SPATIAL STATISTICAL ANALYSIS OF PARTICIPANTS IN THE INDIVIDUAL PENSION SYSTEM OF TURKEY

Vural YILDIRIM [1,*], Yeliz MERT KANTAR [2]

[1] Department of Statistics, Faculty of Science, Anadolu University, Eskisehir, Turkey
[2] Department of Statistics, Faculty of Science, Eskisehir Technical University, Eskisehir, Turkey

ABSTRACT

The individual pension system is known as the personal future investment system that allows participants to regularly save for their retirements. Thus, the role of this system among incomes in retirement has significantly grown in the last decade in Turkey as in other countries. In Turkey, the participation in the individual pension system has been voluntary since the beginning of this system. Thus, it is important to study the factors affecting the participants of the individual pension system and also the distribution of the participants in Turkey. In this study, participants' data in the individual pension system of Turkey (2013), provided by the Pension Monitoring Center, is examined at provincial level by means of GIS and spatial statistics tools. Firstly, spatial distribution, spatial local and global clustering for the participation rate have been researched. Also, the participant rate in Turkey with the effects of variables measured at provincial level (the labor rate, unemployment rate, sex ratio, urbanization rate, deposit rate, illiteracy rate and human development index) is studied by using the spatial econometrics models. The result shows that there are global and local spatial autocorrelations between the participation rates in Turkey and also the spatial lag model provides better results than the classical regression model and some other spatial models in terms of all criteria.

Keywords: Participation rate in the individual pension system (IPS), Turkey, GIS (geographic information system), Spatial econometrics, Spatial autocorrelation

1. INTRODUCTION

Although the individual pension system (IPS) has been implemented for many years in the world, it began operating in Turkey on October 27, 2003 [1]. This system enables individuals to save their earnings voluntarily during work time. Many factors affecting the participation in IPS have been discussed in many ways in the literature. For example, Cetin and Sevuktekin [2] studied factors affecting people's entrance to IPS in Bursa, Turkey. Kara et al. [3] analyzed the risk-taking behavior of the IPS participants in Turkey. Also, Gokcen and Yalcin [4] considered some models for private pension funds by panel regression and time series. Cetin and Sevuktekin [5] studied with spatial data to define most different provinces in value of income between the years 2007 and 2014. By using spatial statistics, they researched the homogeneity or spatial interactions between provinces in Turkey without spatial econometric analysis. On the other hand, Sahin et al. [6] examined the effects of social variables on individuals's regular contributions to the system and analyzed the gender gap by using the generalized linear model. Yildiz et al. [7] investigated impacts of socioeconomic and demographic attributes on the persistence of individuals' payments depending on their own private pension schemes by classifying the individuals according to their genders. Similarly, Comlekci and Gokmen [8] studied on participants' demographic features in order to determine the effect about the participation in IPS. More recently, Bauer et al. [9] investigated the effect of interconnections on strategic investment decisions for 191 Dutch pension funds using spatial autoregressive model.

In Turkey, although participating within IPS has been voluntary since the beginning of the system, many regulations for IPS have been issued in the last decades. In this study, we examined the participants' rate in IPS in Turkey at the level of provincial in order to understand and explain the regional differences in terms of IPS. For this purpose, spatial distribution of the participant rate in IPS is investigated by means of spatial autocorrelation measure, spatial maps and spatial econometrics models.

Within the above described framework, this study is organized as follows: Spatial statistics and spatial econometric models are briefly presented in section 2. Data Description is provided in Section 3. Spatial analysis of participant rate in IPS of Turkey is conducted in Section 4. Results and suggestions are provided in Section 5.

2. METHODS: SPATIAL MAPS, SPATIAL STATISTICS AND SPATIAL ECONOMETRIC MODELS

In this section, the used spatial statistics tools and some econometrics models are briefly introduced.

2.1. Spatial Autocorrelation Statistics

While spatial statistics cover a set of techniques for the analysis of spatially located data, spatial econometrics accounts for estimating spatial effects in regression models [10, 11]. Main objective in analysing spatial data is to detect the presence of spatial autocorrelation in data. There are several measures to test spatial autocorrelation. Most commonly-used measure is Moran I test given in Equation 1.

