Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Estimation of the Durbin Spatial Semi-Parametric Regression Model using the Sub-Segment Regression Method

Anwar Sabah Aghmais[1,*], Ahmed Abd Ali Akkar[1]

[1] Statistics Department, College of Management and Economic, Mustansiriyah University, Iraq

ARTICLE INFO

Article history:
Received 5 October 2023
Received in revised form 5 July 2024
Accepted 28 August 2024
Available online 1 October 2024

Keywords:
Durbin regression; Spatial semi-parametric model; Sub-segment regression method

ABSTRACT

This study dealt with the penalty segment (PS) regression method to estimate the smoothing functions m(x) and m(x*), the Durbin semi-parametric spatial regression model, using the modified spatial adjacency matrix, under the standard adjoining rock, this method is one of the important non-parametric methods, were used to estimate non-parametric regression functions, it was applied using real-world data for a sample size of 75 observations, to find out the effect of average value, and it was analysed The research covered a period of five consecutive years (2018-2022), during which time it encompassed all 15 governorates of Iraq (with the exception of the three governorates that make up the Kurdistan Region). In consideration of the geographical impacts that exist across the governorates, the findings demonstrated that there was a discernible relationship between the rate of electrical loads and the temperature as well as the population density.

1. Introduction

The topic of spatial semi-parametric regression models, has received a lot of attention lately by researchers, as it is considered one of the important standard models that are used in several fields (economic, geographical, health, social.... etc.) [1]. After the science of statistics was limited to analysing data temporally, researchers found in the mid-seventies, it is important to study the effect of place on the data of the phenomenon under study [2]. it was found that the researcher could make a mistake when neglecting the problem of spatial dependence, represents the inherent characteristic of spatial data, and therefore, semi-parametric spatial models have all the positive characteristics, one of the most important of these qualities is the high flexibility in application [3]. These models are used to study the relationship between the dependent variable and one or more independent variables, when the mathematical relationship between the dependent variable and one or more of the independent variables is known, with at least one variable whose relationship to the dependent variable is unknown, therefore, in this research, one of the semi-parametric spatial models was used,

* *Corresponding author.*
E-mail address: Anwarsabah789@uomustansiriyah.edu.iq

https://doi.org/10.37934/araset.52.1.89105

which is the Durbin spatial parametric regression model, using the modified spatial adjacencies matrix under the Rock adjacency criterion.

There is a substantial amount of published research that focuses on estimating future house prices (as well as pinpointing the aspects of the market that have the greatest bearing on these prices). To achieve this goal, a variety of models have been utilised, ranging from more conventional hedonic regressions [4] to more complex strategies such as generalised additive models (GAMs) with integrated model selection using penalised regression splines [5] geographically weighted regression (GWR) with a non-Euclidean distance metric [6] mixed GWR models (Helbich e The traditional hedonic home pricing models, on the other hand, are either linear or, in the majority of cases, intrinsically linear (they include squared components, natural logarithms and so on) [7]. Furthermore, they do not take into consideration the spatial character of property data. Property data [8] and/or property features [9] inherently contain both geographic autocorrelation and spatial heterogeneity. As a result, you need to take these into consideration while specifying the hedonic home price model, along with any potential nonlinear interactions that may exist between some of the covariates and the response variable [10]. These nonlinear relationships can be specified not only according to conventional practise in standard nonlinear house price models (which are inherently linear), whereby the researcher suggests the parametric form of such relationships, but also by relaxing this parametric form assumption and adopting a nonparametric form. Conventional practise in these models involves the researcher suggesting the parametric form of such relationships [11].This manner, the researcher does not need to indicate the kind of relationship (function) that already exists between the variables; rather, it is the model that specifies this (smooth) function, which is a characteristic that offers a specific benefit when working with enormous databases. When we refer to ''nonlinear'' or ''nonlinearities'' in the following, we are indicating that there is a nonlinear relationship between variables of the last type (smooth and nonparametric), regardless of whether or not the model includes squared terms, logarithms, etc., as is the case in many hedonic house price models. In other words, nonlinearity indicates that there is not a linear relationship between the variables. According to the research that has been done, housing prices are nonstationary2 for a number of reasons. Some of these reasons include the similarities that exist between neighbouring properties and the comparable behaviour of homebuyers [12]. Other reasons include the effect of human communication and market needs [13]. The authors [14] point out that, for the reasons described above, the past few years have witnessed the rise of three effective lines of study that take into consideration spatial impacts in the real estate industry. These successful lines of research are known as spatial econometrics, geographical local modelling and geostatistical [15].

Hence the problem arose when studying the various phenomena and not paying attention to the spatial effect, importance in interpreting and estimating the model, especially predictions, may be inaccurate if neglected. The aim of this research is to estimate the semi-parametric Durbin spatial regression model in the presence of spatial effects.

2. Research methodology

2.1 Durbin Semi-Parametric Spatial Regression Model

Because it incorporates a geographically backward dependent variable as well as a spatially backward explanatory variable, the Durbin spatial semi-parametric regression model is regarded to be one of the most significant spatial standard models for usage in a variety of applications. The mathematical formula for this model may be found as follows [16]:

$$\underline{\mathrm{Y}} = \theta \mathrm{W}\underline{\mathrm{Y}} + \mathrm{m}(\mathrm{X}) + \mathrm{m}(X^{*}) + \underline{\varepsilon} \quad (1)$$

whereas :
Y: represents the vector of the dependent variable with dimension nx1
θ: represents the spatial dependence parameter.
:X^* stands for WX
W: represents the spatial adjacencies matrix.
X: represents the matrix of explanatory variables with dimension nxk.
WX: represents the matrix of spatially retarded explanatory variables.
m(X): represents the smoothing function without the spatial adjacency's matrix.
m(WX): represents the smoothing function with the spatial adjacency's matrix.
WY: represents the vector of the spatially retarded dependent variable.
(ε): represents the vector of unobserved errors, i.e. ε~N(0,σ^2 I_n)
Here, the adjacency criterion of ROCK will be relied upon to build the adjacency matrix.

