A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 292, 2013

Maciej Beręsewicz[*]

PLS REGRESSION USING SPATIAL WEIGHTS ON THE EXAMPLE OF SPATIAL MODELING SUPPORT FOR POLITICAL PARTIES IN ELECTIONS 2011 TO THE SEJM OF THE REPUBLIC OF POLAND

1. INTRODUCTION

Aim of this paper is to present and evaluate of Partial Least Square Regression model that takes into account spatial autocorrelation. In the second part Partial Least Square (PLS) method is presented and further Spatial PLS Regression is introduced. The third part contains description of dataset used for presenting PLSR method. Dataset used in this section refers to election for Sejm in Wielkopolskie voivodship in 2011 in Poland. The fourth paragraph contains results of application of PLS Regression and comparison to linear models which also takes into account spatial information (spatial lag and error).

Scientists face many problems with data. One is multicollinearity of data, which causes problems during evaluation of econometric models. This issue can be measured using Variance Inflation Factor (VIF) which identify variables that may cause multicollinearity. There are several ways to deal with this problem. One way is to transform correlated variables in order to create new ones. The other is to remove variables from dataset (this is not recommended solution) or use models which takes into account correlation among predictors. Such models can be based factor analysis (or orthogonal vectors) that creates new variables based on orthogonal transformation of input dataset.

Another issue can be latent structure of dataset, which means that in dataset may be variables which are not only highly correlated but also interrelated. Often we deal with not observed variables which cannot be measured directly but using variables form original dataset such as economy, factors affecting unemployment or satisfaction of customers. There are several methods which can extract latent variables. Most are based on decomposition matrix of variables or correlation coefficients such as Factor Analysis, Principal Components Analysis. One particular method is Partial Least Square Regression (PLS Regression) which will be presented in the second part of this paper.

Furthermore, there are situations when we have more variables than observations. This often occurs in social and natural sciences. For example, when we analyze economy in countries in Europe, the number of describing variables

[*] M.A., Poznan University of Economics, Department of Statistics.

countries can be much higher than the number of countries themselves. This is connected with both issues mentioned earlier – correlation and latent structure. Many (correlated) variables can describe latent variables which describes structure of dataset.

In the analysis there is one issue that should be taken into account – spatial factor. It can cause heterogenity of variance. This means that variables can differ between regions (voivodeships, provinces, counties, communities and so on) and it is due to spatial relationships. In such case it is crucial to include spatial factor into further analysis to measure influence of spatial relationship on variable of interest. This problem can be resolved using spatial models. They take into account spatial component to describe analyzed phenomenon. In Suchecki (2010) one can find extensive description of spatial models.

2. SPATIAL PARTIAL LEAST SQUARE REGRESSION

2.1. PLS REGRESSION MODEL

In this section three models will be discussed PLS Regression and its Spatial development. Partial Least Square Regression (further PLSR) which is extension of multiple linear regression taking into account latent structure of dependent and independent variables (only if there is more than one dependent variable).

PLS method was introduced by Herman Wold, Norwegian statistician, whose work mainly concerned mathematical economics and econometrics. In Wold (1985) we can find introduction to Partial Least Square method to compute orthogonal factors in iterative procedure.

There are other works Wold (1966), Wold (1981) concering iterative methods to estimate principal components. Helland (1990) introduced Partial Least Square regression based on Wold's PLS model. In a nutshell, Partial Least Square procedure is as follows:

- take standardized (centred) dataset, both **Y** and **X**,
- in iterative procedure obtain factors,
- perform OLS Regression using obtained factors.

In first step we need to standardize (center) both X dataset, containing dependent variables and Y which stands for independent variable. This means that factors are produced using standardized values of input dataset, not correlation/covariance matrix as in standard Factor /Principal Components Analysis.

In second step we conduct iterative procedure to obtain factors. There are two popular methods for that – Nonlinear Iterative Partial Least Square (NIPALS) and SIMPLS, the first one will be discussed in detail. NIPALS (PLS1) algorithm as mentioned earlier need standardized (centred) data. The procedure is presented below.

