Slovak University of Technology in Bratislava
Faculty of Mechanical Engineering
17th Conference on Applied Mathematics **APLIMAT 2018**
Proceedings

SPATIAL LAG MODEL FOR APARTMENT PRICES IN PARDUBICE REGION

POZDÍLKOVÁ Alena (CZ), MAREK Jaroslav (CZ)

Abstract. The article is devoted to modelling the relationship of apartment prices between neighbouring small municipalities within the Pardubice region. A spatial lag model will be used for the calculation, and the unknown parameters will be estimated using the least squares method. This model is often used to describe geo-informational phenomena.

Keywords: spatial modelling, spatial lag model, factor model, weights matrix, estimation of unknown parameters, exploratory visualization, spatial lag vs spatial error models

Mathematics Subject Classification: 90J15

1 Introduction

The aim of the article is modeling the relationship between housing prices in small municipalities within the Pardubice region. Specifically, it models the average price of apartments in the municipalities of Pardubice (Česká Třebová, Heřmanův Městec Hlinsko, Holice, Choceň Chrudim, Chvaletice, Lázně Bohdaneč, Letohrad, Litomyšl, Moravská Třebová, Opatovice nad Labem, Pardubice, Polička, Přelouč, Skuteč, Králíky, Svitavy, Ústí nad Orlicí, Vysoké Mýto). The "explaining variable" in this model is the average advertised price of residential units. The "explanatory variables" are the population, the unemployment rate in the district, and the distance from the regional capital. Several of the proposed factor models will be tested for their quality by using a modified index of determination.

It will also be tested to see if the model must also include a spatial factor. This spatial factor, as well as the availability matrix (known from graph theory), allow us to grasp the issue of neighbourhood regions.

2 Housing prices modelling

Econometrics attempts to measure some economic quantities based on the analysis of the real data by using various mathematical-statistical methods. The object of great interest is the measurement of supply and demand using the factor models. The first aim of this paper is to assemble a model for describing the relationships between housing prices and suitable explained variables. The estimators are computed by using a two-stage Least squares method. The second aim is to explore spatial relationship of housing prices throughout the Pardubice region municipalities. A similar problem is solved in [Meen, 1996], where the author studied spatial interactions in UK regional house prices.

The reason for modelling is to determine if the price can be explained by the unemployment rate and the distance from the regional center. Furthermore, we want to prove that the spatial lag model has greater ability to grasp the problem of apartment price modelling. That means whether or not the index of determination is greater.

The non-spatial model estimated by conventional regression procedures is not a reliable representation and should be avoided when there is a spatial phenomenon to be analyzed. [Augustin, 1998]

Special attention must be given to the possibility that errors and variables in the model show spatial dependence in data analysis. The influences of spatial autocorrelation could be drawn from statistical tests. If the interest focuses on obtaining proper statistical inference (estimates of unknown parameters, statistical hypothesis testing, forecasting) from the dependent data, spatial autocorrelation can be considered a nuisance.

2.1 Formulation of regression model

Firstly, we propose a simple regression model, which will feature spatial coefficient. We will work with tools from graph theory, which allow us to describe the existence or nonexistence of a boundary between the municipalities. [Anselin, Rey, 1991], [Augustin, 1998], and [LeSage, 2009] show formulas for the spatial lag model that is given by:

$$\boldsymbol{Y} = \lambda \boldsymbol{W}\boldsymbol{Y} + \boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I}) \,. \quad (1)$$

Here the symbols denote:

- λ… coefficient in a spatial autoregressive structure for the disturbance,
- $\boldsymbol{\gamma}$ … vector of unknown parameters,
- **X** … matrix of explanatory variables,
- $\boldsymbol{W}$ ... known spatial weight matrix, where $W_{ij} = 1$, if boundary between i-th and j-th region exists, $W_{ij} = 0$ if there is no boundary,
- $\boldsymbol{\varepsilon}$ … vector of independent possibly heteroscedastic spatial errors.

Explained variables

In [Kotatkova, 2015], we can see that suitable model for housing prices use three explained variables: distance from the regional center, population, and unemployment rate.

We will use following notation:

- **Y**… the average advertised price of 1 m^2 of apartment. Price per 1m^2 is determined as a proportion of the purchase price and floor area (except of balkonies, and terraces). [Reality, 2017] Data obtained automatically – by parsing of real estate webside.
- $\boldsymbol{x}_1$ … distance from the regional center, cf. [Routeplanner, 2017],
- $\boldsymbol{x}_2$ … average wage throughout the whole municipality, cf. [Payroll, 2017],
- $\boldsymbol{x}_3$ … unemployment rate (in the selected municipality), cf. [Unimployment, 2017].