3.2 Spatial Contiguity Matrix

The basis for spatial regression is knowing whether one region is spatially adjacent to another region, The method of determining the proximity of the place is considered one of the important methods, that were used to calculate the spatial dependence of the data, it was calculated using special matrices, and among these matrices is the regular spatial adjacency matrix, which is characterized by the following:

A positive matrix (it does not contain negative values).
A square matrix of dimensions (n*n).

Only the numbers 0 and 1 make up the components of the matrix that consists of just two numbers. A weight of 0 indicates that the two regions are not next to one another, while a weight of 1 show that the two regions are in continuous proximity to one another. [17] This is the mechanism by which the spatial dependency is determined. It is not necessary for there to be symmetry. The following is how it may be stated mathematically:

$$W = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix}$$

$$w_{ij} = \begin{Bmatrix} 1 & if w_i \ and \ w_j \ are \ adjacent \\ 0 & otherwise \end{Bmatrix} \quad (2)$$

Where (w_ij) is a measure used to compare the degree of convergence or proximity and to know the relationship between region (i) and region (j).

2.3 Adjust Contiguity Matrix

It is called the average because it relies on the usual adjacency matrix in its calculation with the addition of some modifications to it, and its most important characteristics are:

A matrix that is positive (meaning it does not include any values with a negative sign).The values of its components are limited to the range between 0 and 1.As a requirement of this class's standard, the total number of components in each row must add up to one. After, it is utilised to detect variables that are geographically trailing behind. It is not necessary for there to be symmetry. The following is the form in which this matrix is defined [18]:

$$W_{ij}^{std} = \left\{ \frac{w_{ij}}{\sum_i w_{ij}} \qquad 0 < W_{ij}^{std} \leq 1 \right\} \tag{3}$$

If regions i and j share the same boundary.

2.4 Rook Contiguity

There were several methods used to construct the matrix of spatial adjacencies, including the use of the Rock adjacency criterion, which is one of the most widely used criteria due to its realism and ease, as the adjacency is calculated under this criterion as follows:

$$W_{Rook} = \begin{cases} w_{ij} = 1 & \text{If the two areas have a common border from any side (north, south, east, west)} \\ w_{ij} = 0 & \text{Except that} \end{cases}$$

We will take the following Figure 1. To construct the adjacency matrix (see Figure 2) based on the Figure 1, the two adjacent regions (with a line only) will be considered adjoining and given weight (1), and the rest of the cases take weight (0).

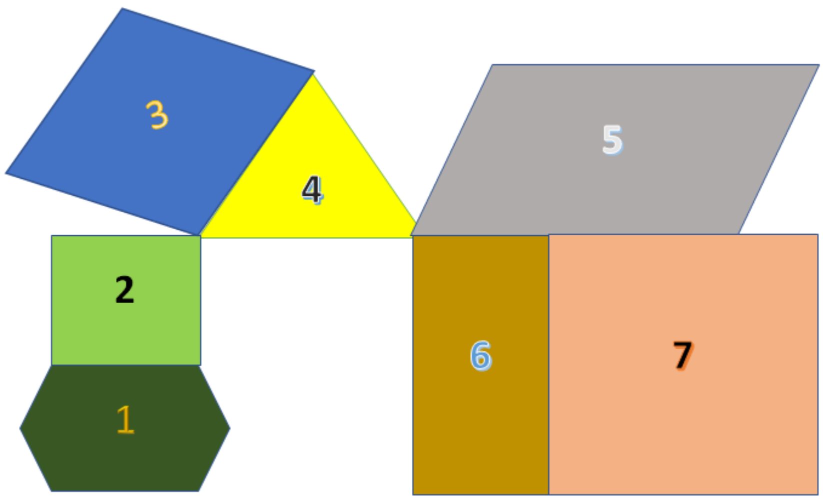

Fig. 1. Example for adjacencies

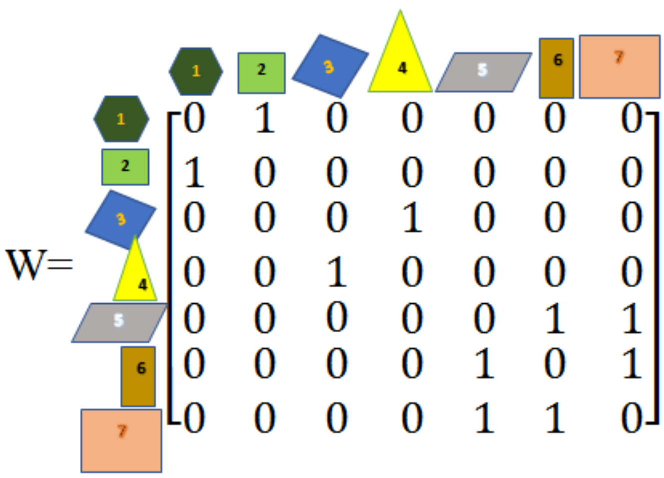

Fig. 2. Adjacency matrix

According to the Rock standard, we note that the area row taking zero values in the region (1,1) because the region does not adjoin itself, while row 1 column 2 takes the weight (1) because the region area adjacent with a line from the top according to Figure (1-1), there is no other adjacent area in other sites, therefore, I took the weight of zero in the remaining values of the first row.

