Journal of Advanced Research in Applied Mathematics and Statistics
Volume 5, Issue 1 - 2020, Pg. No. 1-13
Peer Reviewed Journal

Research Article

Determinants of Per Capita Personal Income in U.S. States: Spatial Fixed Effects Panel Data Modeling

Ahmed H Youssef[1], Mohamed R Abonazel[2], Ohood A Shalaby[3]

[1,2]Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt.
[3]National Center for Social and Criminological Research, Cairo, Egypt.
DOI: https://doi.org/10.24321/2455.7021.202001

I N F O

Corresponding Author:
Mohamed R Abonazel, Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt.
E-mail Id:
mabonazel@hotmail.com
Orcid Id:
https://orcid.org/0000-0001-6010-001X
How to cite this article:
Youssef AH, Abonazel MR, Shalaby OA. Determinants of Per Capita Personal Income in U.S. States: Spatial Fixed Effects Panel Data Modeling. *J Adv Res Appl Math Stat* 2020; 5(1): 1-13.

Date of Submission: 2020-02-05
Date of Acceptance: 2020-04-16

A B S T R A C T

Over the last decades, the Per Capita Personal Income (PCPI) variable was a common measure of the effectiveness of economic development policy. Therefore, this paper is an attempt to investigate the determinants of personal income by using spatial panel data models for 48 U.S. states during the period from 2009 to 2017. We utilize the three following models: spatial autoregressive (SAR) model, Spatial Error (SEM) Model, and Spatial Autoregressive Combined (SAC) model, with individual (or spatial) fixe deffects according to three different known methods for constructing spatial weights matrices: binary contiguity, inverse distance, and Gaussian transformation spatial weights matrix. Additionally, we pay attention for direct and indirect effects estimates of the explanatory variables for SAR, SEM, and SAC models. The second objective of this paper is to show how to select the appropriate model to fit our data.

The results indicate that the three used spatial weights matrices provide the same result based on goodness of fit criteria, and the SAC model is the most appropriate model among the models presented. However, the SAC model with spatial weights matrix based on inverse distance is better compared to other used models. Also, the results indicate that percentage of individuals with graduate or professional degree, real Gross Domestic Product (GDP) per capita,and number of nonfarm jobs have a positive impact on the PCPI, while the percentage of individuals without degree or bachelor's degree have a negative impact on the PCPI.

Keywords: Goodness of Fit Criteria, Gross Domestic Product, Labor Force, Maximum Likelihood, Spatial Autoregressive Combined Model, Spatial Error Model, Spatial Weights Matrix

Introduction

Spending and income are a traditional way to measure the economy's health. National personal income levels are very closely linked to the GDP, they work as a key indicator on consumption levels, inflationary pressures, and market conditions. Therefore, the ability to measure trends of income and spending is very important for investors because it is an indication about the overall economy power and future demand for both goods and services in the market.

Many states throughout the nation now include some

measure of income in their formal performance measures. The use of income measures in benchmarking economic development policy is attractive for a number of reasons. Economic theory suggests that wages are closely linked to individual productivity, and hence, are a potential measure of accumulated economic development efforts of all types. Like wise, higher personal income leads to an increased demand for goods and services, resulting, in part, in greater employment, investment, and production within a region. Therefore, for policymakers who wish to monitor and assess public policy, personal income is an attractive choice for informative, yet low-cost, data to collect and observe, see Hicks et al. (2013). Since the economic development of a region usually affects the surrounding area, therefore, when study the PCPI, we can include the spatial effects in the model, see Purwaningsih et al. (2017).

According to what is posted on U.S. Bureau of Economic Analysis (BEA) site, personal income statistics in USA tell us lot information about how U.S. workers and businesses are faring. In addition, these statistics help assess and compare the economic well-being of state residents. The PCPI, an area's personal income divided by its population, can be used to compare incomes from one state to another or to the nation overall.

In general, the "panel data" terminology refers to the pooling observations on a cross section of households, countries, firms, etc., over several time periods. See, e.g.,Baltagi (2008), Youssef and Abonazel (2017), Abonazel (2017, 2018, 2019), and Youssef et al. (2020) for details on panel data modeling. Panel data models have taken an important role in the literature of analyzing personal income determinants. Additionally, spatial data analysis was witnessed explosive growth in last decades in econometrics and in many applied fields. The attention to space, location, and interaction between them became an important feature of scholarly work because of the increased availability of data sets in which a number of spatial units are followed over time.

This paper is organized as follows: Section 2 reviews some literature about main determinants of the personal income. Section 3 presents the used framework of spatial panel data (SPD) models in our empirical study. Section 4 introduces SPD models (SAR, SEM, and SAC) with fixed effects (FE). Section 5 presents our empirical study on U.S. states. Finally, Section 6 includes the concluding remarks.

Main Determinants of Personal Income

In this section, we draw attention to some of the essential points to take into account when thinking about how to model the determinants of personal income.

Educational Attainment

The human capital is one of the most important ingredients of the economic modeling in much literature. It can be referred to human capital as all worker characteristics that can potentially increase the productivity and efficiency in the production process, in this context, there are much of the literature of labor economics which were shown the positive impact of human capital on individual earnings, see Card (1999) and Akgüç (2011). According to such studies, the pre-labor market investments in schooling potentially boost the individual earnings through increasing the productive skills. There are other studies concerned with investigating the causal relationships between family income and educational attainment, see Blanden and Gregg (2004).

Since the availability of the international educational attainment datasets, many empirical studies have used the educational attainment variables as a proxy for the human capital stock in country. Bennett (2018) mentioned that "Educational attainment is a significant determinant of wages, and the highest income earners in the U.S. states are often the most educated. An efficient educational system should integrate every student in the country into the labor force by ensuring they have the skills necessary to compete in the labor market and make a good wage".

Kalogirou and Hatzichristos (2007) presented a spatial modeling framework for income estimation in the municipality of Athens. They found that the data to be spatially autocorrelated, and the most interesting variable as the determinant of income was the proportion of people witha master's or doctorate degree.

Economy's Size

The GDP can give us an overall picture of the economy of the state. It is an indicator of an economy's size in a country, see Abonazel and Abd-Elftah (2019) and Abonazel and Rabie (2019). In addition to, economic prosperity can be measured as via GDP per capita (GDPPC). There are several empirical studies interested in investigating the casual relationship between GDP and income inequality, see, e.g., Brosio et al. (2016), Brueckner and Lederman (2017), and Chang et al. (2018).

Labor Force Type

Non-farm employment is believed to reduce poverty, and inequality in the distribution of incomes. Since farm employment is generally related with an increase drisk of poverty, Möllers and Buchenrieder (2011) studied the impact of rural non-farm employment on the income distribution among small family farms in transition in Croatia.

