the simulations were run - with WinBUGS call. The rest of the code has a more ordered character and includes the processes of loading the necessary packages (RODBC, sae, R2WinBUGS, maptools, spdep and MASS), setting the gamma parameters for σ_u^2 (fitdistr function is called here), reading the input data for particular region (functions from RODBC package were used and functions from 'maptools' and 'spdep' for digital maps were applied here), fitting the EBLUP model (ordinary and spatial version) using 'sae' package (eblupFH, mseFH, eblupSFH and mseSFH functions are used here) and – after completing the simulations in WinBUGS – arranging the results and estimating the mean and variance (previously using read.coda function) as well as saving the results to the file (standard cat and format functions were used here).