2.1.1. NIPALS (PLS1) Algorithm

Let $j=1, \mathbf{X}_1=\mathbf{X}, \mathbf{y}_1=\mathbf{y}, \mathbf{w}_j=\mathbf{X}^{\mathrm{T}}{}_j\mathbf{y}_j / \left\|\mathbf{X}^{\mathrm{T}}{}_j\mathbf{y}_j\right\|$, $\mathbf{t}_j=\mathbf{X}_j\mathbf{w}_{j,}\ \hat{\mathbf{c}}_j=\mathbf{t}^{\mathrm{T}}{}_j\mathbf{y}_j / \mathbf{t}^{\mathrm{T}}{}_j\mathbf{t}_j$, $\mathbf{p}_j=\mathbf{X}^{\mathrm{T}}{}_j\mathbf{t}_j / \mathbf{t}^{\mathrm{T}}{}_j\mathbf{t}_j$, $\mathbf{X}_{j+1}=\mathbf{X}_j-\mathbf{t}_j\mathbf{p}^{\mathrm{T}}{}_j$, $\mathbf{y}_{j+1}=\mathbf{y}_j-\mathbf{t}_j\hat{\mathbf{c}}_j$,
where: $\mathbf{X}_{nxk}$ – matrix of dependent variables, $\mathbf{y}_{nx1}$ – vector of independent variable, n – number of observations, k – number of variables, g – number of factors (det(determined a priori).

If (j=g) STOP else j=j+1. As a result we get: $\mathbf{W}_{\mathrm{kxg}}$, $\mathbf{P}_{\mathrm{kxg}}$, $\mathbf{T}_{\mathrm{nxg}}$ with columns $\mathbf{w}_{\mathrm{j}}$, $\mathbf{p}_{\mathrm{j}}$, $\mathbf{t}_{\mathrm{j}}$ and $\hat{\mathbf{c}}_{gx1}$ with elements $\hat{\mathbf{c}}_j$. Matrix **W** is orthogonal, **T** has orthogonal columns. The PLS Regression model is as follows:

$$\begin{aligned}\hat{\mathbf{Y}} &= \mathbf{T}\hat{\mathbf{c}} = \hat{\mathbf{X}}\mathbf{W}(\mathbf{P}^{\mathrm{T}}\mathbf{W})^{-1} \\ \hat{\mathbf{X}} &= \mathbf{T}\mathbf{P}^{\mathrm{T}} = \sum_{j=1}^{g} t_j p_j^{\mathrm{T}}\end{aligned}, \tag{1}$$

2.2 SPATIAL PLS REGRESSION

Now let focus Spatial PLS regression models that takes into account spatial factors in autoregressive or error form. This model can be easily extended to other spatial (see: Bivand (2008), LeSage (2009) ,Suchecki (2010), Suchecki (2012)). First model concerns autoregressive influence of spatial factor:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}_{\mathrm{pls}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

where: $\mathbf{X}_{\mathrm{pls}}$ is matrix of factors obtained from NIPALS procedure, **W** is weight matrix, **y** is independent variable, $\boldsymbol{\beta}$ is vector of coefficients and ρ is autoregressive parameter and $\boldsymbol{\varepsilon}\sim N(\mathbf{0},\boldsymbol{\sigma}^2\mathbf{I})$.

Second model is Spatial PLS lag model

$$\begin{aligned}\mathbf{y} &= \mathbf{X}_{\mathrm{pls}}\boldsymbol{\beta} + \boldsymbol{\xi} \\ \boldsymbol{\xi} &= \lambda\mathbf{W}\boldsymbol{\xi} + \boldsymbol{\varepsilon}\end{aligned}, \tag{3}$$

where λ is lag parameter, remaining parameters are described as before.