2.2 Estimation in spatial lag model

The residual regression estimation is divides into two steps [Bhatacharjee, 2006], [LeSage, 2009]:

First stage - Ordinary least squares regression of parameters γ in model (2) with estimate (3):

$$\boldsymbol{Y} = \boldsymbol{X\gamma} + \boldsymbol{u}, \boldsymbol{u} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\sigma}^2 \boldsymbol{V}), \quad (2)$$

$$\widehat{\boldsymbol{\gamma}} = (\boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{Y}, \quad (3)$$

$$\hat{u}_t = Y_t - \boldsymbol{X}_{t.} \widehat{\boldsymbol{\gamma}}. \quad (4)$$

Second stage – spatial lag parameter estimation:

$$\hat{u}_{kt} = \rho_k \sum_{j=1, j \neq k}^{k} w_{kj} \hat{u}_{jt} + \varepsilon_{kt}, \quad (5)$$

where the second stage regressions are estimated separately for each region.

2.3 Measuring of model quality

In this section, we describe characteristics for measuring the quality of the regression model. Let $\bar{y}$ be the mean estimate based on measurements $y_1, \ldots, y_n$ and let s_y^2 be the estimate of dispersion, $s_y = \frac{1}{n} \sum (y_i - \bar{y})^2$. Following quantities express variation of the dependent variable. The total sum of squares SS_T of Y: $SS_T = \sum (y_i - \bar{y})^2$, the sum of squares due to regression SS_R:

$$SS_R = \sum (\hat{y}_i - \bar{y})^2.$$

The goodness of fit can be defined by

$$R^2 = \sqrt{\frac{SS_R}{SS_T}}. \quad (6)$$

The square root of goodness of fit is the correlation coefficient $R = \sqrt{R^2}$. In the case of several explanatory variables, an adjusted measure of R^2 is used usually:

$$R^2 = 1 - (1 - R^2) \left(\frac{N-1}{N-k-1} \right). \quad (7)$$

2.4 Moran's I test

The Moran's statistic has been used as a way of spatial dependency description. Calculations for Moran's I are based on a weighted matrix W and measured values Y:

$$I = \frac{N}{S_0}\left(\frac{e^T W e}{e^T e}\right), \tag{8}$$

with e as a vector of OLS residuals and $S_0 = \sum_i \sum_j w_{ij}$, a standardization factor that corresponds to the sum of the weights for the nonzero cross-products. The statistic shows a striking similarity to the familiar Durbin–Watson test. [Anselin, 2001]

Those interested in derivation of previous formula are referred to paper. [Moran, 1950]

2.5 Test statistic for spatial lag model: Lagrange multiplier test for spatial lag model

If we would like to explore in our model existence of spatial dependency, then the problem of lag coefficient statistical significance appears.

Our null hypothesis is that the spatial arrangement does not affect to value of dependent variable, i.e. the spatial coefficient is equal to zero. For detail see [Anselin, 2001], [Anselin, Rey, 1991].

The Lagrange Multiplier (or Score) test statistic is calculated using the following formula:

$$LM_{err} = [\boldsymbol{e}^T \boldsymbol{W} \boldsymbol{e}/(\boldsymbol{e}^T \boldsymbol{e}/N)]^2/[tr(\boldsymbol{W}^2 + \boldsymbol{W}^T \boldsymbol{W})]. \tag{9}$$

This statistic has an asymptotic $\chi^2(1)$ distribution and, apart from a scaling factor, corresponds to the square of Moran's I. From several simulation experiments [Anselin, Rey, 1991]; it follows that Moran's I has slightly better power than the LM_{err} test in small samples, but the performance of both tests becomes indistinguishable in medium and large size samples. The LM/RS test against a spatial lag alternative was outlined in and takes the form [Anselin, 2001]:

$$LM_{lag} = [\boldsymbol{e}^T \boldsymbol{W} \boldsymbol{e}/(\boldsymbol{e}^T \boldsymbol{e}/N)]^2/D, \tag{10}$$

where

$$D = [(\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T)(\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta})/\sigma^2] + tr(\boldsymbol{W}^2 + \boldsymbol{W}^T\boldsymbol{W}).$$

3 Numerical studies

Example 1. The aim of our study will be to find a suitable model for describing the price of 1m^2 by using some / some variables: distance, population, unemployment. At the same time, we investigate whether spatial dependence exists between variables. This study is realized in the municipalities of Pardubice: Česká Třebová, Heřmanův Městec Hlinsko, Holice, Choceň Chrudim, Chvaletice, Lázně Bohdaneč, Letohrad, Litomyšl, Moravská Třebová, Opatovice nad Labem, Pardubice, Polička, Přelouč, Skuteč, Králíky, Svitavy, Ústí nad Orlicí, Vysoké Mýto. Data are available from [Reality, 2017], [Routeplanner, 2017], [Payroll, 2017], [Unemployment, 2017].