Tran (2015) studied the impact of non-farm employment on household income by using logistic regression analysis among ethnic minorities in the Northwest Mountain and Vietnam. Hisresults indicated that non-farm employment provides a windowto get out of poverty. There are many other studies confirmed the same results, e.g., De Janvry et

ISSN: 2455-7021
DOI: https://doi.org/10.24321/2455.7021.202001

al. (2005) explained that, without non-farm employment, poverty in rural regions in China would be much deeper, as well as, inequality in income distribution would be higher.

The Framework of Spatial Panel Data Models

Anselin (1988) defined spatial econometrics as a set of techniques that deal with the distinctive properties of space in the statistical analysis of regional science models. Elhorst (2014) also defined spatial econometrics which is a subfield of econometrics that deals with spatial interaction effects among geographical units, such as cities, regions, countries, and so forth depending on the nature of the study. Additionally, Lee and Yu (2010a) referred to theses interaction effects may be among economic units in space, where the space can be physical or economic in nature.

Since spatial econometrics deals with interaction effects among spatial units, such provinces, regions, etc. In modeling terminology, three different types of interaction effects are defined that explain why an observation associated with a particular location may be relay on observations in other locations (see Elhorst, 2014):

- Endogenous interaction effects among the dependent variable: which measures whether the dependent variable (y) of unit (A) depends on the dependent variables of other units (B) where (B≠A) and vice versa. This effect can be denoted by ($W_{NT}y$).
- Exogenous interaction effects among explanatory variables: which measures whether the dependent variable (y) of unit (A) depends on the explanatory variables (X) of other units (B) where (B ≠ A). This effect can be denoted by ($W_{NT}X$). Note that if the number of explanatory variables is (K), the maximum number of exogenous interaction effects is also (K).
- Interaction effects among the error terms (u): this effect can be denoted by ($W_{NT}u$) indicating that units may behave similarly because they share the same unobserved characteristics or face similar unobserved environments.

Model Specification

A full static model, with all types of interaction effects, takes the following form:

$$y_t = \lambda W_N y_t + X_t\beta + W_N X_t\theta + \mu + u_t; \qquad u_t = \rho W_N u_t + \varepsilon_t; \quad t = 1, \cdots, T, \tag{1}$$

where y_t is an $(N\times 1)$ vector consisting of one observation of the dependent variable for every spatial unit $(i = 1, \cdots, N)$ in the sample at time t, $W_N y_t$ indicates to the endogenous interaction effects, X_t is an (N×K) matrix of exogenous explanatory variables, $W_N X_t$ refers to the exogenous interaction effects, u_t is the error terms of the model, which is assumed to be serially correlated and to be spatially correlated, and $W_N u_t$ reflects the interaction effects among error terms. Where W_N is a (N×N) non-negative matrix of known constants describing the spatial arrangement of the units in the sample, λ is the spatial autoregressive coefficient, ρ is called the spatial autocorrelation coefficient, while Θ just as β, represents a (K×1) vector of fixed but unknown parameters, and μ is a (N×1) vector contains spatial specific effects. Model (1) can be rewritten as in the following reduced form:

$$y_t = S_N^{-1}\left(X_t\beta + W_N X_t\theta\right) + S_N^{-1}\mu + S_N^{-1}B_N^{-1}\varepsilon_t, \tag{2}$$

where

$$S_N = \left(I_N - \lambda W_N\right), \tag{3}$$

$$B_N = \left(I_N - \rho W_N\right). \tag{4}$$

Vega and Elhorst (2013) mentioned that there is a large gap between econometric theoreticians and practitioners in terms of interest level in these interaction effects. Theoreticians are mainly focused on the models containing interaction effects among endogenous variables or interaction effects among error terms, because of all the econometric problems associated with these models estimation process. Whereas the spatial econometric models with exogenous interaction effects do not suffer from any econometric problems, standard estimation techniques are sufficient under these circumstances. In contrast, practitioners often interest in the spatial model with exogenous interaction effects and take it as point of departure due to their focus on spillover effects. Table 1 summarizes linear spatial econometric models.

Table 1. SPD Models with Different Combinations of Spatial Interaction Effects

Model	Spatial Interaction Effects	
	Term	Number
SAR: Spatial Autoregressive Model	$W_N y_t$	1
SEM: Spatial Error Model	$W_N u_t$	1
SAC: Spatial Autoregressive Combined Model	$W_N y_t$ & $W_N u_t$	2
SLX: Spatial Lag of X Model	$W_N X_t$	K
SDM: Spatial Durbin Model	$W_N y_t$ & $W_N X_t$	K+1
SDEM: Spatial Durbin Error Model	$W_N X_t$ & $W_N u_t$	K+1
GNS: General Nesting Spatial Model	$W_N y_t$ & $W_N X_t$ & $W_N u_t$	K+2

ISSN: 2455-7021
DOI: https://doi.org/10.24321/2455.7021.202001

Spatial Weights Matrix

In order to take into account the spatial dependence in the regression model, a spatial weights matrix is used. It is an important component of spatial econometric models, which determines the intensity and the structure of the spatial dependence between locations exogenously.

Formally, the spatial weights matrix (W_N) is defined as a $(N \times N)$ positive matrix which has zero diagonal elements. Its rows and columns correspond to the cross-sectional observations. An element w_{ij} of the matrix represents the prior strength of the interaction between location (i) (row number i in W_N) and location (j) (column number j in W_N).

To generalize the usage of spatial weights matrix in a panel data setting, the weights are assumed to remain constant over time. Using the subscript to specify the dimension of matrix, with (W_N) as the weights for the cross-sectional dimension, the full $(NT \times NT)$ weights matrix then becomes:

$$W_{NT} = \begin{bmatrix} W_N & 0 & \cdots & 0 \\ 0 & W_N & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & W_N \end{bmatrix} = I_T \otimes W_N, \quad (5)$$

where $\otimes$ is Kronecker product.

The problem of choosing the optimal spatial weights matrix is still in the developing phase. There are several methods to define the spatial matrix. But, in our study, we only use the following three different methods to construct spatial weights matrix:

- The simplest version of a spatial weights matrix is a binary contiguity matrix; when two states share a common border, common vertex, or both, the corresponding entry in the spatial weights matrix is one and zero otherwise. The elements on the main diagonal are zero by definition. Anselin (1988) mentioned that this definition of contiguity obviously emphasizes the need for the existence of a map, which is used to discern the boundaries. This matrix induces a simple spatial structure that might not reflect actual spatial linkages in an appropriate way. Therefore, it can be constructed spatial weights matrices with general weights by utilizing data on geographic distances between states.
- The spatial relations can be based on a simple transformation by taking the inverse of the distance separating the cross-sectional observations. This method enables the construction of weights matrix that respects Tobler's law: the weights are greater (smaller) as the observations are spatially closer (further apart), for more details, see Dubé and Legros (2014).
- The relation who based on Gaussian transformation is not only dependent on the distance between observations, but also dependent on a threshold distance. The explicitly advantage of this transformation is taking into account a threshold distance.