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \tag{1}$$

where n is the total number of total observations x_i is the value measured in the spatial district i, x_j is the measured in the another spatial district j, $\bar{x}$ is the mean value of x_i, ($i=1,…,n$). Also, w_{ij}, which shows interaction between the district i and the district j, is the ijth component of the spatial weight matrix. There are many kinds of spatial weigh matrix in literature. Binary matrices based on geographic arrangement of the observations or contiguity such as rook and queen continuity are most commonly used in spatial data analysis. Also, the spatial weights can be also defined based on distance, the structure of a social network, economic distance and k nearest neighbours [12, 13]. In this study, we consider the queen contiguity matrix to express spatial structures between locations. The Queen contiguity network for the 81 cities in Turkey is demonstrated in Figure 1.
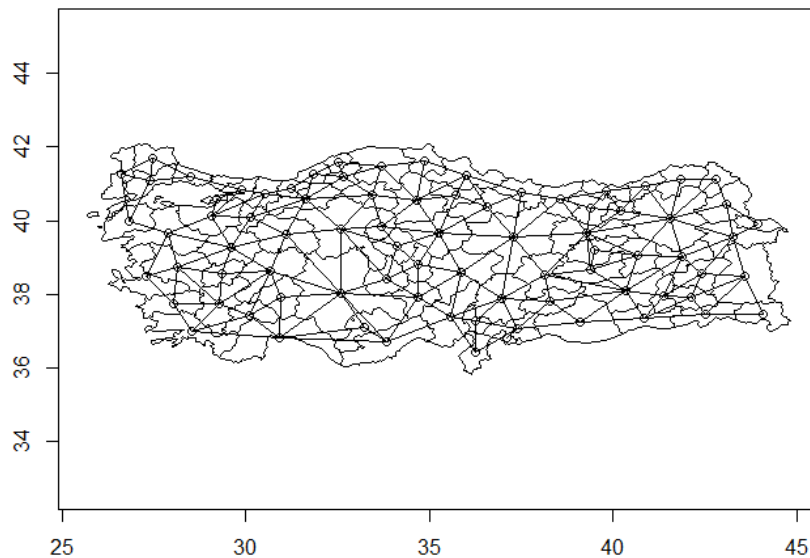
Figure 1. The Queen contiguity network for the 81 cities in Turkey

LISA (Local Indicators of Spatial Association), the local variant of Moran's I, is also used to quantify spatial autocorrelation and clustering within small areas and it is given in Equation 2.

$$I_i = \frac{(x_i - \bar{x})}{s} \sum_j w_{ij} \frac{(x_j - \bar{x})}{s} \tag{2}$$

2.2. Spatial Econometric Models

Spatial econometrics that deals with spatial autocorrelation and spatial heterogeneity in regression models for spatial data [14] is a subfield of econometrics. The models in spatial econometrics are generally called as spatial regression models, which add a spatial weighting matrix to a general linear model. There are several models which may contain spatial interaction among the dependent and/or independent variables and/or error terms. These models are listed as follows:

Spatial lag Model (SLM) (or known as spatial autoregressive model, SAR) is provided by Equation 3.
$$y = \rho W y + X\beta + \varepsilon \tag{3}$$
where $\varepsilon \sim N(0, \sigma^2)$, y is a vector of nx1 dependent variables; X is a nxk matrix of the independent variables; β is a vector of kx1 vector of unknown parameters of the model and ρ is a coefficient of the spatial autoregressive structure for Wy.

Spatial Error Model (SEM) with the spatial autoregressive parameter for the error term is presented as follows:
$$y = X\beta + u \tag{4}$$
$$u = \lambda W u + \varepsilon$$
where $\varepsilon \sim N(0, \sigma^2)$, $u \sim N(0, \Sigma)$ and λ is the spatial autoregressive parameter for the error term.

The formulation of Σ considers both heteroscedastic and auto-correlated error terms for $\lambda \neq 0$, it follows the classic homoscedastic situation for $\lambda=0$ and Σ is expressed in Equation 5.

$$\Sigma = (I_n - \lambda W)^{-1}(I_n - \lambda W')^{-1}\sigma^2 \tag{5}$$

Spatial Lag of X Model (SLX) is given in Equation 6.
$$y = X\beta + WX\theta + \varepsilon \tag{6}$$
where $\varepsilon \sim N(0, \sigma^2)$.

Spatial Lag Combined Model (SAC), which consists of SLM and SEM is presented as follow:
$$y = \rho W y + X\beta + u \tag{7}$$
$$u = \lambda W u + \varepsilon$$
where u and ε are distributed normally as in SEM.

