Small Area Estimation of Income Using Spatio-Temporal Models

University of Łódź
Alina Jędrzejczak

Statistical Office in Łódź
Jan Kubacki

11.07.2018 Warsaw, II Congress of Polish Statistics

Summary

In the presentation the comparison of estimation results for spatial and spatio-temporal small area model is presented. The analysis was conducted for income-related variables coming from the Polish Household Budget Survey and explanatory variables coming from the Polish Local Data Bank. The properties of EBLUPs (Empirical Best Linear Unbiased Predictors) based on spatio-temporal models, which utilize spatial correlation between neighbouring areas and historical data, were compared and contrasted with the 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, eblupSTFH functions were used for point estimation together with mseFH, mseSFH and pbmseSTFH functions for MSE estimation. In the case of sae2 package eblupRY function was applied. The precision of direct estimators was determined using the Balanced Repeated Replication method. The results of the analysis indicate that for the implemented spatio-temporal small-area models visible estimation error reduction was achieved, especially when significant space and time autocorrelations have been observed. The results are even slightly better than those achieved by means of the Rao-Yu model. In the computations three author-defined functions were used, which allowed to perform the extract of random effects for spatial, spatio-temporal and Rao-Yu models and made it possible to obtain their decomposition with respect to spatial and temporal parts, what indicates the novelty of the paper. This comparison was carried out using choropleth maps for spatial effects and distributions of temporal random effects for considered years.

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 the inferences are connected with larger 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.

One of these techniques is the spatio-temporal EBLUP technique presented by Marhuenda, Molina and Morales (2013). It is based on the assumption that the spatial relationships between domains can be modelled by sum of two components:

- the simultaneous autoregressive process SAR
- time-related process described by AR(1) scheme.

Moreover, when an evident correlation exists between survey and administrative data, also the bias of the estimates can be reduced.

Objectives

In the presentation we compare several approaches to the spatial and spatio-temporal modelling implemented for small area estimation. In our opinion, spatio-temporal estimation can be useful with respect to the traditional EBLUP approach. Better efficiency of such models is expected due to taking into account spatial and time-related dependencies between domains. The models using time-related dependencies can additionally be helpful in the analysis of the dynamics of the observed phenomena, what can be supplementary related to the econometric models, including the panel models (Jędrzejczak, Kubacki (2016)).

Small area estimation using spatio-temporal Fay-Herriot model

The methodology of spatio-temporal small area model was described in Marhuenda, Molina, and Morales (2013). It assumes, that the area parameter for domain d at current time t is estimated borrowing strength from the T time instants and from the D domains.

Let θ_{dt} represents the variable of interest determined for area d and time t where d=1,…,D, and t=1,…,T. If the direct estimator of this quantity is denoted by $\hat{\theta}_{dt}^{DIR}$, and the sampling errors can be expressed as e_{dt} and it is assumed that are independent and normally distributed with known variances , the spatio-temporal model can be written as below.

$$\hat{\theta}_{dt}^{DIR} = \theta_{dt} + e_{dt}$$

The above relationship is valid for all considered *d* and *t*. This equation can be expressed also using the model which incorporated the spatio-temporal relationships.

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

Random effects in spatio-temporal Fay-Herriot model

Here $\mathbf{X}_{dt}$ represents the vectors of p auxiliary variables, dependent linearly with ϑ_{dt} with regression coefficients expressed as $\boldsymbol{\beta}$. The area-time random effects can be expressed by $(u_{2d1}, \dots, u_{2dT})^T$ and it is assumed that they are identically and independently distributed for each area, and is subject to the AR(1) process with autocorrelation parameter ρ_2, that can be described as.

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

The area-related random effects can be expressed by $(u_{11}, \dots, u_{1D})^T$ and is subject to the SAR process with variance parameter ${\sigma_1}^2$, spatial autocorrelation ρ_1 and proximity matrix
$\mathbf{W} = (w_{d,l})$, which can be obtained from an original proximity matrix $\mathbf{W}^0$, whose diagonal elements area 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 neighbor 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

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

Spatio-temporal model in terms of general mixed model

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

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

$$\boldsymbol{e} = \underset{1 \le d \le D}{col}\left(\underset{1 \le t \le T}{col}\left(e_{dt}\right)\right) \quad \boldsymbol{u_1} = \underset{1 \le d \le D}{col}\left(u_{1d}\right) \quad \boldsymbol{u_2} = \underset{1 \le d \le D}{col}\left(\underset{1 \le t \le T}{col}\left(u_{2dt}\right)\right)$$

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 ones and has length T, and $\otimes$ 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 one can describe the general mixed model as follows

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

$\boldsymbol{\delta} = (\sigma_1^2, \rho_1, \sigma_2^2, \rho_2)$ is the vector defined in terms of model variance components used in the model above. We can also use the following relationships for vector **e** related with 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 $\psi = diag_{1 \le d \le D}(diag_{1 \le t \le T}(\psi_{dt}))$.