Table 1 shows the regions characteristics. The explanatory data consists of three variables. We will consider a lag model with explanatory variables $\mathbf{x_1}$, $\mathbf{x_2}$, $\mathbf{x_3}$, described in section 1.1. First we try to explain the dependent variable in models 1, 2, 3 using only one variable $\mathbf{x_1}$, $\mathbf{x_2}$ or $\mathbf{x_3}$. We determine the indices determination and decide which variable in the regression model (1) is best explained-cost apartments. Then we build a 4 model using the two best variables. From this model will be model with spatial coefficient created – model 5, see Table 3 in Results. Measured data are shown in table 1. In our case the number of measurement is N=15. Data was obtained on April 1st 2017.

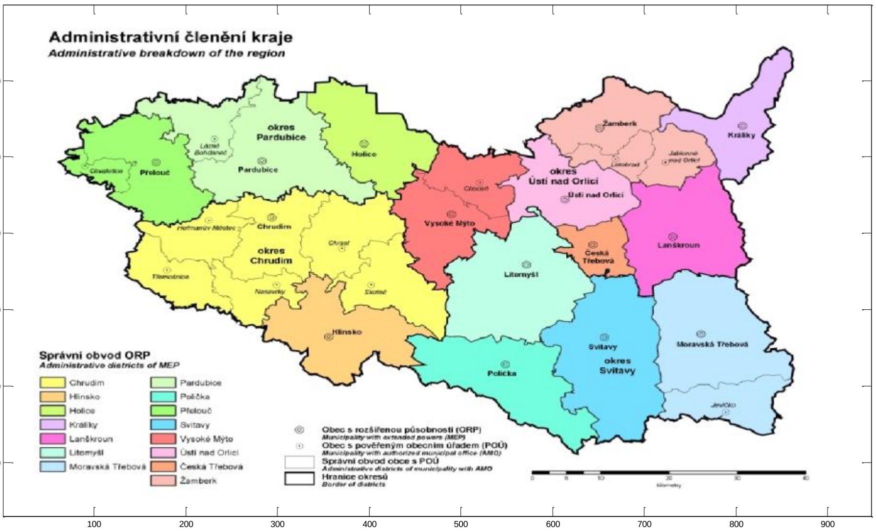

Fig. 1. Map of Pardubice Region.

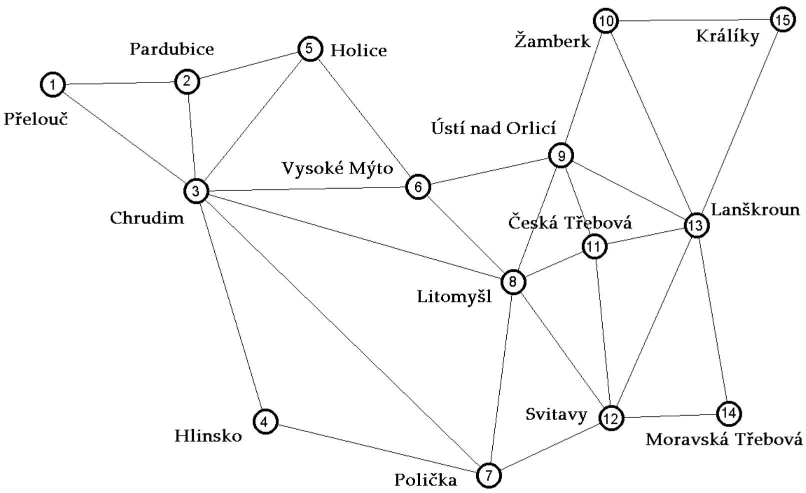

Fig. 2. Pardubice Region – description by graph.