Spatial Durbin Error Model (SDEM), which takes into account the spatially lagged independent variable as well as the spatial autoregressive parameter for the error term, is given in Equation 8.
$$y = X\beta + WX\theta + u \tag{8}$$
$$u = \lambda W u + \varepsilon$$
where u and ε are distributed normally as in SEM.

Spatial Durbin Model (SDM) with only the spatially lagged independent variable is
$$y = \rho W y + X\beta + WX\theta + \varepsilon, \tag{9}$$
where $\varepsilon \sim N(0, \sigma^2)$.

General Nesting Spatial Model (GNS) which considers all spatial autocorrelation in the variables, dependent, independent and error is provided in Equation 10.

$$y = \rho W y + X\beta + WX\theta + u \tag{10}$$
$$u = \lambda W u + \varepsilon$$

where u and ε are distributed normally as in SEM.

Generally, the Lagrange multiplier or likelihood ratio tests are used to determine the best spatial model among SEM, SLM, SAC and also the maximum likelihood (ML) estimation as well as the generalized method of moments and two-stages least squares method, is an accepted method for estimating parameters of spatial econometric models [15, 16]. Thus, in this study, we conducted the ML estimation method to estimate the parameters of the considered models.

3. DATA DESCRIPTION

The number of participants within IPS in Turkey is provided by the Pension Monitoring Center. The participant rate is calculated as a ratio of participants to the population according to provinces (%). In this study, we use the participant rate data collected for 2013 since the other factors related to IPS such as the labor rate and unemployment rate are not announced at the level of provincial in Turkey after 2013. While the map for the provinces in Turkey is given in Figure 2, the choropleth map of the participant rate in IPS is demonstrated in Figure 3. It is seen from Figure 3 that Istanbul and Muğla have the highest rate in Turkey, on the other hand, Muş and Ağrı indicate the lowest rate. It is apparently observed that the east and southeast of Turkey have lowest rates of participants, while Antalya, Muğla and İzmir, located in the west and southwest regions, Ankara, capital of Turkey and İstanbul, an important industrial city of Turkey, have the highest participation rates.
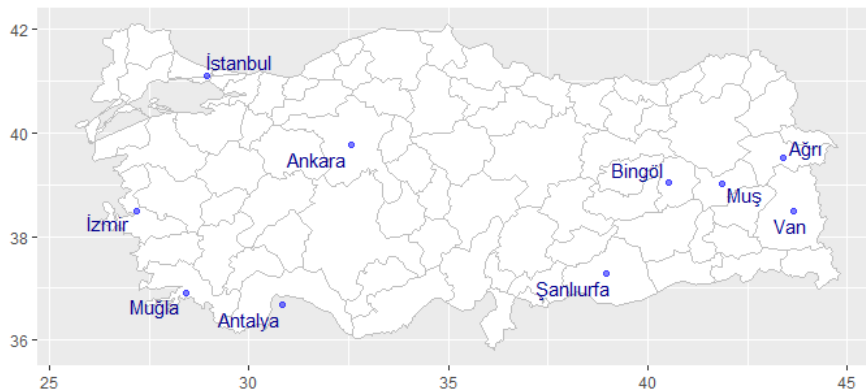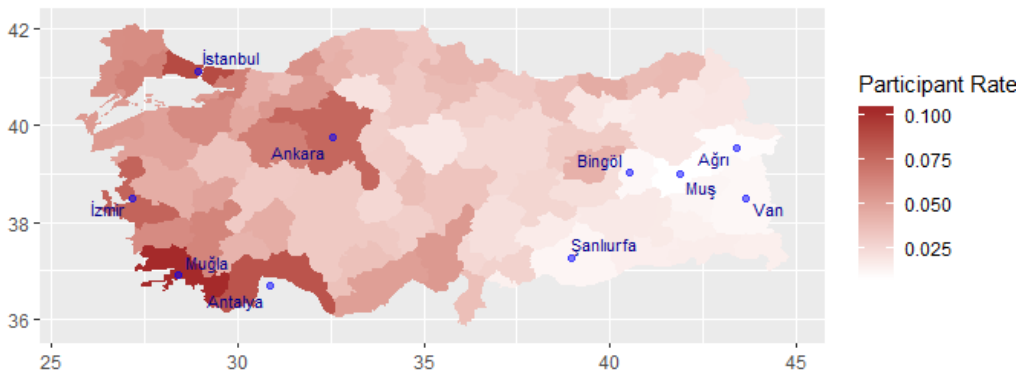