As for the area row so took the weight (1) at the location of the second row, the first column, because the area area adjacent with a single line from the bottom according to Figure (1-1), the rest of the values of the second row are all weighted (0) because there is no other juxtaposition.

for the area class . The location of the third row, fourth column, take the weight (1) as the area area adjacent with a line or a border from the southeast side, as for the rest of the sites in the third row, all of them took the weight (0) because there is no proximity between them and the region.

Also describe the area it took the weight (1) in the location of the fourth row, the third column, because the area area adjacent a border or a line from the north-west side, while the rest of the sites of the region have taken the weight (0) because there is no contiguousness between them and the region.

As for the area row the location of the fifth row, sixth column, took weight (1) because the area area adjacent from the bottom side, with a line or a border, as well as the location of the fifth row, the seventh column, weight (1) was also taken because the area area adjacent with a line from the bottom, as for the rest of the values of the region row they all took the weight (0) because there is no adjacent between them and the rest of the regions.

As for the area row its elements were (1) in the location of the sixth row, the fifth column, because the two regions and adjacent to a line or border from the top, as well as the location of the sixth row, the seventh column, taking weight (1) also because the region area adjacent from the right side with a common line, while the rest of the row weights were (0) because there were no other adjacent areas.

Finally, the row area took weight (1) at the location of the seventh row, fifth column, because the region shares a line from the top with the region as for the rest of the row elements of the region all of them took the weight (0) because there were no contiguous ones.

2.5 Maximum Likelihood Estimation Method for Estimating the Parameter Of (SDM)

The estimation of spatial regression models is most often performed by greatest likelihood estimation (MLE), where the probability of the joint distribution of the greatest value and of all observations is raised with respect to the number of relevant parameters, whereas, the MLE method includes many approximate theoretical properties such as consistency, efficiency, approximate normality, also, this method is considered one of the most important methods of estimation, because it achieves the best estimate of the parameter among several possible estimates. The estimation problems associated with spatial regression models vary with respect to spatial lag and spatial error cases [19]. The spatial regression parameter (θ) in the model is estimated according to Eq. (1) as follows:

$$Y = \theta WY + \mathrm{m}(\mathrm{X}) + \mathrm{m}(X^{*}) + \underline{\varepsilon}$$

whereas :

$\varepsilon \sim N(0, \sigma^2)$

$$\varepsilon = Y(I - \theta W) - m(X) - m(X^*) \quad (4)$$

By maximizing the function of greatest possibility:

$$L(Y/\theta) = (2\pi\sigma^2)^{\frac{-n}{2}} \, | I - \theta W | \, exp\{-1/2 \, \sigma^2 [(Y(I - \theta W) - m(X) - m(X^*))^T (Y(I\text{-}\theta W) - m(X) - m(X^*))] \quad (5)$$

$$\sigma^2 = \frac{e^T e}{n}$$

$$\sigma^2 = \frac{1}{n}[(Y(I - \theta W) - m(X) - m(X^*))^T (Y(I - \theta W) - m(X) - m(X^*))] \quad (6)$$

We take the natural logarithm of both sides to the Eq. (6), so the formula becomes as follows:

A=(I –θW)

$$M = I - X\,(X^T X)^{-1} X^T - X^*\left(X^{*T} X^*\right)^{-1} X^{*T} \quad (7)$$

$$\ln L\left(\frac{Y}{\theta}\right) = C - \frac{n}{2}\ln[e_0^T \, e_0 - 2\theta e_0^T \, e_l + \theta^2 e_l^T e_l] - \frac{n}{2}\ln|\mathrm{I} - \theta \mathrm{W}| \, m \quad (8)$$

The maximization of the above logarithmic potential function is equivalent to the two agency minimizations:

$$\min_{\{\theta\}} \left\{\frac{Y^T A^T M \, A \, Y}{|A|^{2/n}}\right\} \quad (9)$$

This is equivalent to the formula below in which we can find (θ):

$$\min_{\{\theta\}} \left\{\frac{e_0^T \, e_0 - 2\theta e_0^T \, e_l + \theta^2 \, e_l^T \, e_l}{\sum \ln(1 - \theta w_i)}\right\} \quad (10)$$

whereas :

e_0 : represents a vector with residual regression model (Y/X)
e_l: represents the regression vector of the regression model (WY/X)
: w_i represents the Eigen Values of the W weight matrix

$$e_0 = Y - \widehat{\mathrm{m}}(\mathrm{X}) - \widehat{\mathrm{m}}(\mathrm{X}^*) \quad (11)$$

$$e_l = WY - \widehat{\mathrm{m}}(\mathrm{X}) - \widehat{\mathrm{m}}(\mathrm{X}^*) \quad (12)$$

By using the iterative method of Eq. (10), we get the value of (θ).

2.6 Penalized Spline Model

It was proposed in the year (1986) by the researchers (Eilers) and (Marks), and it means that the smoothing is done by using slices (B) with nodes whose number is usually less than the number of data while the distances of the nodes are equal, which is a different approach that can be considered a compromise between the smoothing and the slice Regression, since the main idea is to represent the curve by employing the slice function as well as to control the smoothing by subtracting the penalty limit from the probabilistic model function [20].