Another is model which takes into account both spatial error and lag (SAC model):

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}_{\mathrm{pls}}\boldsymbol{\beta} + \boldsymbol{\xi}, \tag{4}$$

3. ELECTION FOR SEJM IN WIELKOPOLSKIE VOIVODSHIP IN 2011

In 2011 in Poland conducted elections for Sejm where 5 parties obtained result higher than 5% threshold: Platforma Obywatelska (PO, 39,18%), Prawo i Sprawiedliwość (PiS, 29,89%), Ruch Palikota (RP, 10,02%), Polskie Stronnictwo Ludowe (PSL, 8,36%) and Sojusz Lewicy Demokratycznej (SLD, 8,24%).

Polish media emphasized that Poland is divided into two parts, east which votes for right wing PiS and west which is for central center/left wing party PO. Figures 1 to 5 on show, that in case of PO, PiS and PSL we can observe strong spatial difference of distribution of votes. In case of RP and SLD we cannot precisely indicate if there are any differences in space by only analyzing plots.

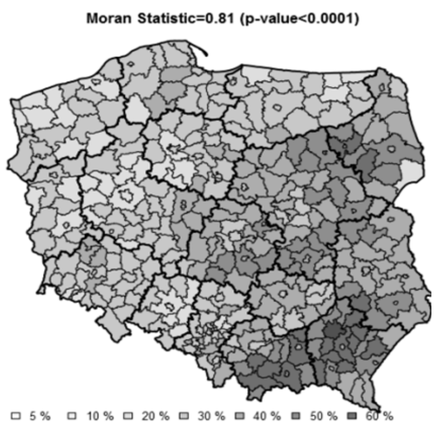

Figure 1. Spatial Distribution of votes for PiS

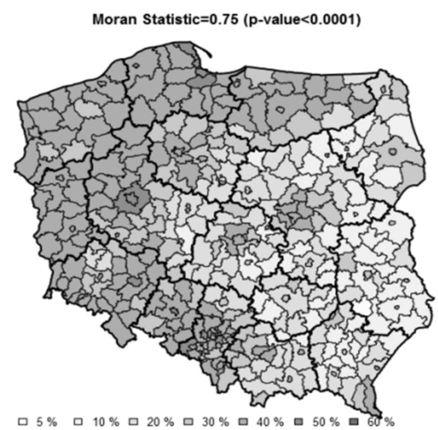

Figure 2. Spatial Distribution of votes for PO

Source: author's own elaboration based on results of elections in Poland 2011.

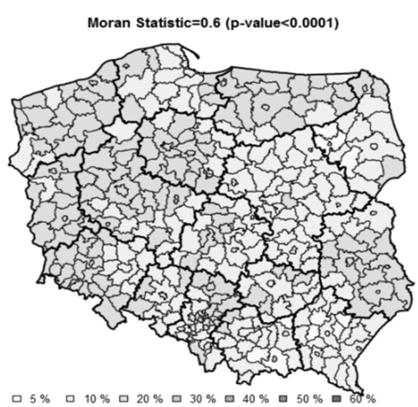

Figure 3. Spatial Distribution of votes for RP

Figure 4. Spatial Distribution of votes for PSL

Source: author's own elaboration based on results of elections in Poland 2011.