Figure 2. Map of Turkey

Figure 3. Spatial distribution of the participant rate within IPS in Turkey

The histogram and statistical density functions (Kernel and Normal) of the participation rate are given in Figure 4. While the average of the participation rates is 0.037, the maximum rate remains the level of 0.102. It is also seen Figure 3 that Muş is the worst in terms of IPS as stated in [5],
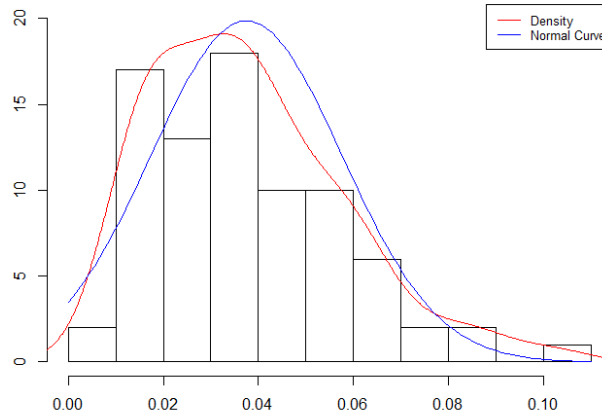
Figure 3. Histogram, Kernel and Normal density graphs of the participant rate within IPS in Turkey

As well as the participant rate in IPS, data which consist of the variables, the labor rate, unemployment rate, sex rate, urbanization rate, deposit rate, illiteracy rate and human development index (HDI) measured at provincial level, are obtained from Turkish Statistical Institute and The Economic Policy Research Foundation of Turkey . The correlation matrix for all variables in a dataset is given in Figure 5.
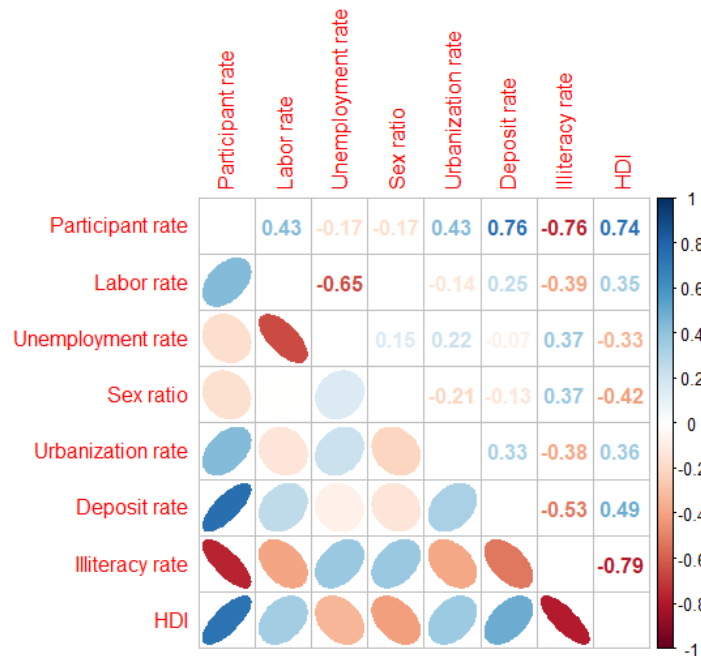

**Figure 5.** Correlation Matrix between the participant rate within IPS and the other variables.

From Figure 5, although the illiteracy rate is highly negative correlated with the participant rate within IPS, there is a positive high correlation not only between deposit rate and participant rate but also between HDI and participant rate within IPS.