This threshold distance can be specified as the average distance, the maximal distance, or even max-min criterion, which meaning that all the observations are taken into consideration in the construction of the spatial weights, for more details, see Dubé and Legros (2014). The threshold distance also can be used in case of inverse distance matrix, as mentioned in Figure 2.

Spatial Panel Data Model Assumptions

We can summarize the general assumptions of SPD models in the following (see Kapoor et al., 2007; Lee and Yu, 2010b):

Assumption 1: Spatial Weights Matrix

- W_N is a non-stochastic spatial weights matrix and row-sum normalized with zero diagonals.
- The admissible parameter space for the true spatial effects λ and ρ is $(-1,1)$.
- The spatial transformation matrix, e.g., $(I_N - \lambda W_N)$ is an invertible on the compact parameter spaces of spatial effects and their inverses are Uniformly Bounded (UB) in the parameter spaces.
- W_N is UB in both row and column sums in absolute value.

Assumption 2: The Error Components

The disturbances $\{\varepsilon_{it}\}$ are $i.i.d.$ across *i* and *t* with zero mean, finite variance (say σ^2), and their higher than fourth moments exist.

Assumption 3: Covariates

The regressors (X_t) are non-stochastic and have full rank and their elements are bounded, uniformly in absolute value.

Assumption 4: N and T

N is large, where T can be finite or large. The case of finite N and large T is of less interest as the incidental parameter problem doesn't occur in this model.

Spatial Fixed Effects Panel Data Models

In this section, we will discuss the main three (SAR, SEM, and SAC) models when the individual effects are fixed.

Spatial Autoregressive Model

Consider the following SARmodel with FE:

$$y_t = \lambda W_N y_t + X_t \beta + \mu + \varepsilon_t. \quad (6)$$

It can be written in a reduced form as:

$$y_t = S_N^{-1} X_t \beta + S_N^{-1} \mu + S_N^{-1} \varepsilon_t. \quad (7)$$

where S_N is defined in (3). By stacking the observations

ISSN: 2455-7021
DOI: https://doi.org/10.24321/2455.7021.202001

across individual and time, the model (6) can rewrite as:

$$y = \lambda\left(I_T \otimes W_N\right)y + X\beta + \left(l_T \otimes I_N\right)\mu + \varepsilon. \quad (8)$$

where l_T is a (T×1) vector of ones.

The consistent estimation of the individual FE is not possible in case of large N because of the incidental parameter problem. Therefore, Elhorst (2003) suggested estimation approach based on specific transformation to eliminating the time invariant individual FE, and then obtain consistent estimators. Instead of the demeaned variables, one may also use the original variables by using Q_0-transformation as follows:

$$y^* = \lambda\left(I_T \otimes W_N\right)y^* + X^*\beta + \varepsilon^*, \quad (9)$$

where

$$y^* = Q_0 y; \qquad X^* = Q_0 X; \qquad \varepsilon^* = Q_0 \varepsilon, \quad (10)$$

$$Q_0 = I_{NT} - \left(\frac{1}{T} l_T l_T^{'} \otimes I_N\right) \quad (11)$$

The focus in this section will base on the maximum likelihood (ML) estimation with transformation approach. The log likelihood function of model (9), as if the disturbances were normally distributed, is:

$$\ln L = -\frac{NT}{2}\ln\left(2\pi\sigma^2\right) + T\ln\left|S_N\right| - \frac{NT}{2\sigma^2}e'e, \quad (12)$$

where $e = y - \lambda\left(I_T \otimes W_N\right)y - X\beta$ and $\left|S_N\right|$ is the Jacobian determinant. This function is very similar to that derived for the SAR cross-section model which was proposed by Anselin and Hudak (1992).

Elhorst (2009) proposed a concentrated likelihood approach that can be maximized from residuals $\left(e_0^* \text{ and } e_1^*\right)$ of the Ordinary Least Squares (OLS) regression of y^* on X^* and the OLS regression of $\left(I_T \otimes W_N\right)y^*$ on X^*. Then the ML estimator of λ is obtained by maximizing the following concentrated log-likelihood function:

$$\ln L_{concentrated} = C + T\ln\left|S_N\right| - \frac{NT}{2}\ln\left[\left(e_0^* - \lambda e_1^*\right)'\left(e_0^* - \lambda e_1^*\right)\right], \quad (13)$$

where C is a constant not depending on λ. Unfortunately, this maximization problem can only be solved numerically, since a closed-form solution for λ not exist. Therefore, an iteration procedure must be used, which require λ be initially fixed to calculate $\hat{\beta}$ and σ^2. Finally, $\hat{\beta}$ and σ^2 are obtained from the first order conditions of the likelihood function by replacing λ with its estimated value from the ML.

Spatial Error Model

Consider the following SEM with FE:

$$y_t = X_t\beta + \mu + u_t; \qquad u_t = \rho W_N u_t + \varepsilon_t. \quad (14)$$

It can be written in a reduced form as:

$$y_t = X_t\beta + \mu + B_N^{-1}\varepsilon_t, \quad (15)$$

where B_N is defined in (4). By stacking the observations across individual and time, the model (14) can rewrite as:

$$y = X\beta + \left(l_T \otimes I_N\right)\mu + u; \qquad u = \rho\left(I_T \otimes W_N\right)u + \varepsilon. \quad (16)$$

The estimation strategy for the cross-sectional SEM, which proposed by Anselin and Hudak (1992), can be easily extended to the panel context. Again a concentrated likelihood approach can be taken but an iterative procedure is needed to estimate the parameters of the SEM. The model is transformed according to (10) to eliminate individual FE. The log-likelihood function of model (16) can be written as:

$$\ln L = -\frac{NT}{2}\ln\left(2\pi\sigma^2\right) + T\ln\left|B_N\right| - \frac{1}{2\sigma^2}e'\left(I_T \otimes B_N^{'}B_N\right)e, \quad (17)$$

where e = y - Xβ.

Elhorst (2009) proposed the following concentrated log-likelihood function of ρ:

$$\ln L_{concentrated} = C + T\ln\left|B_N\right| - \frac{NT}{2}\ln\left[e(\rho)'e(\rho)\right], \quad (18)$$

where

$$e(\rho) = y^* - \rho\left(I_T \otimes W_N\right)y^* - \left[X^* - \rho\left(I_T \otimes W_N\right)X^*\right]\beta. \quad (19)$$

Maximizing this function with respect to ρ yields the ML estimator of ρ, given β and σ^2. An iteration procedure must be used in which the set of parameters β and σ^2.