Covariance matrix of spatio-temporal model

Random effects can satisfy the following relationships, where the covariance matrix for the model **G** is used. It can be expressed as $\mathbf{u} \sim N\{\mathbf{0}_n, \mathbf{G}(\boldsymbol{\delta})\}$, and **G** can be expressed as the block diagonal matrix, that has the following form

$$\mathbf{G}(\boldsymbol{\delta}) = diag\{\sigma_1^2 \boldsymbol{\Omega}_\mathbf{1}(\rho_1), \sigma_2^2 \boldsymbol{\Omega}_\mathbf{2}(\rho_2)\}.$$

Here we have the following relationships for matrices $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$.

$$\boldsymbol{\Omega}_1(\rho_1) = \{(\mathbf{I}_D - \rho_1 \mathbf{W})^T (\mathbf{I}_D - \rho_1 \mathbf{W})\}^{-1}$$
$$\boldsymbol{\Omega}_2(\rho_2) = diag_{1 \le d \le D}\{\boldsymbol{\Omega}_{2d}(\rho_2)\}$$

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

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

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

BLUP estimator for spatio-temporal Fay-Herriot model

The vector $\boldsymbol{\beta}$ and the random effects $\mathbf{u}$ can be obtained using BLUP estimator $\widetilde{\boldsymbol{\beta}}(\boldsymbol{\delta})$ and the following equations, that uses **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

$$\widetilde{\boldsymbol{u}}_{\mathbf{1}}(\boldsymbol{\delta}) = \sigma_1^2\boldsymbol{\Omega}_1(\rho_1)\mathbf{Z}_1^T\mathbf{V}^{-1}(\boldsymbol{\delta})\{\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\delta})\}$$
$$\widetilde{\mathbf{u}}_2(\boldsymbol{\delta}) = \sigma_2^2\boldsymbol{\Omega}_2(\rho_2)\mathbf{V}^{-1}(\boldsymbol{\delta})\{\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\delta})\}$$

REML estimation method for spatio-temporal model

The Restricted Maximum Likelihood (REML) method uses maximization method for likelihood function, which corresponds to the joint probability density function as a vector of n-p linearly independent contrasts expressed as $\mathbf{F}^{\mathbf{T}}\mathbf{y}$ where **F** is the n×(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 maximize the log likelihood function $\ell_R(\boldsymbol{\delta};\mathbf{y})$=log$f_R(\boldsymbol{\delta};\mathbf{y})$ using Fisher scoring algorithm. In this algorithm scoring vectors that has the form $S_R(\boldsymbol{\delta})$=$\partial\ell_R(\boldsymbol{\delta};\mathbf{y})/\partial\boldsymbol{\delta}$ and the Fisher information matrix that has the form

$$\mathfrak{I}_R(\boldsymbol{\delta}) = -E\left\{\frac{\partial^2\ell_R(\delta;y)}{\partial\delta\partial\delta'}\right\} = (\mathfrak{I}_{rs}^R(\boldsymbol{\delta}))$$

Fisher scoring algorithm for spatio-temporal model

The first order derivative of $\ell_R(\boldsymbol{\delta};\mathbf{y})$, with respect of δ_r can be given as below.

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

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

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

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 included in sae package. This method can be summarized as follows:

1. Using the available data $\{(\hat{\theta}_{dt}^{DIR}, x_{dt})$, t=1,..,T, d=1,...,D} obtain the estimates of the STFH model and obtain model parameter estimates for **β** and **δ**.

2.Generate bootstrap area effects $\{u_{1d}^{*(b)}$, d=1,...,D } from the SAR process, using $(\hat{\sigma}_1^2, \hat{\rho}_1)$ as true values of (σ_1^2, ρ_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, with acting $(\hat{\sigma}_2^2, \hat{\rho}_2)$ as true values of parameters (σ_2^2, ρ_2)

4. Calculate true bootstrap quantities, using the formula

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

Determining the MSE of spatio-temporal estimates using parametric bootstrap method – 2nd part of algorithm

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 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.

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

Diagnostics of Rao-Yu model and spatio-temporal model of income from social security benefits based on sample and administrative data for 2003-2011 years

Model/ explanatory variable	Coefficient estimates	Standard error	t-statistics	p-value
Rao-Yu model	σ_2^2 = 91.169	σ_1^2 =98.381		ρ_2=0.4586
Intercept	42.7990	15.1360	2.8276	0.0047
Average monthly gross wages and salaries	0.0083	0.0115	0.7164	0.4738
Average retirements pay	0.1648	0.0235	7.0148	0.0000
GDP per capita (Poland 100%)	-0.2901	0.1569	-1.8487	0.0645
Spatio-temporal model	σ_2^2 = 91.046	σ_1^2 =72.305	ρ_1 =0.7991	ρ_2=0.4288
Intercept	52.0950	16.7740	3.1058	0.0019
Average monthly gross wages and salaries	0.0164	0.0116	1.4170	0.1565
Average retirements pay	0.1477	0.0235	6.2917	0.0000
GDP per capita (Poland 100%)	-0.3955	0.1445	-2.7374	0.0062