The penalty slice method addresses the problem of appropriate determination of the number of nodes and their locations by taking all the different values of the explanatory variable (X) as nodes. What is required for smoothing, and this problem is addressed by setting what is called the penalty constraint, so this method is called (the regression of the penalty segment), so if we assume we have a model that represents the simple linear relationship between the two variables (X,Y), which can be formulated as follows.

$$Y_i=\beta_0+\beta_1 X_i+u_i \tag{13}$$

whereas:

Y: represents the vector of the dependent variable of order (n * 1)

β_0,β_1: represent the two parameters of the model

Xi: represents the vector of the explanatory variable of rank (n*2).

We note from this that Eq. (13) consists of base functions, which consist of only two functions (1, x), which represent the first and second columns, respectively, of the matrix of base functions.

$$X=\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

Assuming we have the model:

$$Y_i=\beta_0+\beta_1 X_i+\beta_2 X_i^2+u_i \tag{14}$$

The base functions of Eq. (14) are (1, x, x^2) and they represent the columns (first, second and third) respectively in an X matrix (the matrix of base functions), which can be written as follows:

$$X=\begin{pmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix}$$

But if we assume that the regression model curve contains more than one slope, then we add the basis function $(x-\tau)_+$. In this case, the model becomes:

$$Y_i=\beta_0+\beta_1 X_i+\beta_2 (x-\tau)_+ + u_i) \tag{15}$$

For this model, the base functions become (1, x, $(x-\tau)_+$), which represents the (first, second and third) columns, respectively.

Therefore, the matrix of X (matrix of base functions) becomes as follows:

$$X=\begin{pmatrix} 1 & x_1 & (x_1-\tau)_+ \\ \vdots & \vdots & \vdots \\ 1 & x_n & (x_n-\tau)_+ \end{pmatrix}$$

We note from the above that the more complex the model becomes, the more it is processed by adding other base functions, in other words, new columns are added to the X matrix of 〖(x-τ)〗 _(+) values [21].

In general, the base functions are:

$$(1, x_i, (x_i-\tau)_+ \ \dots (x_i-\tau_k)_+) \tag{16}$$

The general form of the X matrix is as follows:

$$X=\begin{pmatrix} 1 & x_1 & (x_1-\tau_1)_+ \cdots & (x_1-\tau_k)_+ \\ \vdots & \vdots & \vdots \quad \vdots & \vdots \\ 1 & x_n & (x_n-\tau_1)_+ \cdots & (x_n-\tau_k)_+ \end{pmatrix}$$

whereas : 〖(x_n-τ_k)〗 _+ is the slice function

The functions in the Eq. (16) are the cut-off functions at the nodes (τ_1, τ_2,……τ_k), and therefore this model is called the slide regression model for the following smoothing function m(x) [22].

$$m(x)=\beta_0+\beta_1 X_i+\sum_{l=1}^{k}\beta_{1l}(x_i-\tau_l) \tag{17}$$

whereas :

β_1l is the base chip parameter.

τ_l : represents the number of nodes where

l=1,2,…………k They are called inner nodes, and τ_l falls within the period (a, b)

i.e.:

$a=\tau_1<\tau_2<\tau_3<\tau_4<\cdots\cdots<\tau_k=b$

Therefore, base functions of degree (p) in light of segmental functions of degree (p) are as follows:

$$(1, x_i, x_i^2, \cdots\cdots\cdots x_i^p, (x_i-\tau_1)_+^p, \cdots\cdots\cdots (x_i-\tau_l)_+^p) \tag{18}$$

While the smoothing function of degree (P) is of the form:

$$m(x)=\beta_0+\beta_1 x_1+\beta_2 x_1^2+\beta_3 x_1^3+\ldots+\beta_p x_1^p+\sum_{l=1}^{k}\beta_{pl}(x_i-\tau_l)^p \tag{19}$$

This model depends on the number of nodes and the functions of the base slice. In this way, the points are divided into (τ) of the nodes with equal dimensions within the period (a, b), as the nodes take all the values of the explanatory variable (X) and can be expressed in the following formula:

$$\tau_l=a+\frac{(a+b)l}{(\tau+1)} \qquad , l=1,2,3,\ldots.k$$

The Eq. (17) is restricted in order to improve the fit of the data and can be written in matrix form as follows:

$$\hat{Y} = \mathrm{m(x)}$$
$$\hat{Y} = X\hat{\beta}$$

There are also several penalty constraints that can be followed, but the least squares constraint is easier to apply and is excellent in reducing data inappropriateness to the model [22].

When using the penalty clause on the preamble, the least squares criterion becomes as follows:

$$\text{Mini } |Y - X\beta|^2 \tag{20}$$

Whereas, formula (20) is subject to the penalty condition { { β^T D β ≤C

whereas :

D: It is a diagonal symmetric matrix (the penalty matrix) consisting of only two numbers, which are (0,1) and with dimensions (p+k+1)(p+k+1)

It can be expressed in matrix form as follows:

$$D=\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \ldots & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & . & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & . & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & 1 \end{pmatrix} = \begin{pmatrix} 0_{(p+1)(p+1)} & 0_{(p+1)(k)} \\ 0_{(k)(p+1)} & I_{(k)(k)} \end{pmatrix}$$

Through the Lagrange multiplier, we can obtain the penal least squares criterion as follows:

$$\mathrm{m(x)} = |\mathrm{Y} - \mathrm{X}\beta|^2 + \lambda^2 \beta^T D\, \beta \tag{21}$$

whereas:

λ is the smoothing parameter and is always greater than zero

: is the sum of squares of |Y-Xβ|² errors

²β^(T) D β λ: is the penalty term

And by using the partial derivative of the penal least squares criterion with respect to β, we can obtain the estimator as follows:

$$\frac{\partial}{\partial \beta}\left(\left\|\mathrm{Y} - \mathrm{X}\hat{\beta}\right\|^2\right) + \frac{\partial}{\partial \hat{\beta}}\left(\lambda^2 \hat{\beta}^T D \hat{\beta}\right) = 0$$

Due to the difference, each part is derived separately:

$$\frac{\partial}{\partial \beta}\left(\left\|\mathrm{Y} - \mathrm{X}\hat{\beta}\right\|^2\right) = 2\mathrm{X}^T(Y - X\hat{\beta})$$

As for the second part:

$$\frac{\partial}{\partial\hat{\beta}}\left(\lambda^2\hat{\beta}^T D\hat{\beta}\right) = 2\lambda^2 D\hat{\beta}$$
$$2\lambda^2 D\hat{\beta} - 2X^T\left(Y - X\hat{\beta}\right) = 0$$
$$\lambda^2 D\hat{\beta} = X^T\left(Y - X\hat{\beta}\right)$$
$$\lambda^2 D\hat{\beta} = X^T Y - X^T X\hat{\beta}$$
$$\lambda^2 D\hat{\beta} + X^T X\hat{\beta} = X^T Y$$
$$(\lambda^2 D + X^T X)\hat{\beta} = X^T Y$$

Therefore, the estimated value of β is as follows:

$$\hat{\beta} = (x^T x + \lambda^2 D)^{-1} x^T \underline{y} \quad (22)$$

Substituting the value of β into Eq. (22), we get:

$$\hat{m}\ (x) = x(x^T x + \lambda^2 D)^{-1} x^T\ \underline{y} \quad (23)$$

whereas, Eq. (23) is what is called the estimation of penalty segments. As for the last term of the Durbin spatial regression model [m(wx)], it is estimated by the same Eq. (22), and for short we will denote it by [m(x^*)], so the new formula for calculating the last term becomes as follows:

$$\hat{m}\ (x^*) = x^*(x^{*T`} x^* + \lambda^2 D)^{-1} x^{*T} \underline{y} \quad (24)$$

〖:x〗 ^* represents the spatial hysteresis variable (wx)

As for the penalty limit parameter λ, it is estimated using the legitimate crossing criterion (CV) and according to the following formula:

$$cv_\lambda = \min_\lambda n^{-1} I_{i,}^{n} (Y_i - \hat{m}_{PS}(X))^2 w(X_i) \quad (25)$$

2.7 Moran's Test to Detect Spatial Dependence

It is one of the most important tests that must be conducted for spatial analysis in statistics, before using spatial modelling, data should be selected for spatial dependence, Moran's test statistic was used to detect spatial dependence between regions, represented by the existence of a spatial connection between values. The Moran's value ranges from -1 to 1, if the value of Moran's test coefficient is close to 1 [23].

3. Results and discussion

Electric energy is of great importance in providing amenities for human life and has a great impact in the field of economic development. In general, there is no precise definition of the electrical load, but what has been agreed upon is that the electrical load is the intensity of the current required to operate the load, or it is everything that is connected to the network and consumes electrical energy, where most electrical loads are considered devices to convert electrical energy into other forms of energy such as (motors, heaters, refrigerators, ... etc.). Studies have also shown that there is an effect of air temperature on the consumption of electrical energy and the hours of shedding loads

High temperatures usually have a negative effect on electrical loads, while other studies have shown the effect of the number of individuals in areas on electrical loads. Therefore, this study came to demonstrate the relationship between temperature and population density on electrical loads.

3.1 Description of the Study Variables and Data Collection

The influence that temperature and population density have on electrical loads was the topic of the research, and the variables that were examined were those representing the following categories of information:

Y: The dependent variable, which represents the rate of electrical loads.
X_1 : The first explanatory variable, which represents the average temperature.
X_2: The second explanatory variable, which represents the population density rate.

Where genuine spatial data was utilised for the Iraqi governorates, with the exception of the governorates of the area, for five years in a row, where the sample size was seventy-five, the phrase "where real spatial data was utilised" was used. The Ministry of Electricity and the official website of the Iraqi Air Force, both of which include a significant number of examples displayed in the following table, are cited as the sources.