According to Millo and Piras (2012), the estimation procedure of SEM with individual FE can be summarized in the following steps:

Step 1: Estimated OLS residuals (of the transformed model) can be used to obtain an initial estimate of ρ.

Step 2: The initial estimate of ρ can be usedto compute a (spatial) feasible Generalized Least Square (GLS) estimator of the regression coefficients, the errorvariance and a new set of estimated GLS residuals.

Step 3: An iterative procedure may then beem ployed: the concentrated likelihood and the GLS estimators are alternately computed until convergence.

Spatial Autoregressive Combined Model

Consider the following SAC model with individual FE:

$$y_t = \lambda W_N y_t + X_t\beta + \mu + u_t;\ u_t = \rho W_N u_t + \varepsilon_t. \quad (20)$$

It can be written in a reduced form as:

$$y_t = S_N^{-1}X_t\beta + S_N^{-1}\mu + S_N^{-1}B_N^{-1}\varepsilon_t. \quad (21)$$

Lee and Yu (2010b) used the transformation, $J_T = \left(I_T - \frac{1}{T}l_T l_T^{'}\right)$ (the deviation from the time mean operator) which is defined in (11), to eliminate the individual effects. Because W_N is time invariant, the variables in the deviation form would still be a SAR.

Let $\left(F_{T,T-1}, \frac{1}{\sqrt{T}}l_T\right)$ be the orthonormal matrix of the eigenvectors of J_T, where $F_{T,T-1}$ is $T(T-1)$ eigenvector matrix corresponding to the eigenvalues of one and $\frac{1}{\sqrt{T}}l_T$ is the T-dimensional

ISSN: 2455-7021
DOI: https://doi.org/10.24321/2455.7021.202001

column vector corresponding to the eigenvalues of zero. For any $T(T-1)$ matrix,

$$\left(y_1^{**}, \ldots, y_{T-1}^{**}\right) = \left(y_1, \ldots, y_T\right) F_{T,T-1}, \quad (22)$$

then equation (20) implies:

$$y_t^{**} = \lambda W_N y_t^{**} + X_t^{**}\beta + u_t^{**}; \; u_t^{**} = \rho W_N u_t^{**} + \varepsilon_t^{**}; t = 1, \ldots, T-1.$$

In this model, we assume that:

$$E\left(\varepsilon_1^{**'}, \ldots, \varepsilon_{T-1}^{**'}\right)'\left(\varepsilon_1^{**'}, \ldots, \varepsilon_{T-1}^{**'}\right) = \sigma^2\left(F_{T,T-1}' \otimes I_N\right)\left(F_{T,T-1} \otimes I_N\right) = \sigma^2 I_{N(T-1)}, \quad (23)$$

The log likelihood function of (23), as if the disturbances were normally distributed, is:

$$\ln L = -\frac{N(T-1)}{2}\ln\left(2\pi\sigma^2\right) + (T-1)\left(\ln|S_N| + \ln|B_N|\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^{**'}\varepsilon_t^{**}, \quad (24)$$

where

$$\varepsilon_t^{**} = B_N\left(S_N y_t^{**} - X_t^{**}\beta\right). \quad (25)$$

Empirical Study

Annual data for 48 U.S. states are used during the period from 2009 to 2017. We employ three specifications of SPD models (SAR, SEM, and SAC models) utilizing three different known methods for constructing spatial weights matrices as shown in Figure 2. In addition to that we pay attention for estimates of direct and indirect effects of explanatory variables in SAR, SEM and SAC models.

The second objective of this study is to show how to choose the appropriate model to fit our data. It is necessary to know whether or not there are spatial effects, and if so, as a second step, it is determined the type of spatial interaction effects should be taken into account for (i) the spatially autocorrelated error term, (ii) the spatially lagged dependent variable or (iii) combination of them.

Data Description

As an empirical application, this paper is concerned with studying the significant impact of three dimensions on the PCPI using data for 48 U.S. states during period from 2009 to 2017. The dataset is limited by the amount of information available for each state involved. We selected a variety of variables that have been shown to correlate with the PCPI

Table 2. Definition of the Variables

Dimension	Variable Name	Name of Variable on the Site	Definition	Measuring Unit	Source
Dependent Variable	PCPI	Per capita personal income	Per capita personal income	1000 dollars	U.S. BEA
Educational Attainment	ND	Some college, no degree	Percentage of individuals without degree	%	U.S. Census Bureau
	BD	Bachelor's degree	Percentage of individuals with bachelor's degrees	%	
	GD	Graduate or professional degree	Percentage of individuals with graduate or professional degree	%	
Economy's Size	GDPPC	Real GDP per capita by state	Real GDP per capita	1000 dollars	U.S. BEA
Labor Force	Population	Population	Number of Population	100000 persons	
	NonFarm	Nonfarm employment	Number of nonfarm jobs		

Table 3. Descriptive Statistics of the Variables

Variable	Mean	Standard Deviation	Minimum	Maximum
PCPI	44.18	7.88	29.94	72.11
ND	21.43	3.25	0.00	27.60
BD	17.79	3.06	0.00	24.80
GD	10.34	2.72	0.00	18.70
GDPPC	49.82	9.41	33.15	76.36
Population	65.28	70.63	5.60	394.00
NonFarm	37.19	39.78	3.72	233.48

ISSN: 2455-7021
DOI: https://doi.org/10.24321/2455.7021.202001

in previous researches. The used softwares in our study are "STATA version 15" and "R version 3.6.1". Table 2 displays the definition of he used variables, and some descriptive statistics of these variables have been presented in Table 3.

To visualize regional differences in income per capita of U.S. states, Figure 1 presents a map of USA which is colored according to the extent of regional income per capita in 2017. Based on these exploratory tools, there is substantial variation in regional income per capita in U.S. states. In 2017, the state with lowest personal income exhibited a 36,567 U.S. dollar per capita (Mississippi) while the highest regional income amounted to 72,110 U.S dollar per capita (Connecticut).

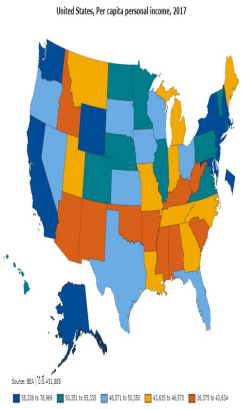

Figure 1. Per Capita Personal Income according to State in 2017

Testing the Multicollinearity of Explanatory Variables

The first step of data processing is to try to ensure that there is no high linear correlation between two or more explanatory variables. Where statistical inferences are not reliable in the case of multicollinearity, because it makes estimates of the regression coefficients inaccurate, inflates their standard errors, deflates the partial t-tests for them, gives false non-significant p-values, and reduces the predictability of the model, see Studenmund (2016). We use the most common method to detect multicollinearity: (i) Pearson correlation matrix between each pair of predictor variables and (ii) the Variance Inflation Factor (VIF) based on there sults of the pooled OLS model, see Paul (2006) and Youssef et al. (2020).