Estimation results for income from social security benefits by region (direct estimates, ordinary EBLUP, Rao-Yu EBLUP and spatio-temporal estimate) in Poland for 2011 year

Region	Direct estimate		EBLUP estimate		Rao-Yu estimate			Spatio-temporal estimate		
	Value	REE	Value	REE	Value	REE	Time-related random effect	Value	REE	Time-related random effect
	zł	%	zł	%	zł	%	zł	zł	%	Zł
Dolnośląskie	315.43	3.15	314.28	2.71	320.99	2.38	-4.534	321.34	2.52	-5.277
Kujawsko-Pomor.	290.86	4.12	295.48	3.27	285.82	2.92	-1.605	285.53	2.64	-0.753
Lubelskie	293.67	1.58	292.69	1.58	292.60	1.47	-1.980	292.70	1.46	-2.046
Lubuskie	341.59	6.63	305.30	3.95	311.14	3.53	5.759	311.89	3.15	5.322
Łódzkie	325.03	3.46	311.73	2.94	322.08	2.53	8.534	321.14	2.48	8.999
Małopolskie	303.54	2.94	306.04	2.54	302.91	2.36	-4.027	303.04	2.24	-3.915
Mazowieckie	289.62	2.32	289.31	2.41	297.11	1.98	-17.887	296.90	1.81	-16.814
Opolskie	301.51	9.07	314.45	3.78	322.61	3.29	-1.428	323.49	3.41	-2.227
Podkarpackie	286.56	2.56	285.20	2.40	286.09	2.15	2.961	285.92	1.86	3.235
Podlaskie	286.75	0.43	286.84	0.43	286.79	0.43	-4.907	286.79	0.36	-4.303
Pomorskie	278.48	5.70	299.42	3.55	297.46	3.17	-6.712	297.02	3.02	-6.393
Śląskie	390.99	1.43	389.81	1.48	388.78	1.30	5.453	388.66	1.30	6.261
Świętokrzyskie	309.59	4.52	299.08	3.37	299.75	3.03	4.204	300.03	2.83	3.940
Warm.-Mazurskie	273.84	5.94	288.56	3.74	285.54	3.28	-1.284	283.64	3.20	0.139
Wielkopolskie	289.86	2.03	292.30	1.99	291.94	1.81	-5.454	292.23	1.61	-6.752
Zach.-pomorskie	305.48	6.39	308.96	3.59	317.75	3.31	0.218	316.74	3.16	0.316

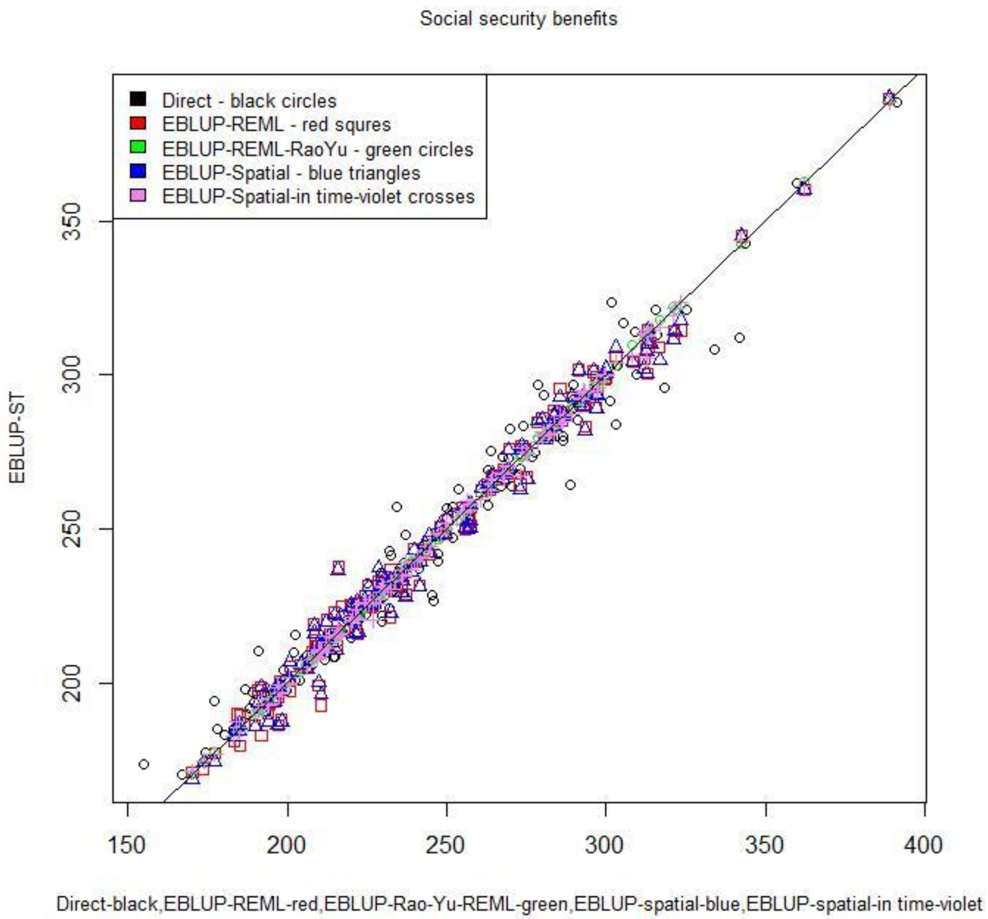
Estimation results for income from social security benefits by region (direct estimates, spatial EBLUP, Rao-Yu EBLUP and spatio-temporal estimate) in Poland for 2011 year