Table 1
Represents the real data

No.	load rate	Tem. rate	Population density	No.	load rate	Tem. rate	Population density
1	2300027.223	25.1277546	8126755	39	15697.91161	22.8519129	1726409
2	2527119.05	24.4445032	8340711	40	17820.48089	23.0951376	1770765
3	80765.42622	24.801484	8558625	41	296348.2405	25.3257824	1595235
4	78934.22727	25.3573658	8780422	42	371384.7094	24.0692227	1637232
5	92063.99706	25.0541496	9006001	43	13616.08942	24.3227169	1680015
6	394023.2869	23.0285286	3729998	44	14067.26237	24.8461414	1723546
7	599995.4397	21.4814479	3828197	45	17502.81656	24.6145247	1767837
8	20433.14679	21.75	3928215	46	381947.754	26.4581926	1471592
9	20447.94195	22.2775391	4030006	47	443643.8348	25.6526125	1510338
10	24130.2454	22.5659456	4133536	48	14503.44978	25.7327626	1549788
11	1357747.78	28.2324925	2908491	49	14140.08065	26.5686583	1589961
12	1532387.883	27.1772554	2985073	50	16351.04704	26.0006446	1630807
13	50419.13813	27.2101598	3063059	51	374320.8808	27.0004897	1378723
14	53121.94977	27.3809068	3142449	52	438586.282	25.7930043	1415034
15	56949.1881	27.5255238	3223158	53	14835.07834	25.9007991	1452007
16	534870.4251	27.8212388	2095172	54	14252.23104	26.6510985	1489631
17	605335.4213	26.4988971	2150338	55	14757.41142	26.2845851	1527911
18	21480.59521	26.6194064	2206514	56	280392.6418	26.6495286	1291048
19	24201.82044	27.4215318	2263695	57	322515.0563	25.9190721	1325031
20	25102.56769	26.8673984	2321851	58	10652.38736	26.0690639	1359642
21	387948.2858	26.29883	2065042	59	10811.57211	26.8102519	1394885
22	501429.96	25.6206842	2119.403	60	11944.3872	26.3617253	1430714
23	16171.50386	25.8299087	2174783	61	387017.582	26.356442	1218732
24	15799.44911	26.5990047	2231136	62	457019.0818	25.432133	1250806
25	17127.07487	26.2322607	2288456	63	15090.41817	25.6818493	1283484
26	258769.246	25.5421516	1771656	64	15398.87572	25.9959829	1316750
27	341695.1423	24.438499	1818318	65	17062.4767	25.4833632	1350577
28	12612.63433	24.7067352	1865818	66	334298.6734	27.6218205	1112673

29	13046.3349	25.0863493	1914165	67	387462.1515	26.2782358	1141966
30	15429.4123	24.5735122	1963346	68	12878.75914	26.2487443	1171802
31	387948.2858	25.3664986	1637226	69	13471.88275	27.5268484	1202175
32	457857.9099	24.2461515	1680328	70	14343.34375	27.2357687	1233053
33	15405.32481	24.4640411	1724238	71	205562.0208	27.5446485	814371
34	15719.5132	25.0546235	1768920	72	226678.7517	26.4499665	835797
35	16521.55526	24.9365525	1814368	73	8386.304329	26.6280177	857652
36	430017.047	23.1734577	1597876	74	8808.344499	27.196871	879874
37	495926.0058	21.8036828	1639953	75	9680.586331	26.5561727	902480
38	15757.87041	22.2753425	1682809				

3.2 Description of the Study Model (SDM)

For the purpose of analysing the relationship between the average electrical loads (Y) and between the average temperatures (X_1) and population density (X_2), the Durbin spatial semi-parametric model was used to comply with the spatial data structure and my agencies

,i=1,.....75 $Y_i = \theta W_j Y_i + \mathrm{m}(X_i) + \mathrm{m}({X_i}^*) + \underline{\varepsilon}$

whereas:

Y_i: It is a vector with dimensions (75 * 1) and represents the average electrical loads for all observations and at the level of all governorates.

θ: represents the spatial dependence parameter with a value of (0.5134).

W_j: represents the modified spatial adjacencies matrix under the ROCK criterion with dimensions (75 * 75).

m(X_i) is a vector whose dimension is (1*75) and represents the anonymous smoothing function.

:m(〖X_i〗 ^*) is a vector whose dimensions are (1*75) and represents the smoothing function in the presence of the adjacency matrix.

W_j Y_i: A vector whose dimensions are (1*75) and represents the spatially retarded dependent variable.

(ε) is a vector whose dimensions are (1*75) and represents random errors.

3.3 Constructing the Modified Spatial Adjacencies Matrix

The size of the modified spatial weights matrix that was built under the Rock adjacency criterion is (75 * 75), which is a standard matrix denoted by (W_ij^std) and Eq. (3) was used to calculate it based on the neighbourhoods between the fifteen governorates.

3.4 Moran Test Results

This test was carried out in order to determine whether or not there was an issue with spatial dependency, and the findings were as follows:

Table 2
Moran's test to detect spatial effect

Methods	Value	P-Value	Decision
Moran's test	23.1364	0.000	Accept H_1

The results of the (I_moran) test after comparison with the value of (P-Value) and at a significant level of 5%, we note that the test value amounted to 23.1364, and the value of (P-Value) was equal to 0.000, and this indicates the rejection of the null hypothesis and the acceptance of the alternative hypothesis, which is that there is a spatial effect on the rate annual. In other words, there is a spatial effect on the rate annual.

3.5 Estimating the Durbin Spatial Semi-Parametric Regression Model (SDM) using the Penal Segment Regression (PS) Method

Here, the regression model (SDM) is estimated according to the modified spatial adjacency matrix (W_ij^std) under the W_Rook adjacency criterion using the penalty slice regression method (PS) after compensating for the values of the smoothing functions according to the results of the (PS) method and finding the value of the parameter θ within the penalty limit Where the estimated values of the Durbin spatial semi-parametric regression model were obtained using the MATLAB statistical program.