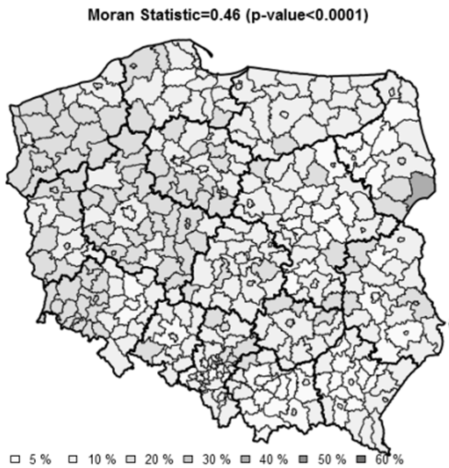

Figure 5. Spatial Distribution of votes for SLD

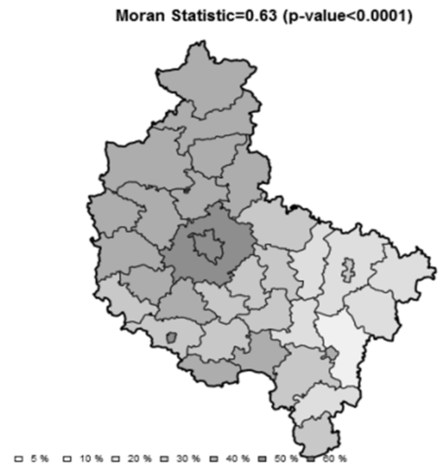

Figure 6. Spatial Distribution of votes for PO in Wielkopolskie voivodship

Source: author's own elaboration based on results of elections in Poland 2011.

In each case Moran I test for global autocorrelation was performed to measure if there is spatial correlation. As fugures show, every statistic is significant and the highest correlation is observed for PO, the lowest autocorrelation is for SLD.

Figure 7. Spatial Distribution of votes for PiS in Wielkopolskie voivodship

Figure 8. Spatial Distribution of votes for RP in Wielkopolskie voivodship

Source: author's own elaboration based on results of elections in Poland 2011.

Wielkopolskie voivodship will be more deeply examined in section 3. Figures 6–10 present results for Moran I test for global autocorrelation and visualization of distribution of votes.

Figure 9. Spatial Distribution of votes for PSL in Wielkopolskie voivodship

Figure 10. Spatial Distribution of votes for SLD in Wielkopolskie voivodship

Source: author's own elaboration based on results of elections in Poland 2011.

Only for Platforma Obywatelska (PO) and Prawo i Sprawiedliwość Moran I statistic is significant at $\alpha=0.05$, this two parties will be concerned in further analysis.

3.2. AUXILIARY DATA OBTAIN FROM LOCAL DATA BANK

In order to use statistical models data was obtained from Polish Statistical Office's Local Data Bank. Dataset contains 37 variables, which were used for creating 4 new ones, concern information about population (i.a. working age), entrepreneurships (i.a. registered in REGON record), job offers, income and spending of local authorities or numbers of nursery schools. Information about attendance for voting was also used for describing share of votes for PO and PiS. All of variables are listed in appendix A.

4. RESULTS

In first step there 2 models were performed – OLS regression for original data and PLS regression. Both were applied for results of votes for PO and PiS. Results are presented in first part of this section. In the second part Lagrange tests for Spatial models were applied to find which model is the best. In the last part models with spatial factor were compared to find out if Spatial PLS regression suits better than Spatial OLS regression for analyzed dataset. All calculations were made in **R** with *sp*, *spdep* packages.

4.1. COMPARISON OF LINEAR MODELS

Table 1 and 2 contains results of estimation for PO. Attendance have positive influence on share of votes and percentage of people after work age have negative influence which is connected with young people who votes for PO.

Table 1. Results of estimation of OLS regression model for Platforma Obywatelska

	Estimate	**Std. Error**	**P-value**
(Intercept)	0.070	0.116	0.5509
Attendance	1.538	0.236	0.0000
PerOfAfterWorkAgePop	– 2.675	0.765	0.0014
Residual standard error: 0.0608 on 32 degrees of freedom. Multiple R-squared: 0.57247, Adjusted R-squared: 0.54575. AIC: – 91.80 BIC: – 85.58 F-statistic: 21.424 on 2 and 32 DF, p-value: 1.246e-06.			

Source: own calculation using **R.**

PLS procedure find 3 significant factors, first is New Enterprises in REGON record, second structure of working age and the last one is Public Sector. All of factor loadings can be found in Appendix B. If we compare these two models, PLS Regression is better in every statistic describing quality of model, R^2 is higher, AIC and BIC are lower also residual error is lower than in OLS Regression.