Region	Direct estimate		Spatial EBLUP estimate		Rao-Yu estimate		Spatio-temporal estimate	
	Value	REE	Value	REE	Value	REE	Value	REE
	zł	%	zł	%	zł	%	zł	%
Dolnośląskie	315.43	3.15	314.25	3.01	320.99	2.38	321.34	2.52
Kujawsko-Pomor.	290.86	4.12	292.73	3.58	285.82	2.92	285.53	2.64
Lubelskie	293.67	1.58	292.67	1.66	292.60	1.47	292.70	1.46
Lubuskie	341.59	6.63	302.03	4.46	311.14	3.53	311.89	3.15
Łódzkie	325.03	3.46	311.33	3.21	322.08	2.53	321.14	2.48
Małopolskie	303.54	2.94	308.70	2.75	302.91	2.36	303.04	2.24
Mazowieckie	289.62	2.32	289.29	2.62	297.11	1.98	296.90	1.81
Opolskie	301.51	9.07	317.78	4.14	322.61	3.29	323.49	3.41
Podkarpackie	286.56	2.56	286.35	2.56	286.09	2.15	285.92	1.86
Podlaskie	286.75	0.43	286.82	0.43	286.79	0.43	286.79	0.36
Pomorskie	278.48	5.70	293.79	4.32	297.46	3.17	297.02	3.02
Śląskie	390.99	1.43	389.90	1.55	388.78	1.30	388.66	1.30
Świętokrzyskie	309.59	4.52	302.13	3.51	299.75	3.03	300.03	2.83
Warm.-Mazurskie	273.84	5.94	283.16	4.17	285.54	3.28	283.64	3.20
Wielkopolskie	289.86	2.03	292.84	2.09	291.94	1.81	292.23	1.61
Zach.-pomorskie	305.48	6.39	304.76	4.41	317.75	3.31	316.74	3.16

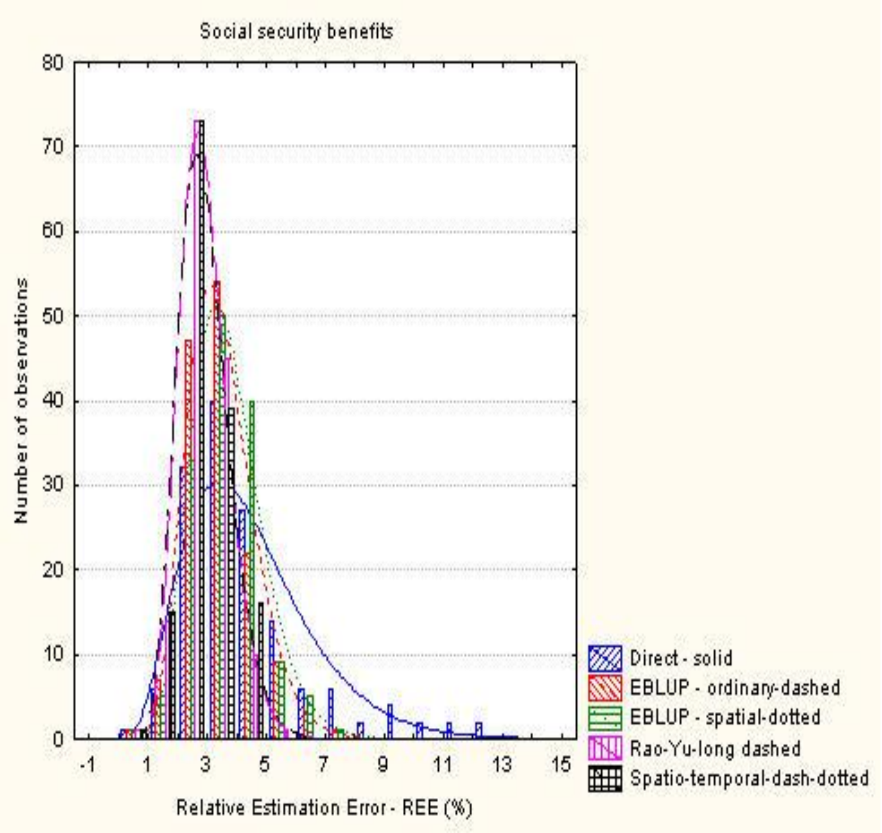
REE reduction estimation results for income from social security benefits by region (ordinary EBLUP, spatial EBLUP, Rao-Yu EBLUP and spatio-temporal estimate) in Poland for 2011 year