4. SPATIAL ANALYSIS OF PARTICIPANT RATE IN THE INDIVIDUAL PENSION SYSTEM OF TURKEY

In order to visualize the spatial patterns of the participant rate within IPS in Turkey, the Moran I scatter plot, as well as the choropleth map, is given in Figure 6. It is seen from Figure 6 that there may be global spatial correlation according to the participant rate. Additionally, the Correlogram plot in Figure 7 supports the spatial dependence.
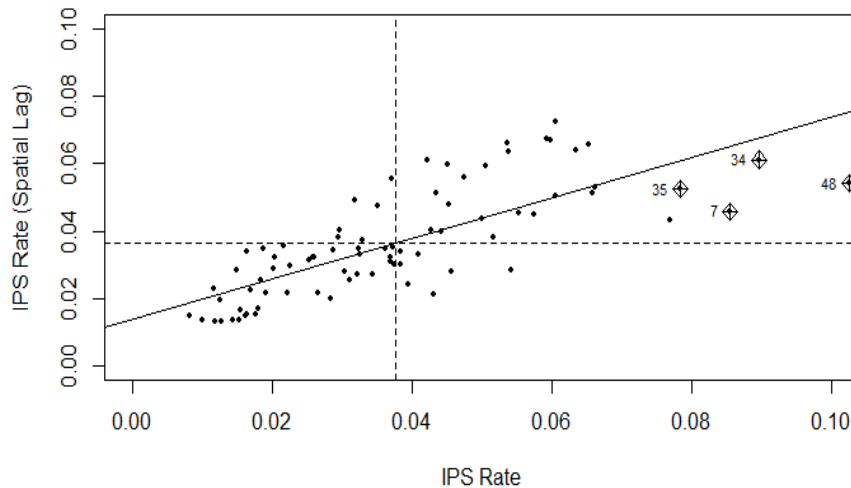
Figure 6. Moran's I scatter plot for the participation rate in IPS
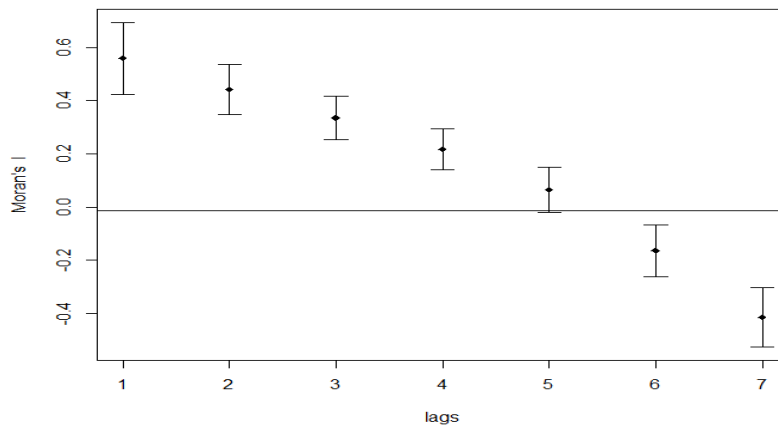
Figure 7. Correlogram plot for the participation rate in IPS
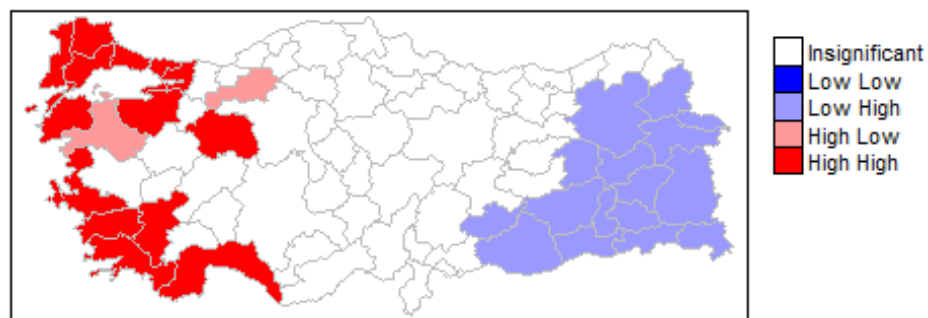
Figure 8. Lisa or Local Moran I plot for the participation rate in IPS

LISA (Local Indicator of Spatial Association) can be used in order to identify geographic locations where spatial local autocorrelation occurs. Through LISA analysis, a cluster map showing places where spatial autocorrelation exists is given for the participation rate in IPS in Figure 8. In this map, red colour, which means high values surrounded by high values (H-H), and light blue color, which accounts for low values surrounded by high values (L-H), indicates spatial clustering of the west and east regions in Turkey. Figure 8 shows the existence of two core regions (west and east) in Turkey in terms of the participation rate within IPS. East of Turkey, which has low-high values (light blue), specifies negative spatial associations and also potential spatial outliers on the map.

Now, we consider the spatial autocorrelation between the participation rate in IPS (y) and its spatially lagged variable (Wy) and also the variables, the labor rate, unemployment rate, sex ratio, urbanization rate, the deposit rate, illiteracy rate and human development index (HDI), OLS (ordinary linear regression) residuals and their spatially lagged variables. It can be obtained from Table 1 that all variables have correlated with corresponding spatially lagged variables. These results show that the spatial econometric models can be used for the estimation of coefficients.