Table 2. Results of estimation PLS regression model for Platforma Obywatelska

	Estimate	**Std. Error**	**P-value**
(Intercept)	0.382	0. 009	0.0000
NewEnterprisesREGON	0.015	0.002	0.0000
StructureOfWorkingAge	0.024	0.005	0.0001
PublicSector	0.031	0.010	0.0036
Residual standard error: 0.0531 on 31 degrees of freedom Multiple R-squared: 0.68414. Adjusted R-squared: 0.65357 AIC: – 100.39 BIC: – 92.62 F-statistic: 22.381 on 3 and 31 DF, p-value: 6.662e-08			

Source: own calculation using **R.**

Table 3. Results of estimation OLS regression model for Prawo i Sprawiedliwość

	Estimate	Std. Error	P-value
(Intercept)	0.022	0.082	0.7864
PerOfAfterWorkAgePop	1.482	0.571	0.0141
Residual standard error: 0.0459 on 32 degrees of freedom Multiple R-squared: 0.18998. Adjusted R-squared: 0.13935. AIC: – 111.51 BIC: – 105.29 F-statistic: 3.753 on 2 and 32 DF, p-value: 0.03435			

Source: own calculation using **R.**

Table 3 and 4 contains results of estimation for Prawo i Sprawiedliwość, in the first table we can see that only one variable is significant, positive influence on votes for PiS have variable concerning percentage of people after work age (opposite to model for PO). It can be explained by that PiS is concerned as right wing party with promoting conservative and catholic values which goes mainly to older people.

Table 4. Results of estimation PLS regression model for Prawo i Sprawiedliwość

	Estimate	Std. Error	P-value
(Intercept)	0.2281	0.0062	< 2e-16
NonPositiveEnterpriseEnv	– 0.003795	0.001686	0.031612
PerOfNonWorkingAge	– 0.015489	0.00366	0.000191
NonPositiveWorkingEnv	– 0.019528	0.006874	0.007881
Residual standard error: 0.0366 on 31 degrees of freedom Multiple R-squared: 0.50036 . Adjusted R-squared: 0.45201 AIC: – 126.423 BIC: – 118.6463 F-statistic: 10.348 on 3 and 31 DF, p-value: 7.069e-05			

Source: own calculation using **R.**

In case of PLS Regression, 3 factors were obtained from iterative procedure. First can be described as Nonpositive Enterprise Environment which contains variables connected with percentage of enterprises removed from the REGON register due to (mostly) bankruptcy or end disband. Percent of nonworking age concern people which are not in working age (younger and older). The last one concerns nonpositive working environment which is connected with unemployment, low number of job offers and average income.

Comparing two models we can see that PLS Regression is far more better than OLS Regression, R2 is higher, AIC and BIC are lower and residual error is also lower for PLS R.

4.2. LAGRANGE TESTS FOR SPATIAL MODELS

In this paragraph are presented Lagrange tests for spatial models (see: Bivand (2008), LeSage (2009), Suchecki (2010), Suchecki (2012)).

Table 5. Results of Lagrange tests for Platforma Obywatelska

Tested model	Statistic	Statistic value	p-value
OLS Regression model	LMerr	9.1073	0.0025
	LMlag	12.5997	0.0004
	RLMerr	0.8762	0.3493
	RLMlag	4.3686	0.0366
PLS Regression model	LMerr	0.2923	0.5887
	LMlag	5.7484	0.0165
	RLMerr	1.4982	0.2209
	RLMlag	6.9543	0.0084

Source: own calculation using **R.**

As we can see in table 5 for OLS Regression significant statistics those for LM Error and LM Lag, for PLS Regression LM Lag and Robust LM Lag which indicates Spatial Error Model.

Results below relate to testing models for PiS. All test statistics are significant for OLS Regression, for PLS Regression only three of them. Analyze of this p-value levels indicates that the best for OLS Reg is model with spatial autocorrelation, for PLS Reg spatial error model.