Table 4 shows that there is strong correlation among the variables "Population and NonFarm" greater than 0.99. Additionally, theresults of VIF in the first time with all

Table 4. Pearson Correlation Matrix and VIF

	ND	BD	GD	GDPPC	Population	NonFarm
ND	1					
BD	0.074	1				
GD	-0.308***	0.733***	1			
GDPPC	-0.179***	0.572***	0.551***	1		
Population	-0.157***	0.088	0.1491**	0.188***	1	
NonFarm	-0.169***	0.124**	0.181***	0.227***	0.997***	1
VIF1	1.59	3.27	3.57	1.90	305.63	28061.42
VIF2	1.47	3.10	3.06	1.66	----	1.08

Notes: VIF1: is VIF for all variables, VIF2: is VIF after removing "Population".The superscripts *** and ** indicate statistical significance at the 0.001 and 0.01 level, respectively.

Table 5. Summary Statistics of the Straight Line Geographic Distances between Centroids of U.S. States, in kilometers

Min. of all Distances	Mean of all Distances	Max. of all Distances	Std. Dev. of all Distances
80.41	1676.80	4231.84	948.20
Links	W1: Binary Contiguity	W2: Inverse Distance	W3: Gaussian Transformation
Total number of links	214	210	210
Min. number of links	1	1	1
Mean number of links	4.46	4.38	4.38
Max. number of links	8	11	11

Source: https://www.mapdevelopers.com/distance_from_to.php (Access Date: 29/8/2019)

ISSN: 2455-7021
DOI: https://doi.org/10.24321/2455.7021.202001

regressors (VIF1) confirmed that there is multicollinearity problem between the regressors; where in most of empirical studies, the general rule of thumb is that VIF values exceeding 5 need further investigation, while VIF values exceeding 10 indicate to serious multicollinearity requiring correction. According to Paul (2006), if there are two regressors are almost linearly correlated, eliminating one regressor may be useful in combating multicollinearity. Therefore, we drop one variable (Population) from the model, the new results of VIF (VIF2) confirmed that there is no multicollinearity problem because all values of VIF2 less than 5.

Spatial Weights Matrix

In order to account for spatial dependence in the regression model, we utilize the following three different ways to construct spatial weights matrix (see Figure 2): We use a binary contiguity spatial weights matrix for the U.S. states, which is available in the splm package (spatial panel data models in R), see Millo and Piras (2012). While we use the

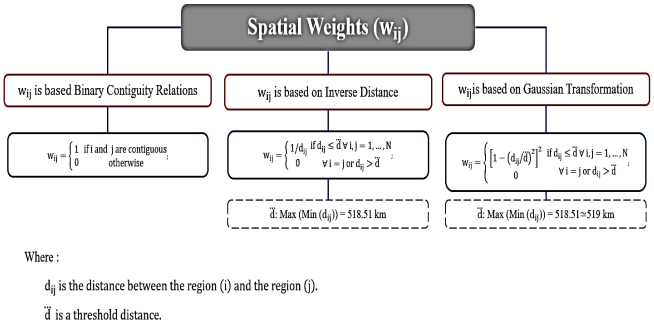

Figure 2. Spatial Weights Matrices in Our Empirical Study

data of the straight-line geographical distances between centroids of U.S. states, which are summarized in Table 5, to create the spatial weights matrices based on inverse distance and Gaussian transformation. The three used spatial weights matrices are row standardized to facilitate interpretation, see Lottmann (2012).

Framework of Spatial Econometric Modeling

We analyze determinants of regional differences for PCPI among U.S. states by using SAR, SEM, and SAC models with spatial weights matrices mentioned in Figure 2. The regional study of income levels may mean that they correlated over space. The presence of spatial autocorrelation implies that the level of personal income in one particular region is correlated with that of neighboring regions. The spatial econometric literature proved that ignoring spatial interaction effects make estimates are biased and inefficient, see Anselin and Bera (1998).

The pooled OLS regression model will yield biased parameter estimates in case of SAR and SLX. However, OLS estimation will produce unbiased and inefficient estimates for SEM. Where ignoring of spatially lagged variable is similar to an omitted variable bias, see Franzese and Hays (2007). As OLS estimation for SEM will lead to inconsistent estimates, see Franzese and Hays (2007), and Anselin and Bera (1998).

In this study, we base on the following framework proposed in Figure 3. To capture spatial dependence in the data, we utilize spatial panel approach. The spatial econometric literature provides various models for data with spatial autocorrelation: the model with spatially lagged dependent variable, SAR, spatially lagged in the error term, SEM,

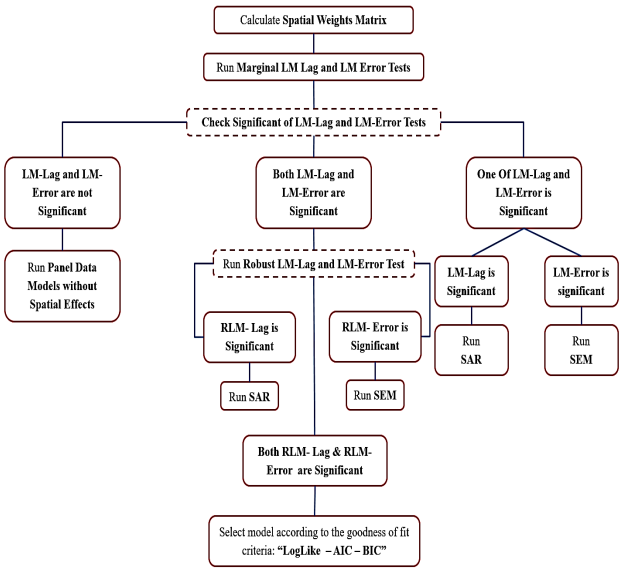

Figure 3. Framework of Our Empirical Study

spatially lagged explanatory variables, SLX, and combinations of them. The SLX model is the simplest model because the additional regressors are exogenous and the error term remains spherical. It does not suffer from any econometric problems. Therefore, we will focus in this paper on SAR, SEM, and SAC models.

Testing the Spatial Dependence

To test the spatial interaction effects in case of a cross-sectional data, Burridge (1980) and Anselin (1988) proposed Lagrange Multiplier (LM) tests for a spatially lagged dependent variable and for spatial error correlation term. Anselin et al. (1996) also proposed robust LM tests for a spatially lagged dependent variable in the local presence of spatial error autocorrelation and for spatial error autocorrelation in the local presence of a spatially lagged dependent variable. These tests are frequently used in empirical studies. Recently, Anselin et al. (2008) developed the classical LM tests for a spatial panel, and Elhorst (2009) developed the robust counterparts of these LM tests for a spatial panel. The classical and robust LM tests rely on the residuals of the non-spatial model and follow a Chi-square distribution with one degree of freedom.