Table 1. Moran I values between the variables and the spatially lagged variables

Variables	Moran I statistic	*p*-value
Participation rate	0.61446	< 0.001
Labor rate	0.61957	< 0.001
Unemployment rate	0.57423	< 0.001
Sex ratio	0.26029	0.00001
Urbanization rate	0.17893	0.00403
Deposit rate	0.20723	0.00026
Illiteracy rate	0.76320	< 0.001
HDI	0.69125	< 0.001
OLS residuals	0.19378	0.00032

Before using spatial models, another step is checking the OLS residuals in terms of heteroscedasticity and normality. We observed that the Breusch-Pagan (BP) test of heteroscedasticity is not significant (BP test value = 8.141, *p*-value = 0.320), while the Jarque-Bera (JB) test (JB test value = 43.133, *p*-value = 0.000) reveals non-normality of error terms. According to the results of all analyses, it seems more suitable to apply spatial regression models for the estimation and modelling of the participation rate within IPS.

Table 2 provides the results of criteria, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Schwarz Information Criterion (SIC) and also model selection tests, the likelihood ratio (LR) test and Wald test for autoregressive parameters. The results of Table 2 show that SLM and SAC has less AIC values than the others, on the other hand, GNS provides the highest logL. According to AIC, BIC and SIC, SLM can be considered to be the most suitable model for estimation of the participation rate within IPS. On the other hand, LR and Wald (except SAC and GNS) tests confirm that the parameters of spatial autocorrelation are significant in all spatial models.

Table 2. Comparison of the models and the results of tests

	Model Selection Criteria				Model Selection Test			
	Log-Likelihood	AIC	BIC	SIC	LR	*p*-value	Wald	*p*-value
OLS	277.203	-538.4066	-519.2510	-9.248383	-	-	-	-
SEM	281.350	-542.7007	-518.7562	-9.297235	-8.294	3.98e-03	14.803	0.000119
SLM	283.055	-546.1094	-522.1649	-9.305164	-11.703	6.24e-04	12.726	0.000361
SLX	283.345	-536.6903	-500.7735	-9.020266	-12.284	9.16e-02	-	-
SAC	283.295	-544.5903	-518.2514	-9.256116	-12.184	2.26e-03	0.663	0.717969
SDM	285.827	-537.6531	-496.9475	-9.000198	-17.247	2.76e-02	0.012	0.011937
SDEM	285.119	-536.2371	-495.5314	-8.979332	-15.830	4.49e-02	0.022	0.022430
GNS	285.926	-535.8528	-492.7527	-8.979106	-17.446	4.22e-02	0.999	0.999999

Table 3. The estimated parameters of the models

Parameters	OLS	SEM	SLM	SLX	SAC	SDM	SDEM	GNS
ρ	-	-	0.304**	-	0.257**	0.345**	-	0.461
(Intercept)	-0.155**	-0.156**	-0.150**	-0.131	-0.154**	-0.067	-0.109	-0.055
Labor rate	0.001**	0.001**	0.001**	0.001*	0.001**	0.001**	0.001**	0.001*
Unemployment rate	0.001**	0.001**	0.001**	0.001*	0.001**	0.001**	0.001**	0.001*
Sex ratio	0.053**	0.056**	0.053**	0.051*	0.056**	0.053**	0.049**	0.055**
Urbanization rate	0.010*	0.011**	0.011**	0.009	0.011**	0.010*	0.009*	0.010*
Deposit rate	0.002**	0.002**	0.002**	0.002**	0.002**	0.002**	0.002**	0.002*
Illiteracy rate	-0.300**	-0.216**	-0.151*	-0.193*	-0.160*	-0.169*	-0.188**	-0.165*
HDI	0.115**	0.124**	0.097**	0.095**	0.104**	0.102**	0.100**	0.102**
lag.Labor rate	-	-	-	0.001**	-	0.001	0.001*	0.001
lag.Unemployment rate	-	-	-	0.002*	-	0.001	0.001*	0.001
lag.Sex ratio	-	-	-	-0.051	-	-0.070	-0.053	-0.074
lag.Urbanization rate	-	-	-	-0.010	-	-0.012	-0.008	-0.013
lag.Deposit rate	-	-	-	0.000	-	0.000	0.000	-0.001
lag.Illiteracy rate	-	-	-	-0.218	-	-0.105	-0.202	-0.071
lag.HDI	-	-	-	0.004	-	-0.048	-0.013	-0.057
λ	-	0.476**	-	-	0.160	-	0.324**	-0.183