Table 6. Results of Lagrange tests for Prawo i Sprawiedliwość

Tested model	Statistic	Statistic value	p-value
OLS Regression model	LMerr	6.4504	0.0111
	LMlag	14.2835	0.0002
	RLMerr	7.2793	0.0070
	RLMlag	15.1124	0.0001
PLS Regression model	LMerr	0.6382	0.4244
	LMlag	9.8465	0.0017
	RLMerr	4.9942	0.0254
	RLMlag	14.2026	0.0002

Source: own calculation using **R.**

4.3. COMPARISON OF SPATIAL MODELS

Next we compare all concerning in paper spatial models. From the analysis of table 7 we can see that including spatial factor for OLS regression improved its results so high that it is better than PLS regression. This means that what PLSR gained in creating orthogonal, factor variables, OLS caught up by include spatial factor.

Table 7. Comparison of estimated models for Platforma Obywatelska

		Df	AIC	LogLik	Test	L.Ratio	p-value
OLS	Linear	4.00	– 91.80	49.90			
	SAR	5.00	– 105.24	57.62	1 vs 2	15.44	0.00
	ERR	5.00	– 107.65	58.83			
	SAC	6.00	– 105.78	58.89	3 vs 4	0.13	0.72

Table 7 (cont.).

		Df	AIC	LogLik	Test	L.Ratio	p-value
PLS	Linear	5.00	– 100.39	55.20			
	SAR	6.00	– 104.15	58.08	1 vs 2	5.76	0.01
	ERR	6.00	– 99.62	55.81			
	SAC	7.00	– 103.90	58.95	3 vs 4	6.28	0.02

Source: own calculation using **R.**

Results for PiS are different because for PLS regression SAC model is the best, which takes into account both spatial error and lag. Other models have higher AIC and lower LogLik values.

Table 8. Comparison of estimated models for Prawo i Sprawiedliwość

		Df	AIC	LogLik	Test	L.Ratio	p-value
OLS	Linear	4.00	– 111.51	59.76			
	SAR	5.00	– 124.76	67.38	1 vs 2	15.25	0.00
	ERR	5.00	– 126.97	68.48			
	SAC	6.00	– 135.74	73.87	3 vs 4	10.77	0.00
PLS	Linear	5.00	– 126.42	68.21			
	SAR	5.00	– 134.35	73.18	1 vs 2	9.9264	0.00
	ERR	6.00	– 127.43	69.72			
	SAC	7.00	– 144.63	79.32	3 vs 4	19.2019	0.00

Source: own calculation using **R.**

To conclude only for PiS, Spatial PLS regression is better than other models. If we include spatial factor in OLS models it gain information which PLS regression obtain for creating factors in iterative procedure.

5. CONCLUSION

Partial Least Square regression is method which can be used in cases when there are more variables than observations, where are correlated structure. Adding spatial factor as spatial lag or error do not improve model as in case of OLS regression. This can mean that PLS Regression get information of spatial associations during iterative procedure or there is other than spatial factor that causes heterogeneity between regions.

One can be connected with population age which vary between regions in Poland, eastern regions of Wielkopolskie voivodship is inhabited mostly by older people and non-working population is higher than in west and center part of this voivodship. This factor should be investigated in further works. What is interesting autocorrelation decrease with diminishing of percentage of support for parties. Another issue is that there should be conducted Monte Carlo simulation to find out real advantages and disadvantages of PLS model. Simulating spatial heterogenity should point if PLS procedure takes some information in iterative procedure.

REFERENCES

Anderson G., Kaufmann P. and Renberg L. (1996), *Non-linear modelling with a coupled neural network-PLS regression system*. J. Chemom. 10.

Hoskuldsson A. (1988), *PLS regression methods*. J. Chemom. 2.

Lorber A., Wangen L. and Kowalski B.R. (1987), *A theoretical foundation for PLS,* J. of Chemometrics 1.

Kopczewska K. (2006), *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa.

Suchecki B. (ed.) (2010), *Ekonometria Przestrzenna Metody i modele analizy danych,* Warszawa, C.H. Beck.

Wold, H. (1966), *Estimation of principal components and related models by iterative least squares,* In P.R. Krishnaiaah (Ed.). Multivariate Analysis. (pp. 391–420) New York: Academic Press.