Table 3
The real and estimated values of the (PS) method

No.	Year	Governorate	Ln(Y)	$\hat{Y}$_PS
1	2018	Baghdad	14.64843	11.4445
2	2019	Baghdad	14.74259	11.37458
3	2020	Baghdad	11.2993	11.41791
4	2021	Baghdad	11.27637	11.48186
5	2022	Baghdad	11.43024	11.45368
6	2018	Mosul	12.88417	11.02855
7	2019	Mosul	13.30468	10.84557
8	2020	Mosul	9.924914	10.88332
9	2021	Mosul	9.925638	10.95212
10	2022	Mosul	10.09122	10.99098
11	2018	Basra	14.12134	11.76758
12	2019	Basra	14.24234	11.6693
13	2020	Basra	10.82813	11.67685
14	2021	Basra	10.88035	11.69795
15	2022	Basra	10.94991	11.71635
16	2018	Thi Qar	13.18978	11.56195
17	2019	Thi Qar	13.31354	11.43511
18	2020	Thi Qar	9.974905	11.45162
19	2021	Thi Qar	10.09418	11.53583
20	2022	Thi Qar	10.13073	11.48507
21	2018	Babylon	12.86863	11.21729
22	2019	Babylon	13.12522	10.00281
23	2020	Babylon	9.691006	11.17744
24	2021	Babylon	9.66773	11.26072
25	2022	Babylon	9.748416	11.22754
26	2018	Anbar	12.46369	10.9159
27	2019	Anbar	12.74167	10.80125
28	2020	Anbar	9.442454	10.83494
29	2021	Anbar	9.476263	10.88026

30	2022	Anbar	9.644031	10.82885
31	2018	Diyala	12.86863	10.79961
32	2019	Diyala	13.03431	10.68227
33	2020	Diyala	9.642468	10.71066
34	2021	Diyala	9.662658	10.77916
35	2022	Diyala	9.712421	10.77065
36	2018	Kirkuk	12.97158	10.77474
37	2019	Kirkuk	13.11418	10.61496
38	2020	Kirkuk	9.665095	10.67689
39	2021	Kirkuk	9.661283	10.74997
40	2022	Kirkuk	9.788104	10.7827
41	2018	Saladin	12.59929	10.77467
42	2019	Saladin	12.82499	10.64193
43	2020	Saladin	9.519007	10.67443
44	2021	Saladin	9.551606	10.73603
45	2022	Saladin	9.770117	10.71502
46	2018	Najaf	12.85304	11.47412
47	2019	Najaf	13.00278	11.39517
48	2020	Najaf	9.582142	11.40786
49	2021	Najaf	9.556769	11.49821
50	2022	Najaf	9.702047	11.44422
51	2018	Wasit	12.83287	11.127
52	2019	Wasit	12.99131	11.00809
53	2020	Wasit	9.60475	11.02361
54	2021	Wasit	9.564669	11.10478
55	2022	Wasit	9.599501	11.0717
56	2018	Qadisiah	12.54395	11.36507
57	2019	Qadisiah	12.68391	11.29454
58	2020	Qadisiah	9.273539	11.31437
59	2021	Qadisiah	9.288372	11.39413
60	2022	Qadisiah	9.388017	11.35291
61	2018	Karbala	12.86623	11.24695
62	2019	Karbala	13.03248	11.15512
63	2020	Karbala	9.621815	11.18572
64	2021	Karbala	9.64205	11.22272
65	2022	Karbala	9.744637	11.1733
66	2018	Maysan	12.71979	11.04881
67	2019	Maysan	12.86737	10.91883
68	2020	Maysan	9.463335	10.92009
69	2021	Maysan	9.50836	11.0524
70	2022	Maysan	9.571041	11.02798
71	2018	Muthanna	12.2335	11.14231
72	2019	Muthanna	12.33129	11.03741
73	2020	Muthanna	9.034355	11.05977
74	2021	Muthanna	9.083455	11.12095
75	2022	Muthanna	9.177878	11.06096