No.	Region	Distance x_1	No. of inhabitants x_2	Unemploity x_3	Prize Y
1.	Přelouč	0.36	9 127	6.13 %	33 113 CZK
2.	Pardubice	0.00	89 638	6.13 %	37 501 CZK
3.	Chrudim	0.25	23 061	8.39 %	31 386 CZK
4.	Hlinsko	0.70	9 831	8.39 %	14 558 CZK
5.	Holice	0.32	6 514	6.13 %	28 041 CZK
6.	Vysoké Mýto	1.46	12 404	6.95 %	21 144 CZK
7.	Polička	1.06	8 783	9.29 %	17 269 CZK
8.	Litomyšl	0.83	10 043	9.29 %	32 101 CZK
9.	Ústí nad Orlicí	1.00	14 226	6.95 %	19 563 CZK
10.	Žamberk	1.04	6 062	6.95 %	21 104 CZK
11.	Česká Třebová	1.05	15 710	6.95 %	18 756 CZK
12.	Svitavy	1.05	17 005	9.29 %	26 279 CZK
13.	Lanškroun	1.30	10 031	6.95 %	24 771 CZK
14.	Moravská	1.23	10 267	9.29 %	19 947 CZK
15.	Králíky	1.42	4 300	6.95 %	9712 CZK

Tab. 1. Prices of housing in selected regions with explanatory variables.

The method used to compute unknown parameters, described in section 2.2, is based on the two step strategy. The existence of the boundaries of Figures 1 and 2 expresses the matrix W. Fig. 1 and Fig. 2 show municipalities in Pardubice region. From these figures weights matrix can be obtained:

$$\boldsymbol{W} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Given that $\boldsymbol{y} = \boldsymbol{\lambda}\, \boldsymbol{W}\boldsymbol{y}$, we obtain a set of equations that we will render in relation to the unknown λ. [LeSage, 2009] And here we have the ability of each equation to get $\lambda_1, \ldots, \lambda_{15}$. These values can be redirected to get one λ.

4 Results

In the first model, we use x_1 as the explanatory variable, in the second model we use the variable x_2 and in the third model we use the variable x_3. The fourth model is based on two best explanatory variables from previous models, such as distance and population. The fifth model is based on model 4 by adding a spatial expression $\boldsymbol{\lambda W y}$. By application of least squares method, following results were achieved:

in model 1: $\hat{\boldsymbol{\gamma}} = (34539, -12459)$, $R^2 = 0.7318$, $LM_{err} = 13.29$, $LM_{lag} = 41.93$,
in model 2: $\hat{\boldsymbol{\gamma}} = (20303, 0)$, $R^2 = 0.5561$, $LM_{err} = 33.40$, $LM_{lag} = 42.21$,
in model 3: $\hat{\boldsymbol{\gamma}} = (31625, -1045)$, $R^2 = 0.1694$, $LM_{err} = 86.89$, $LM_{lag} = 52.33$,
in model 4: $\hat{\boldsymbol{\gamma}} = (31545, -10460, 0.076)$, $R^2 = 0.7512$, $LM_{err} = 24.60$, $LM_{lag} = 40.54$,
in model 5: $\hat{\boldsymbol{\gamma}} = (31545, -10460, 0.076)$, $R^2 = 0.8053$., estimates of spatial coefficients $\boldsymbol{\lambda}$ are given in Tab. 3.

No.	Model 1 $\hat{Y}$	Model 1 $\hat{\varepsilon}$	Model 2 $\hat{Y}$	Model 2 $\hat{\varepsilon}$	Model 3 $\hat{Y}$	Model 3 $\hat{\varepsilon}$
1	30053.65	3059	22176.22	10937	25220.88	7892
2	34538.83	2962	38704.53	-1203	25220.88	12280
3	31424.12	-38	25036.77	6349	22859.77	8526
4	25817.64	-11260	22320.74	-7763	22859.77	-8302
5	30552.00	-2512	21639.79	6401	25220.88	2820
6	16348.92	4795	22848.96	-1705	24364.20	-3220
7	21332.46	-4064	22105.60	-4837	21919.50	-4651
8	24197.99	7903	22364.27	9737	21919.50	10182
9	22079.99	-2517	23223.00	-3660	24364.20	-4802
10	21581.63	-478	21547.00	-443	24364.20	-3260
11	21457.04	-2701	23527.66	-4772	24364.20	-5608
12	21457.04	4822	23793.51	2486	21919.50	4360
13	18342.33	6428	22361.80	2409	24364.20	406
14	19214.45	733	22410.25	-2463	21919.50	-1972
15	16847.27	-7135	21185.27	-11473	24364.20	-14652

Tab. 2. Results.

Obtained results illustrates the differences between models. We can see that model 3 is very poor. The best results are calculated in model 4. In this model null hypothesis is rejected, e.g. spatial dependency is significant.