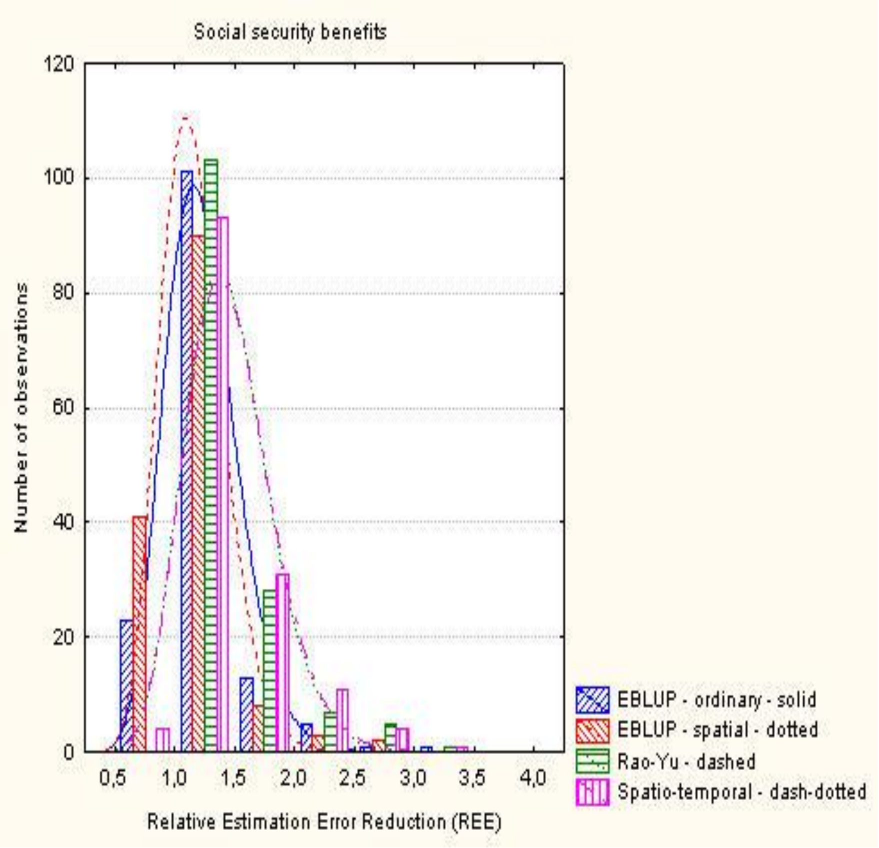
Region	EBLUP	Spatial EBLUP estimate		Rao-Yu estimate		Spatio-temporal estimate	
	REE reduction	REE reduction	Spatial related REE reduction	REE reduction	Time-related REE reduction	REE reduction	Spatio-temporal related REE reduction
Dolnośląskie	1.1607	1.0462	0.9014	1.3225	1.1394	1.2517	1.0784
Kujawsko-Pomor.	1.2614	1.1501	0.9118	1.4116	1.1191	1.5627	1.2389
Lubelskie	0.9988	0.9484	0.9495	1.0760	1.0773	1.0841	1.0854
Lubuskie	1.6803	1.4860	0.8843	1.8792	1.1183	2.1051	1.2528
Łódzkie	1.1774	1.0802	0.9174	1.3711	1.1645	1.3954	1.1851
Małopolskie	1.1566	1.0697	0.9249	1.2449	1.0764	1.3102	1.1328
Mazowieckie	0.9627	0.8869	0.9213	1.1695	1.2149	1.2797	1.3293
Opolskie	2.3983	2.1911	0.9136	2.7535	1.1481	2.6582	1.1084
Podkarpackie	1.0677	1.0000	0.9366	1.1910	1.1155	1.3773	1.2900
Podlaskie	1.0009	0.9963	0.9954	1.0043	1.0034	1.1954	1.1943
Pomorskie	1.6039	1.3200	0.8230	1.7968	1.1202	1.8904	1.1786
Śląskie	0.9649	0.9209	0.9545	1.0976	1.1376	1.0991	1.1391
Świętokrzyskie	1.3398	1.2872	0.9607	1.4916	1.1132	1.5969	1.1919
Warm.-Mazurskie	1.5878	1.4240	0.8969	1.8080	1.1387	1.8541	1.1677
Wielkopolskie	1.0221	0.9718	0.9508	1.1243	1.1000	1.2657	1.2383
Zach.-pomorskie	1.7820	1.4503	0.8139	1.9298	1.0829	2.0238	1.1357

Scatterplot for estimation of results obtained using direct estimates, EBLUP models, spatial EBLUP models, Rao-Yu models and spatio-temporal model for income from social security benefits by region in Poland for 2003-2011 years