Wold H. (1981), *The fix-point approach to interdependent systems,* Amsterdam: North Holland.

Wold H. (1985), *Partial least squares,* (in:) Samuel Kotz and Norman L. Johnson (eds.), *Encyclopedia of statistical sciences,* Vol. 6, New York: Wiley, pp. 581–591.

Maciej Beręsewicz

PLS REGRESSION USING SPATIAL WEIGHTS ON THE EXAMPLE OF SPATIAL MODELING SUPPORT FOR POLITICAL PARTIES IN ELECTIONS 2011 TO THE SEJM OF THE REPUBLIC OF POLAND

Space has an important role in the reality around us, especially in the context of socio-economic research. One of the best examples in which the geographic location of one of the most significant factors is the support for political parties. Interesting from the standpoint of policy research is to analyze factors influencing the results of the political party in a particular spatial or administrative unit. The article focuses on the analysis of electoral data for counties.

This was motivated by the high availability of data at a county level, which may be obtained from the Local Data Bank. However, collinearity which occurs in data that affect the support of political parties, limits the use of ordinary linear models. It results in failure of taking into account most of the information contained in the data. In the article will be presented Spatial Partial Least Squares Regression (SPLSR) which takes into account the spatial factor and collinearity.

Author will assess SPLSR model with known spatial linear models with spatial lag and error to compare fit, information criteria and errors. Aim of the article is to show, if taking into account collinearity of predictors significantly improve modelling the support for political parties, which SPLSR model does.

REGRESJA PLS Z UWZGLĘDNIENIEM WAG PRZESTRZENNYCH NA PRZYKŁADZIE MODELOWANIA POPARCIA DLA PARTII POLITYCZNYCH W WYBORACH DO SEJMU 2011 ROKU

Przestrzeń odgrywa ważną rolę w otaczającej nas rzeczywistości, zwłaszcza w kontekście badań społeczno-ekonomicznych. Jednym z przykładów, którym położenie geograficzne badanej jednostki jest znaczące jest poparcie dla partii politycznych, widoczne przy okazji wyborów dokonywanych przez obywateli. Interesujące z punktu widzenia badań politycznych jest analizowanie czynników wpływających na wyniki danej partii politycznej w określonej jednostce przestrzennej czy administracyjnej. W artykule skupiono się na analizie danych wyborczych w ujęciu powiatów.

Podyktowane było to dużą dostępnością informacji na stosunkowo niskim poziomie agregacji przestrzennej, które można pozyskać z Banku Danych Lokalnych GUS.

Występująca współliniowość zmiennych, które wpływają na poparcie partii politycznych ogranicza jednak stosowanie zwykłych modeli liniowych co skutkuje nieuwzględnieniem części informacji. Metodą, która pozwala na uwzględnienie współliniowości jest regresja PLS (*Partial Least Squares Regression*), która nie była wcześniej proponowana w modelowaniu przestrzennym. W artykule zostanie zaprezentowany model Przestrzennej Regresji Metodą Cząstkową Najmniejszych Kwadratów (SPLSR) uwzględniający czynnik przestrzenny. Następnie dokonana zostanie analiza porównawcza SPLSR z modelami klasycznej regresji liniowej uwzględniającej czynnik przestrzenny.

Celem artykułu jest ocena modelu SPLSR w badaniach społeczno-ekonomicznych na przykładzie modelowania poparcia dla partii politycznych. Obliczenia zostaną wykonane w programie R z wykorzystaniem pakietów pls, sp, spdep, maps.