Signif. codes   '**' 0.05   '*' 0.1 '

Table 3 displays the OLS estimates of classical linear model and the ML estimates of the spatial models. Illiteracy rate has maximum effect among variables for all models and labor and unemployment rate has minimum except autoregressive parameters. These three variables are also significant in all models. Illiteracy rate has the only negative effect on the participation among independent variables except lagged and spatial autocorrelated coefficients. Illiteracy rate has maximum effect in OLS meanwhile, has minimum effect in SLM however lagged illiteracy rate is insignificant in SLX, SDM, SDEM and GNS. Labor and unemployment rates are the only lagged variables which are significant in SLX and SDEM among all lagged variables. HDI has the 2nd greatest effect among independent variables and it is also significant in all models, however lagged HDI is insignificant in SLX, SDM, SDEM and GNS. Surprisingly, urbanization rate is the only variable not significant in SLX in all models among independent variables. Illiteracy rate and HDI shows different values among all models considering OLS, although there are slight differences for other variables.

The spatial autocorrelation coefficients ρ and λ are insignificant in GNS; likewise, λ is insignificant also in SAC. This is supported by results of Wald test too. The highest value of ρ is 0.345 in SDM and the lowest value is 0.257 in SAC considering significance. The spatial autocorrelated coefficient λ is significant only in SEM and SDEM so its highest value is 0.476 in SEM and lowest value is 0.324 in SDEM. Lagged variables reduced spatial autocorrelation in error term. According to AIC, SLM is the best model fitting the data. Estimated coefficient value of spatial dependence has values as 0.304 in SLM. This means that the participation rate in a region is effected by rate as 0.304 of the neighbour's rates.

Table 4. Impact measures of variables for SLM

SLM	Impact measures		
Variables	Direct	Indirect	Total
Labor_Rate	0.00076	0.00031	0.00107
Unemployment_Rate	0.00091	0.00037	0.00127
Sex_Ratio	0.05454	0.02215	0.07669
Urbanization_Rate	0.01100	0.00447	0.01546
Deposit_Rate	0.00171	0.00070	0.00241
Illiteracy_Rate	-0.15457	-0.06277	-0.21733
HDI	0.09867	0.04007	0.13873

Lastly, estimated impact measures for each of the explanatory variables are presented in Table 4 by following the LeSage and Pace [12] and Elhorst's [17] approaches. While direct effects measure the impact of explanatory variables in the studied location, indirect effects measure the impact of explanatory variables in the studied location on the values in neighbouring locations [18]. As seen in Table 4, direct effects are more important than indirect effects for all variables. Thus, the values in a given area have bigger role than the value of its neighbourhood on IPS rate. Illiteracy rate has the maximum effect and labor rate has the minimum effect on IPS rate, this is similar to the estimated coefficients of SLM.

5. CONCLUSIONS

In this study, we have studied the participation rate within IPS via spatial statistics, spatial maps and spatial econometrics models. Firstly, we have observed that the participation within IPS has spatial structure in Turkey by means of spatial autocorrelation statistics and spatial maps. Also, global and local Moran's I statistics show that there is local and global autocorrelation between IPS rate due to spatial clustering of the west and east regions in Turkey. Particularly, the lowest participation rate within IPS are seen in the East and southeast Anatolia as indicated in [19]. Particularly, while Muş is the worst province as stated in [5], İstanbul, Ankara, Izmir, Muğla and Antalya are best ones in terms of IPS. When it is taken into account LISA, two clusters in Turkey are observed. West represents positive local autocorrelation and thus positive neighbouring effect. Also, various factor such as labor rate, unemployment rate, urbanization rate, sex ratio, the deposit rate, illiteracy rate and human development index are evaluated for modelling IPS rates. As well as these factor, demographic properties, age groups and also other factors can also affect the participation in IPS [8, 20]. On the other hand, it is clear that the impact of the policy shift on the voluntary participation in the system [21]. With these factors, linear model and also spatial econometrics models are used to model the participant rate in Turkey at provincial level since spatial autocorrelation for these variables are observed and non-normality occur in OLS residual. Such spatial analyses are important because classical regression models, which ignore specification of spatial effects, can lead to inaccurate inferences concerning predictor variables. The present analysis also points to a significant spatial dependence. The results of AIC and LR test indicate that SLM and SEM are more appropriate than classical regression model for this study to model the rate within IPS. Especially, it shows that the SLM is the best among the tested models with regard to all criteria and tests. Thus the spatial econometric model is strongly recommended for estimation of the variables concerning IPS.