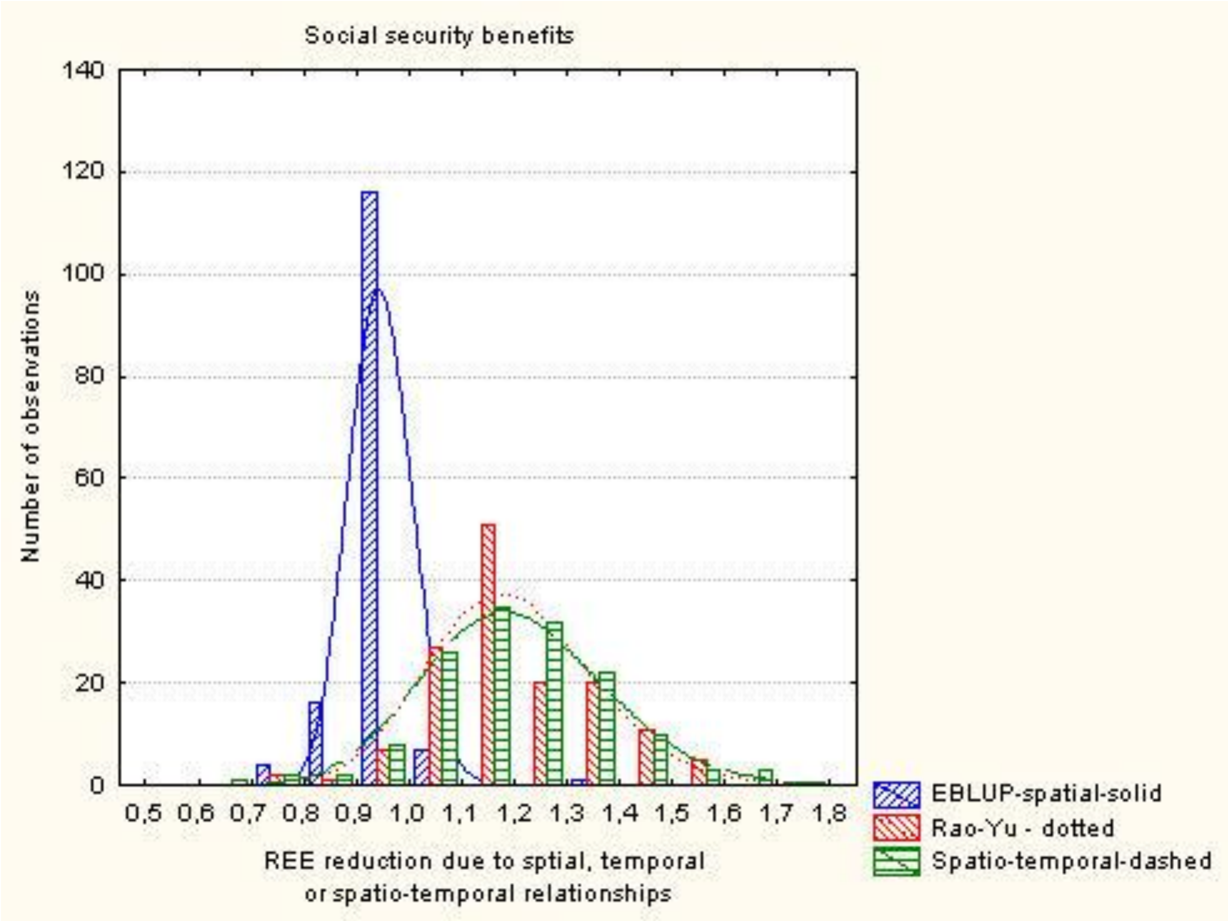


Distribution for estimation of relative estimation error (REE) – left - and REE reduction – right - obtained using direct estimates, EBLUP models, spatial EBLUP models, Rao-Yu models and spatio-temporal model for income from social security benefits by region in Poland for 2003-2011 years