APPENDIX A

Table 9. Variable names

x01	Total Population
x02	The Working Age Population
x03	The Population Of Working Age
x04	Total Marriages
x05	Jobs offers
x06	The Unemployment Rate
x07	Average Monthly Gross Wynagordzenie
x08	Properties For 1000 Population
x09	Total Nursery
x10	Nursery – places - Total
x12	Participation Of Persons In Households Benefiting From Social Environment In The Total Population
x13	Entities Of The National Economy In General
x14	The Types Of Economic Entities, The Public Sector - In Total
x15	The Types Of Economic Entities: The Private Sector - In Total
x16	Number of Entreprises
x17	Newly Registered Entreprises From REGON register - Total
x18	Newly Registered Entreprises From REGON register, The Public Sector - In Total
x19	Newly Registered Entreprises From REGON register Private Sector - In Total
x20	Unsubscribe Entreprises From REGON register Total
x21	Unsubscribe Entreprises From REGON register Public Sector - Total
x22	Unsubscribe Entreprises From REGON register Private Sector - In Total
x23	Registred Entreprises in REGON register For 10 Thousand. Population
x24	Newly Registred Entreprises in REGON At 10 Thousand. Population
x25	Unsubscribe Entreprises From REGON For 10 Thousand. Population
x32	Revenue District Budgets - Total
x33	Capita Income In 1 District Budgets
x34	Budget Expenditure District
x35	District Budget Expenditure - Total Capital Expenditure
x36	Budget Expenditure Capital Expenditure Capital District
x37	Budget Expenditure In Total Current Expenditure District
UdzLudProd	Percentage of Working Age People
UdzLudPo-prod	Percentage of Afer Working Age People
UdzPrzedPub	Percentage of Before Working Age People

Source: developed by author, on basis of The Local Data Bank http://www.stat.gov.pl/bdl/app/strona.html?p_name=indeks

APPENDIX B

Table 10. Loadings for Platforma Obywatelska and Prawo i Sprawiedliwość

	Lodings for Platforma Obywatelska			Lodings for Prawo i Sprawiedliwość		
Variables	Comp 1	Comp 2	Comp 3	Comp 1	Comp 2	Comp 3
Attendance	0.234	0.087	0.040	– 0.310	0.063	0.068
x04	0.239	– 0.165	– 0.115	– 0.280	0.234	0.084
x05	0.115	0.112	– 0.155	– 0.157	0.006	0.280
x06	– 0.151	– 0.203	0.256	0.226	0.062	– 0.319
x07	0.177	– 0.138	0.052	– 0.203	0.191	– 0.022
x08	0.147	0.160	– 0.320	– 0.224	– 0.065	0.239
x09	0.245	– 0.153	– 0.084	– 0.291	0.225	0.030
x10	0.251	– 0.128	– 0.065	– 0.303	0.211	0.033
x12	– 0.174	– 0.232	0.324	0.257	0.065	– 0.413
x13	0.253	– 0.148	– 0.072	– 0.301	0.224	0.013
x14	0.232	– 0.252	0.112	– 0.254	0.298	– 0.105
x17	0.251	– 0.148	– 0.073	– 0.300	0.225	0.021
x18	0.234	– 0.239	0.197	– 0.260	0.281	– 0.236
x20	0.256	– 0.146	– 0.041	– 0.305	0.227	– 0.007
x21	0.176	– 0.248	0.454	– 0.180	0.271	– 0.344
x23	0.248	0.174	– 0.064	– 0.348	– 0.007	0.060
x24	0.224	0.168	0.079	– 0.316	– 0.018	– 0.078
x25	0.137	0.298	0.135	– 0.224	– 0.152	– 0.133
x32	0.237	– 0.241	0.099	– 0.264	0.289	– 0.163
x33	0.166	– 0.071	0.173	– 0.193	0.152	– 0.114
x34	0.236	– 0.244	0.102	– 0.262	0.289	– 0.171
x35	0.231	– 0.255	0.107	– 0.253	0.293	– 0.188
x36	0.230	– 0.256	0.108	– 0.253	0.293	– 0.190
UdzLudProd	0.152	0.329	0.346	– 0.256	– 0.188	– 0.271
UdzLudPoprod	0.109	– 0.319	– 0.112	– 0.074	0.320	0.145
UdzPrzedPub	– 0.155	– 0.228	0.553	0.240	0.091	– 0.440

Source: own calculation using **R.**