After estimation of a pooled OLS model, spatial dependence

ISSN: 2455-7021
DOI: https://doi.org/10.24321/2455.7021.202001

Table 6. Results of Classical and Robust LM Tests with Different Spatial Weights Matrices

Test	W1: Binary Contiguity		W2: Inverse Distance		W3: Gaussian Transformation	
	Lag	Error	Lag	Error	Lag	Error
LM	149.49***	192.98***	133.62***	137.96***	105.11***	109.30***
Robust LM	22.53***	66.03***	25.53***	29.88***	20.63***	24.82***

Note: The superscript *** indicates statistical significance at the 0.001 level.

tests (LM-lag and LM-error tests) can be applied to specify whether estimation of a spatial model is warranted. The SEM and SAR models can be compared as a robustness check. If the LM test is rejected for the absence of spatial lag or spatial error in the model, it proves that a spatial panel model is the suitable method for the analysis.

The LM-lag test is used to determine if there is any ignored spatial autocorrelation in the dependent variable, while the LM-error test is employed to detect spatial dependence in error terms. Both of these tests are one-way tests in that they only considered the specific type of spatial dependence that is being tested. The robust varieties of these tests (robust LM-lag and robust LM-errortests) are designed to take into account the other type of spatial dependence, that is, the robust LM-lag test takes into account the potential presence of spatial dependence in the error term, while the robust LM-error test takes into account the potential of spatial dependence in the dependent variable.

Table 6 shows that classic and robust LM tests of spatial lag and spatial error terms are significant at the 0.001 level. However, the spatial error term is more significant than the spatial lag term in the three types of spatial weights matrices. So, we will include spatial lag and spatial error terms in our model, and then select the best model based on goodness of fit criteria.

Model Selection

As in a lot of empirical research, the models are comparable in terms of the goodness of fit criteria, such as the Log-pseudo likelihood, Overall R-Squared, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC), see Barufi et al. (2012). However, AIC and BIC are of the

Table 7. Results of Estimated SPD Models with Individual FE

	FE without Spatial	W1: Binary Contiguity			W2: Inverse Distance			W3: Gaussian Transformation		
		SAR	SEM	SAC	SAR	SEM	SAC	SAR	SEM	SAC
ND	-1.57***	-0.26**	-0.92***	-0.13	-0.27***	-0.11***	-0.20*	-0.34***	-1.26***	-0.23**
BD	-1.11**	-0.51*	0.16	-0.40*	-0.69***	0.60**	-0.73***	-0.78***	0.83***	-0.86***
GD	5.95***	1.59***	1.96***	1.10***	1.80***	1.52***	1.71***	2.15***	1.60***	2.02***
GDPPC	0.44***	0.31***	0.28***	0.30***	0.30***	0.30***	0.29***	0.33***	0.34***	0.29***
Non Farm	0.16***	0.13***	0.11***	0.12***	0.13***	0.09***	0.12***	0.13***	0.08***	0.12***
Mean of FE	----	-7.90	23.09	-9.56	-6.09	22.08	-6.34	-5.43	19.86	-5.39
λ	----	0.68***	----	0.74***	0.67***	----	0.70***	0.62***	----	0.67***
ρ	----	----	0.90***	-0.31***	----	0.87***	-0.11*	----	0.84***	-0.17*
Log-like	-792 .274	-508. 843	-529. 617	-545. 906	-545.906	-545.906	-542. 372	-542. 372	-567 .345	-567.35
R-Squared	0.5400	0.5018	0.5700	0.5122	0.4346	0.6137	0.4149	0.4695	0.6487	0.4331
AIC	1596.54	1031.68	1073.23	1029.16	1022.91	1091.06	1022.88	1067.47	1139.18	1067.28
BIC	1620.96	1059.34	1100.88	1060.77	1059.67	1118.71	1054.48	1099.13	1166.83	1098.88

Notes: Within estimator is used in case of FE without spatial effects. ML estimation transformed approach is used in case of SPD models. The superscripts ***, **, and * indicate statistical significance at the 0.001, 0.01 and 0.05 level, respectively

ISSN: 2455-7021
DOI: https://doi.org/10.24321/2455.7021.202001

best methods to select the most adequate weighting matrix according to the simulation study of Herrera et al. (2019).

Table 7 shows that the values of AIC and BIC of SAC models are smaller than the values of AIC and BIC of SLM and SEM models in most cases of used spatial weights matrices. Therefore, we can say that the SAC model is more appropriate than SLM and SEM models for our data. On purely statistical grounds, all three SAC models for the three types of spatial weights matrices seem to provide very similar levels of goodness of fit criteria. However, the SAC model with spatial weights matrix based on inverse distance returns a slightly higher Log-pseudo likelihood, and lower AIC and BIC. On the other hand, the SAC model with binary contiguity matrix records the higher R-Squared.

Interpretation of the Results

One should remember that the traditional interpretation of the regressioncoefficient; the effect of the explanatory variable on a dependent variable, is only valid in case of regression models without spatial interaction effects. However, in models with a spatial interaction effects (such as: SAR, SAC, SDM, and GNS) in order to fully explain the effect of changes, direct and indirect effects must be calculated and interpreted as model coefficients. This is because of appearing y in both sides of the equation, where y appears on the left-hand side and λWy appears on the right-hand side, for more details, see LeSage and Pace (2009) and Kopczewska et al. (2017).

LeSage and Pace (2009) and Elhorst (2010) defined the spatial direct effects as an impact of change of X_k in i location on change of y in i location, on the other hand, the indirect effect (spillover effect) is an impact of change of X_k in i location on change of y in j location $(i \neq j)$. Where these effects can be derived from the partial derivatives matrix of the expected value of y_t with respect to the k^{th} explanatory variable of X_t:

$$M = \begin{bmatrix} \frac{\partial E(y_{1t})}{\partial x_{1kt}} & \cdots & \frac{\partial E(y_{1t})}{\partial x_{Nkt}} \\ \vdots & \vdots & \vdots \\ \frac{\partial E(y_{Nt})}{\partial x_{1kt}} & \cdots & \frac{\partial E(y_{Nt})}{\partial x_{Nkt}} \end{bmatrix}. \quad (27)$$

Since both the direct and indirect effects are different for each unit in the sample, the presentation of these effects is considered a problem. In our case, there are 48 states (spatial units) and five explanatory variables; we obtain five different squared matrices (of order 48) of direct and indirect effects. To reduce the number of these matrices for direct and spillover effects, LeSage and Pace (2009) proposed to report one summary indicator for the direct effect measured by the average of the diagonal elements of M, and one summary indicator for the spillover effect measured by the average row or column sums of the off-diagonal elements in the same matrix M in (27). The average row effect represents the impact on a specific element of the dependent variable as a result of a unit change in all elements of an exogenous variable, while the average column effect represents the impact of changing a specific element of an exogenous variable on the dependent variable of all other units.