ACKNOWLEDGEMENTS

REFERENCES

[1] Acikgoz E, Uygurturk H, Korkmaz T. Analysis of factors affecting growth of pension mutual funds in Turkey. International Journal of Economics and Financial Issues 2015; 5(2): 427-433.

[2] Cetin I, Sevuktekin M. Factors affecting people's entrance to individual pension system in Bursa. Journal of Accounting and Finans 2015; 171-191.

[3] Kara S, Yildiz Y, Karan MB. Analysis of risk-taking behaviour of individual pension system participants: the case of Turkey. Journal of Economics, Finance and Accounting 2015; 2(3): 375-396.

[4] Gokcen U, Atakan Yalcin A. The case against active pension funds: evidence from the Turkish private pension system. Emerging Markets Review 2015; 23: 46–67.

[5] Çetin I, Sevüktekin M. Türkiye'de bireysel emeklilik sistemi'nin iller bazinda mekansal ekonometrik analizi. In: IBAD 2016 1st International Scientific Researches Congress on Humanities and Social Sciences; 19-22 August 2016; Madrid, Spain.

[6] Sahin S, Rittersberger-Tilic H, Elveren, AY. The individual pension system in Turkey: a gendered perspective. Ekonomik Yaklasim 2016; 21(77): 115-142.

[7] Yildiz Y, Arslan-Ayaydin O, Karan MB. Payment persistence of participants in Turkish private pension scheme and gender differences. International Journal of Economics and Finance 2016; 8(10): 159-166.

[8] Çömlekçi İ, Gökmen O. Bireysel emeklilik sistemine katilmada etkili olan faktörler: TR42 bölgesinde bir araştirma. Journal Of International Social Research2017; 10(49): 579-588.

[9] Bauer R, Bonetti M, Broeders D. Pension funds interconnections and herd behaviour. DNB Working Paper No. 612, Amsterdam, 2018.

[10] Anselin L. Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic Publishers, 1988.

[11] Anselin L. Thirty years of spatial econometrics. Papers in Regional Science 2010; 89: 3-25. doi:10.1111/j.1435-5957.2010.00279.x

[12] Lesage JP, Pace RK. Introduction to Spatial Econometrics. USA: Taylor and Francis Group, 2009.

[13] Getis A, Aldstadt AJ. Constructing the spatial weights matrix using a local statistic. Geogr Anal 2004; 36(2): 90–104.

[14] Anselin L. Spatial dependence and spatial heterogeneity: Model specification issues in the spatial expansion paradigm. In Applications of the Expansion Method 2003; 264-279.

[15] Arbia G. Spatial Econometrics—Statistical Foundations and Applications to Regional Convergence, Berlin, Germany: Springer, 2006.

[16] Arbia G. Spatial Linear Regression Models. In: A Primer for Spatial Econometrics. London, UK: Palgrave Texts in Econometrics, 2014.

[17] Elhorst JP. Applied Spatial Econometrics: Raising the Bar. Spatial Economic Analysis 2010; 8:10: 9-28.

[18] Kopczewska K, Lewandowska A. The price for subway access: Spatial modelling of office rental rates in London. Urban Geography 2018; 39:10: 1528-1554.

[19] EGM. Emeklilik Gozetim Merkezi. Bireysel Emeklilik Sistemi Gelişim Raporu, http://www.egm.org.tr/bes2016gr.htm, Emeklilik Gozetim Merkezi, Istanbul, Turkey, 2016.

[20] Apak S. Türkiye'de bireysel emeklilik sisteminin gelişimi. Ekonomi Bilimleri Dergisi 2010; 2:2: 121-129.

[21] Ertuğrul HM, Gebesoglu PF, Atasoy BS. Mind the gap: Turkish case study of policy change in private pension schemes. Borsa Istanbul Review 2018; 18:2: 140-149.

[22] Yildirim V. Spatial econometric models: robust estimation for spatial error model. PhD, Anadolu University, Eskisehir, Turkey, 2018.