No.	Model 4 $\hat{Y}$	Model 4 $\hat{\varepsilon}$	Model 5 $\hat{Y}$	Model 5 $\hat{\varepsilon}$	λ
1	28473.63	4639	28453	4660	0.13
2	38362.04	-861	38221	-720	-0.03
3	30683.92	702	30722	664	0.03
4	24970.61	-10413	26327	-11769	-0.43
5	28693.34	-653	28618	-577	-0.02
6	17216.34	3928	18045	3099	0.14
7	21125.15	-3856	20647	-3378	-0.15
8	23626.88	8474	23472	8629	0.38
9	22166.71	-2604	20468	-905	-0.11
10	21127.44	-24	21127	-23	-0.00
11	21756.54	-3001	19794	-1038	-0.12
12	21855.02	4424	23488	2791	0.20
13	18709.55	6061	16142	8629	0.32
14	19459.73	488	19669	278	0.02
15	17018.47	-7306	15086	-5374	-0.32

Tab. 3. Results.

5 Conclusions

Paper presented to the readers the solution of one type of spatial regression problem. This is a situation where it should be a spatial dependence of measured data to consider. To using this model, we are inspired by studies from [Kotatkova, 2015] and [Bhatacharjee, 2006]. First paper is devoted to regression model for description of flat prices in Czech Republic by suitable variables. Second paper studied spatial diffusion in housing starts across regions in England by regressing OLS residuals of the regression relationship for each region on residuals from all the other regions. Our results are analogous like in these studies. We proofed that use of spatial model is necessary.

Acknowledgement

This paper was supported by internal support of University of Pardubice.

References

[1] ANSELIN, L., *Spatial Econometrics*, 14th chapter, A Companion to Theoretical Econometrics, Copyright by Blackwell Publishing Ltd, 2001, 310-330.

[2] ANSELIN, L., REY, S., Properties of Tests for Spatial Dependence in Linear Regression. *Geographical analysis* 23, 1991, 112–131.

[3] AUGUSTIN, N. H. et al.: The role of simulation in modelling spatially correlated data. *Environmetrics* 9, 1998, 175–196.

[4] BHATACHARJEE, A., JENSEN-BUTLER, C.: *Estimation of Spatial Weights Matrix, with an Application to Diffusion in Housing Demand.* Glasgow: School of Economics and Finance, University of St. Andrews, U.K., 2006.

[5] KOTATKOVA STRANSKA, P., MAREK, J.: Simultaneous equations and modelling of new flat demand. In: *Mezinárodní vědecká konference Hradecké ekonomické dny 2015* (Jedlička, P., ed.). Univerzita Hradec Králové, Fakulta informatiky a managementu, Hradec Králové, 2015, díl I, 11–16.
[6] LESAGE, J., PACE, K. R.*: Introduction to spatial Econometrics.* CRC Press/Taylor & Francis Group, Boca Raton, 2009.
[7] MEEN, G.: Spatial Aggregation, Spatial Dependence and Predictability in the UK Housing Market. *Housing Studies* 11, 1996, 345–372.
[8] MORAN, P. A. P. : Notes on Continuous Stochastic Phenomena. *Biometrika* 37 (1), 1950, 17–23. DOI:10.2307/2332142. JSTOR 2332142. http://www-hsc.usc.edu/~mereditf/Spatial%20Statistics%206.pdf
[9] Reality: Real estate data – prize per 1 m2 of flat. [online], [cit. 8. 5. 2017]. URL: http://www.sreality.cz/hledani/prodej /byty/
[10] Routeplanner: Routeplanner SKODA auto. [online], [cit. 8. 5. 2017]. URL: www.planovac-tras-mapy.cz/
[11] Payroll. [online]. [cit. 8. 5. 2017]. URL: http://www/brno.czso.cz/xb/redakce.nsf/i/mesta_a_obce
[12] Unemployment. ČSÚ. [online]. [cit. 8. 5. 2017]. URL: http://www.czso.cz

Current address

Pozdílková Alena, Mgr. Ph.D.
Department of Mathematics and Physics, University of Pardubice
Faculty of Electrical Engineering and Informatics
Studentská 85, 530 02 Pardubice, Czech Republic
E-mail: alena.pozdilkova@upce.cz

Marek Jaroslav, Mgr. Ph.D.
Department of Mathematics and Physics, University of Pardubice
Faculty of Electrical Engineering and Informatics
Studentská 85, 530 02 Pardubice, Czech Republic
E-mail: jaroslav.marek@upce.cz