In SAC model, the direct effect is the average of main diagonal elements of the following matrix, and the indirect effect is the average of row off-diagonal elements in the same matrix:

$$M_{SAC} = (I_N - \lambda W_N)^{-1} \beta_k I_N. \quad (28)$$

Based on the above, the estimates of SAC model cannot be interpreted as partial derivatives in the traditional regression model. In our study, to assess the magnitudes of impacts arising from changes in the five explanatory variables, Table 8 provides the measures of direct, indirect, and the total effect for each variable.

Table 8. Direct and Indirect Effects of SAC model usingW2

Variable	Direct Effect	Indirect Effect	Total effect
ND	-0.2675**	-0.4166*	-0.6841*
BD	-0.9504***	-1.4803***	-2.4306***
GD	2.2326***	3.4774***	5.7100***
GDPPC	0.3813***	0.5939***	0.9751***
NonFarm	0.1599***	0.2491***	0.4091***

Notes: The superscripts ***, **, and * indicate statistical significance at the 0.001, 0.01 and 0.05 level, respectively.

The first column of Table 8 contains the direct effects, which measure how much the dependent variable (PCPI) changes in a state when a particular explanatory variable changes in that same state. We note that all five included explanatory variables are significant at the 0.01 level or less.

The second column in Table 8 shows the indirect or spillover effects of changes in our explanatory variables. It is important to note that the scalar summaryused to calculate the indirect effects summarizes spillovers over all states in the sample. We note that all five included explanatory variables are significant at the 0.001 level and one of them is significant at 0.05.

Finally, the last column of Table 8 shows the point estimates for the total effects, which are defined as the sum of the direct and indirect effects. We can note that all total effects are statistically significant at different significance levels (0.05 and 0.001).

ISSN: 2455-7021
DOI: https://doi.org/10.24321/2455.7021.202001

In general, we conclude the following notes from Table 8:

- The direct effect of increasing ND in specific state by 1% directly decreases PCPI by 267.5 dollars in the same state. In addition, the indirect effect of increasing ND in neighboring states has negative effect on the PCPIby 416.6 dollars. The total effect from ND is negative and consists mostly of indirect impact. In other words, the total effect of ND has a point estimate of -0.6841 which indicates that a 1% increase in ND in the own and surrounding states decreases the PCPI by 684.1 dollars.
- The impact of BD has a negative direct and indirect effect on the PCPI, indicating that we expect an increase in the PCPI in states with a low level of BD. The indirect effect resulting from BD in near by states is close to twice the magnitude of the direct effect, indicating a large spillover effect from BD. The total effect is negative, with about two-third of the spillover effects of BD in neighboring states.
- The direct and indirect effects of GD are positive (as expected); this means that the increase in GD in a specific state by 1% directly increases the PCPI by 2232.6 dollars in the same state and indirectly increases it in other states by 3477.4 dollars.
- The direct and indirect effects of GDPPC are positive (as expected); with the increase of the GDPPC by 1000 dollar in a particular state, the PCPI will increase by 381.3 dollars on average in the same state, and increase by 593.9 dollars on average in other states.
- The impact of NonFarm has a positive direct and indirect effect on the PCPI; when the number of nonfarm jobs increases by 100000 in a particular state, the PCPI increases by 159.9 dollars in the same state, and increase by 249.1 dollars on average in other states.

Summary and Conclusion

This paper is an attempt to investigating of the PCPI determinants. We utilize SPD approach by using annual data for 48 U.S. states over the period from 2009 to 2017 based on U.S. Census Bureau and Bureau of Economic Analysis. We employ the different three spatial models (SAR, SEM, and SAC) with individual FE according to three different known methods for constructing spatial weights matrices (binary contiguity, inverse distance and Gaussian transformation spatial weights matrix). To model regional income, we use five explanatory variables according previous empirical researches. These variables are: the percentage of people with (no degree, bachelor's degree, and graduate or professional degree), real GDP per capita, and number of non-farm jobs. We can summarize our empirical study in the following:

- For each type of spatial weight matrices, we test the spatial interaction effects by using classical LM-Lag and LM-Error tests, the results indicate to reject the null hypothesis of the absence of a spatial lag term and spatial error term at the 0.001 level of significance in all our models.
- Robust counterparts of LM tests have been applied. The results show that both of robust LM-Lag and robust LM-Error test are significant at the 0.001 level. Therefore, we included spatial lag and spatial errors terms in the model. This means that the SAC models are the appropriate models of our data.
- the most appropriate model to fit our data is selected based on goodness of fit criteria (Log-pseudo likelihood, overall R-squared, AIC, and BIC). The results show that the SAC model with inverse distance matrix is the best model; because it has the smallest values of AIC and BIC.
- We estimate the direct and indirect effects of final selected model to for interpretation purposes. The results show that the direct and indirect effects of all five included explanatory variables are significant at the 5% level or less.

Finally, we point to the need for more studies to discover more effective mechanisms that help countries in improve the population's standard of living and their personal incomes. We suggest studying the impact of other economic indicators on the PCPI, such as "the governmental expenditure on education, and foreign direct investment". Future research also could follow a number of promising paths for identifying the underlying determinants of the PCPI at state-level in USA, by using random effects SPD approach.

References

1. Abonazel MR. Bias Correction Methods for Dynamic Panel Data Models with Fixed Effects. *International Journal of Applied Mathematical Research* 2017; 6(2): 58-66.
2. Abonazel MR. Different Estimators for Stochastic Parameter Panel Data Models with Serially Correlated Errors. *Journal of Statistics Applications and Probability* 2018; 7(3): 423-434.
3. Abonazel MR. Generalized Estimators of Stationary Random-Coefficients Panel Data Models: Asymptotic and Small Sample Properties. *Revstat Statistical Journal* 2019; 17(4): 493–521.
4. Abonazel MR, Abd-Elftah AI. Forecasting Egyptian GDP Using ARIMA Models. *Reports on Economics and Finance* 2019; 5(1): 35 – 47.
5. Abonazel M, Rabie A. The Impact of Using Robust Estimations in Regression Models: An Application on the Egyptian Economy. *Journal of Advanced Research in Applied Mathematics and Statistics* 2019; 4(2): 8-16.
6. Akgüç M. The Effects of Different Stages of Education on Income across Countries. Working paper, *Toulouse*