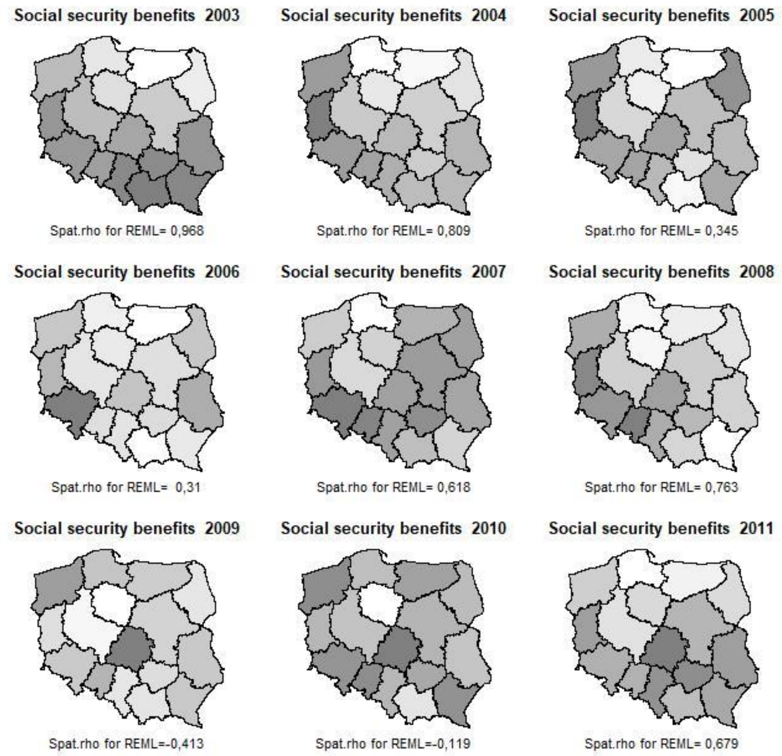


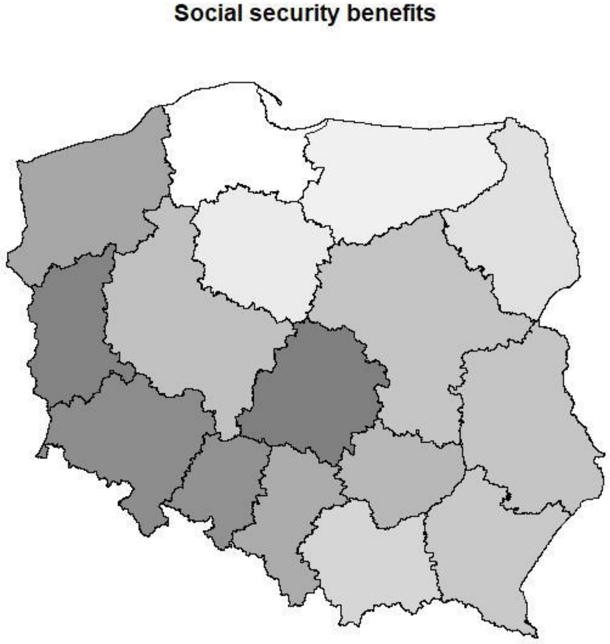
Distribution for estimation of relative estimation error (REE) reduction due to spatial or time or space-time obtained using spatial EBLUP models, Rao-Yu models and spatio-temporal model for income from social security benefits by region in Poland for 2003-2011 years



Statistical Office in Łódź

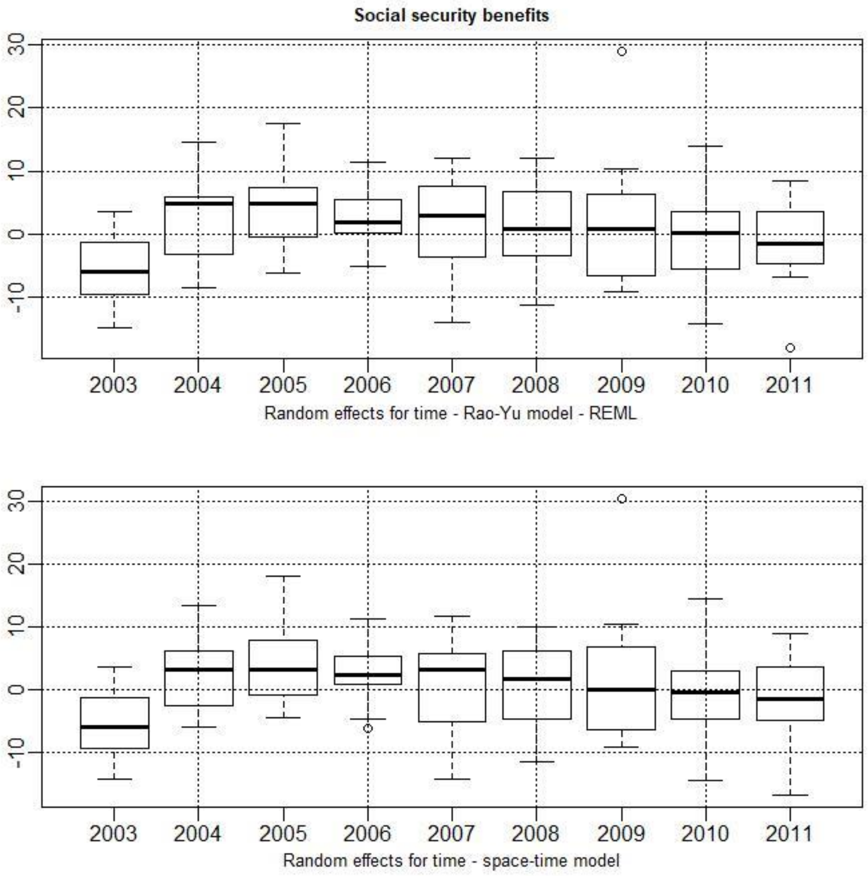
Left - Choropleth maps for estimation of random effects related with space for spatial EBLUP models
Right - Choropleth map for estimation of random effects related with space (u_1) for spatio-temporal model
for income from social security benefits by region in Poland for 2003-2011 years