ISSN: 2455-7021
DOI: https://doi.org/10.24321/2455.7021.202001

School of Economics TSE, 2011: 1-28.
7. Anselin L, Hudak S. Spatial Econometrics in Practice: A Review of Software Options. *Regional Science and Urban Economics* 1992; 22(3): 509-536.
8. Anselin L. Spatial Econometrics: Methods and Models. *Springer Science and Business Media*, 1988.
9. Anselin L, Bera AK. Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics, in Handbook of Applied Economic Statistics, *New York* and NY, 1998: 237-289.
10. Anselin L, Bera AK, Florax R et al. Simple Diagnostic Tests for Spatial Dependence. *Regional Science and Urban Economics* 1996; 26(1): 77-104.
11. Anselin L, Gallo J, Jayet JH. Spatial Panel Econometrics. In The Econometrics of Panel Data. *Springer,* Berlin, Heidelberg, 2008.
12. Baltagi B. Econometric Analysis of Panel Data. *John Wiley & Sons,* 2008.
13. Barufi AM, Haddad E, Paez A. Infant Mortality in Brazil, 1980-2000: A Spatial Panel Data Analysis. *BMC Public Health* 2012; 12(1): 181.
14. Bennett P. Examining the Effects of Education on Financial Savings Behavior, 2018.
15. Blanden J, Gregg P. Family Income and Educational Attainment: A Review of Approaches and Evidence for Britain. *Oxford Review of Economic Policy* 2004; 20(2): 245-263.
16. Brosio G. Jiménez JP, Ruelas I. Looking at the Nexus between Personal Income Distribution and Regional GDP Inequality in Decentralized Systems. Documento Presentado en las Quintas Jornadas Iberoamericanas de Financiación Local, Santiago de Compostela, 2016.
17. Brueckner M, Lederman D. Inequality and GDP Per Capita: The Role of Initial Income, World Bank, 2017.
18. Burridge P. On the Cliff-Ord Test for Spatial Correlation. *Journal of the Royal Statistical Society: Series B* (Methodological) 1980; 42(1): 107-108.
19. Card D. The Causal Effect of Education on Earnings, Handbook of Labor Economics, 1999; 3: 1801-1863.
20. Chang S, Gupta R, Miller SM. Causality between Per Capita Real GDP and Income Inequality in the US: Evidence from a Wavelet Analysis. *Social Indicators Research* 2018; 135(1): 269-289.
21. De Janvry A, Sadoulet E, Zhu N. The Role of Non-farm Incomes in Reducing Rural Poverty and Inequality in China, 2005.
22. Dubé J, Legros D. Spatial Econometrics Using Microdata. *John Wiley & Sons*, 2014.
23. Elhorst JP. Specification and Estimation of Spatial Panel Data Models. *International Regional Science Review* 2003; 26(3): 244-268.
24. Elhorst JP. Spatial Panel Data Models. Handbook of Applied Spatial Analysis. *Springer*, 2009.
25. Elhorst JP. Applied Spatial Econometrics: Raising the Bar. *Spatial Economic Analysis* 2010, 5(1): 9-28.
26. Elhorst JP. Spatial Econometrics: from Cross-Sectional Data to Spatial Panels. Heidelberg: *Springer*, 2014.
27. Franzese RJ, Hays JC. Spatial Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data. *Political Analysis* 2007; 15: 140-164.
28. Herrera M, Mur J, Ruiz M. A Comparison Study on Criteria to Select the Most Adequate Weighting Matrix. *Entropy* 2019; 21(2): 160.
29. Hicks MJ, Devaraj S, Faulk D et al. The Causes of State Differences in Per Capita Income: How Does Indiana Fare. Ball State University Center for Business and Economic Research, 2013.
30. Kalogirou S, Hatzichristos T. A Spatial Modelling Framework for Income Estimation. *Spatial Economic Analysis* 2007; 2(3): 297-316.
31. Kapoor M, Kelejian HH, Prucha IR. Panel Data Models with Spatially Correlated Error Components. *Journal of Econometrics* 2007; 140(1): 97-130.
32. Kopczewska K, Kudła J, Walczyk K. Strategy of Spatial Panel Estimation: Spatial Spillovers between Taxation and Economic Growth. *Applied Spatial Analysis and Policy* 2017; 10(1): 77-102.
33. Lee LF, Yu J. Some Recent Developments in Spatial Panel Data Models. *Regional Science and Urban Economics* 2010a; 40(5): 255-271.
34. Lee LF, Yu J. Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects. *Journal of Econometrics* 2010b; 154(2): 165-185.
35. LeSage J, Pace RK. Introduction to Spatial Econometrics. Boca Raton: CRC Press, Taylor and Francis Group, 2009.
36. Lottmann F. Explaining Regional Unemployment Differences in Germany: A Spatial Panel Data Analysis. No. 2012-026. SFB 649 Discussion Paper, 2012.
37. Millo G, Piras G. splm: Spatial Panel Data Models in R. *Journal of Statistical Software* 2012; 47(1): 1-38.
38. Möllers J, Buchenrieder G. Effects of Rural Non-farm Employment on Household Welfare and Income Distribution of Small Farms in Croatia. *Quarterly Journal of International Agriculture* 2011; 50(892-2016-65199): 217-235.
39. Paul RK. Multicollinearity: Causes, Effects and Remedies. M. Sc. Thesis, Indian Agricultural Statistics Research Institute, New Delhi, India, 2006.
40. Purwaningsih T, Ghosh A, Chumairoh C. Spatial Data Modeling in Disposable Income per Capita in China Using Nationwide Spatial Autoregressive (SAR). *International Journal of Advances in Intelligent Informatics* 2017; 3(2): 98-106.
41. Studenmund AH. Using Econometrics: A Practical Guide.

ISSN: 2455-7021
DOI: https://doi.org/10.24321/2455.7021.202001

7th Edition, Pearson 2016.
42. Tran TQ. Nonfarm Employment and Household Income among Ethnic Minorities in Vietnam. *Economic Research-Ekonomska Istraživanja* 2015; 28(1): 703-716.
43. Vega SH, Elhorst JP. On Spatial Econometric Models, Spillover Effects, and W. In 53rd ERSA Congress, Palermo*,* Italy, 2013.
44. Youssef AH, and Abonazel MR. Alternative GMM Estimators for First-Order Autoregressive Panel Model: An Improving Efficiency Approach. *Communications in Statistics-Simulation and Computation* 2017; 46(4): 3112-3128.
45. Youssef AH, Abonazel MR, Ahmed EG. Estimating the Number of Patents in the World Using Count Panel Data Models. *Asian Journal of Probability and Statistics* 2020; 6(4): 24-33.

ISSN: 2455-7021
DOI: https://doi.org/10.24321/2455.7021.202001