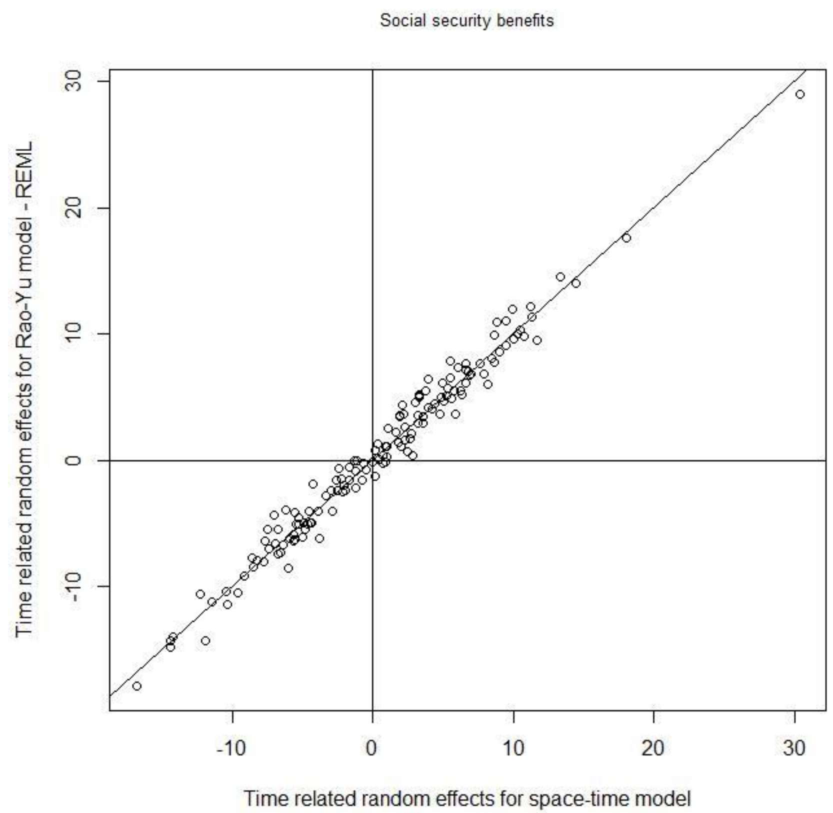




Space related random effects for space-time model - Spat.rho = 0,799

Left - Distributions of random effects related with time (u_2) for Rao-Yu model and spatio-temporal models (top- Rao-Yu, bottom-spatio-temporal) Right - scatterplot for random effects related with time (u_2) for spatio-temporal model and Rao-Yu model obtained for income from social security benefits for 2003-2011 years





Conclusions

The presentation shows a procedure of efficient estimation for small areas based on the application of the spatio-temporal model to the general linear mixed model with spatially correlated random effects and significant correlation with time. In particular, the spatial Simultaneous Autoregressive Process, using spatial neighborhood as auxiliary information and AR(1) process for time-related random error, 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 administrative data. 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 spatio-temporal and Rao-Yu technique shows more REE reduction than the other model techniques. The calculations, where some additional assumptions on the spatial relationships were made, also confirm efficiency gains for spatial-based estimators. However, such a correspondence does not always occur for all the years, so one should be conscious that for lower ρ_2 values the benefit of using the spatial method may be ambiguous.

Conclusions

The presented spatio-temporal model improves the precision of small-area estimates not only in relation to direct estimates, what is easy to obtain, but also in comparison with other indirect techniques based on small-area models, also ordinary spatial EBLUP. That small area model approach, using spatio-temporal EBLUP procedures based on a general linear mixed model, presents a well-known advantage of taking into account the between-area variation beyond that explained by the auxiliary variables included in classical regression models.

Further benefits can be expected when time-dependent nonlinear relationships are taken into account, for example such as GARCH relationships or nonlinear dependence on explanatory variables. Previously conducted analysis of nonlinear models (see Jędrzejczak, Kubacki (2016), Jędrzejczak, Kubacki (2017)) may be a starting point for more detailed comparisons between Rao-Yu method, nonlinear models and econometric panel models.

Literature

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/

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 from 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, 2012

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, 2017, pp. 69-81 http://www.dbc.wroc.pl/Content/37066/Jedrzejczak_Analiza_Rozkladow_Dochodu_Rozporzadzalnego_Wedlug_2017.pdf

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 http://stat.gov.pl/download/gfx/portalinformacyjny/en/defaultlistaplikow/3478/31/1/000_sit_17_3_11_10__2016.pdf

Literature

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, 2012

Marhuenda, Y. Molina, I. and 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 .

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 .

Literature

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 https://ww2.amstat.org/sections/SRMS/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

More information at:
https://www.sciencedirect.com/science/article/pii/S0167947312003222

and at: https://books.google.pl/books?id=_tRZCgAAQBAJ

.