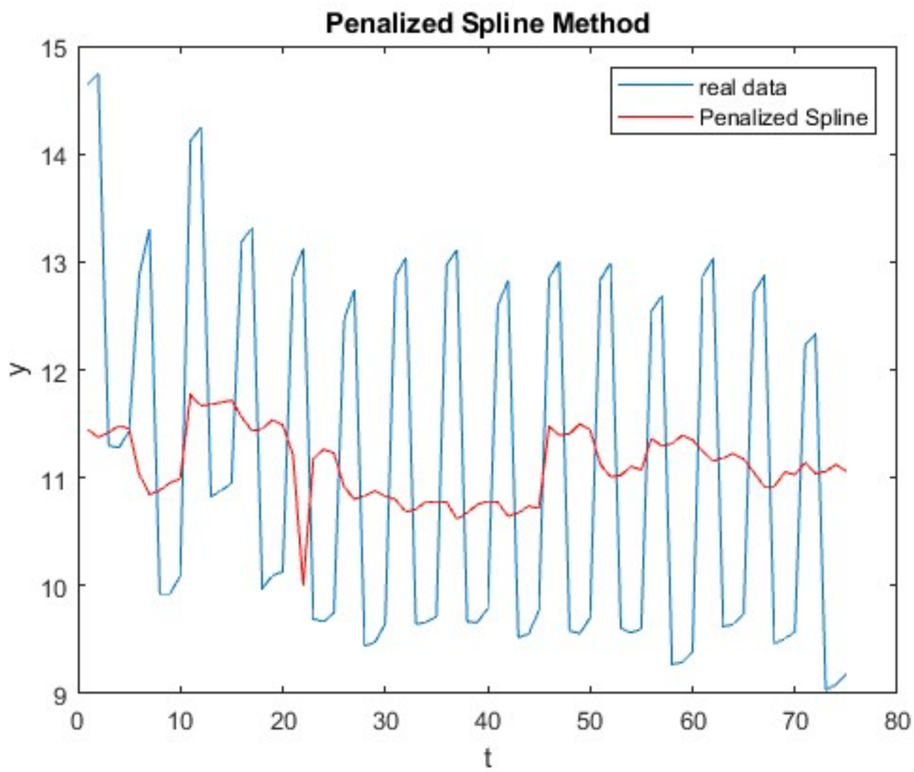

Fig. 3. The real and estimated values of the penalty slide regression method

4. Conclusions

i. By the Moran test to detect the spatial effect, the results showed that there is a spatial effect on the annual rate of electrical loads.
ii. The value of the mean absolute relative error (MABE) was shown under the estimation of the Durban semi-parametric spatial regression model (SDM), and by using the modified adjacency matrix under the Rook adjacency criterion through the estimation by the positional linear estimator method LLE, the value was equal to (2.5953). These results were shown under the estimation of the positional linear estimator method LLE.
iii. The value of the mean absolute relative error (MABE) was shown under the estimation of the Durban semi-parametric spatial regression model (SDM) and by using the modified adjacencies matrix under the Rook adjacency criterion through the estimation by the penal slice regression method PS equal to (1.6946).
iv. The value of the mean absolute relative error (MABE) was demonstrated under the estimate of the Durban semi-parametric spatial regression model (SDM), and by utilising the modified adjacencies matrix under the Rook adjacency criteria through the estimation by the SH reduction approach, the value was equal to (2.5370). Both of these methods were used to display the value of the MABE.
v. By the tables and figures, the estimated values showed that they are close to the real values, which indicates that the Durban spatial semi-parametric regression model represents the electric loads well.

Acknowledgement
This research was not funded by any grant.

References

[1] Fritz, Willy. "Numerical simulation of the peculiar subsonic flow-field about the VFE-2 delta wing with rounded leading edge." *Aerospace Science and Technology* 24, no. 1 (2013): 45-55. https://doi.org/10.1016/j.ast.2012.02.006

[2] Basile, Roberto, and Bernard Gress. "Semi-parametric spatial auto-covariance models of regional growth behaviour in Europe." *Available at SSRN 585221* (2004). https://doi.org/10.2139/ssrn.585221

[3] Lambert, Dayton M. "Dynamic panel estimation of a regional adjustment model with spatial-temporal robust covariance." *Letters in Spatial and Resource Sciences* 13 (2020): 245-265. https://doi.org/10.1007/s12076-020-00257-y

[4] Thompson, Marten Lee. "Gaussian processes in semi-parametric models." PhD diss., university of minnesota, 2023.

[5] Li, Tizheng, Qingyan Yin, and Jialong Peng. "Variable selection of partially linear varying coefficient spatial autoregressive model." *Journal of Statistical Computation and Simulation* 90, no. 15 (2020): 2681-2704. https://doi.org/10.1080/00949655.2020.1788560

[6] Cubadda, Gianluca, Francesco Giancaterini, Alain Hecq, and Joann Jasiak. "Optimization of the Generalized Covariance Estimator in Noncausal Processes." *arXiv preprint arXiv:2306.14653* (2023).

[7] Jalil, Nawal Aswan Abdul, Hussein Kadhim Sharaf, and Sadeq Salman. "A simulation on the effect of ultrasonic vibration on ultrasonic assisted soldering of Cu/SAC305/Cu joint." *Journal of Advanced Research in Applied Mechanics* 36, no. 1 (2017): 1-9.

[8] Alyaseri, Nagham Hameed Abdulkhudhur, Mazen Dawood Salman, Rabab Wahhab Maseer, Emad Kamil Hussein, Kussay Ahmed Subhi, Saleemah Abdullah Alwan, Jasim Gshayyish Zwaid *et al.,* "Exploring the Modeling of Socio-Technical Systems in the Fields of Sport, Engineering and Economics." *Revista iberoamericana de psicología del ejercicio y el deporte* 18, no. 3 (2023): 338-341.

[9] Abdullah, Yussra Malallah, Ghadeer Salim Aziz, and Hussein Kadhim Sharaf. "Simulate the Rheological Behaviour of the Solar Collector by Using Computational Fluid Dynamic Approach." *CFD Letters* 15, no. 9 (2023): 175-182. https://doi.org/10.37934/cfdl.15.9.175182

[10] Nabilah, Nur Amira, Cheng Yee Ng, Nauman Riyaz Maldar, and Fatin Khalida Abd Khadir. "Marine hydrokinetic energy potential of Peninsular Malaysia by using hybrid site selection method." *Progress in Energy and Environment* (2023): 1-10.

[11] Hazian, Nur Qamarina, Nur Ainun Mokhtar, Nurulbahiyah Ahmad Khairudin, and Ragheed Hussam Yousif. "In Silico Molecular Docking Simulation of Chromolaena Odorata Phytoconstituents Against Matrix Metalloproteinase Proteins–9 (MMP-9)." *Journal of Research in Nanoscience and Nanotechnology* 7, no. 1 (2022): 1-6. https://doi.org/10.37934/jrnn.7.1.16

[12] Bachi Al-Fahad, Imad O., Hussein Kadhim Sharaf, Lina Nasseer Bachache, and Nasseer Kassim Bachache. "Identifying The Mechanism Of The Fatigue Behavior Of The Composite Shaft Subjected To Variable Load." *Eastern-European Journal of Enterprise Technologies* 123, no. 7 (2023). https://doi.org/10.15587/1729-4061.2023.283078

[13] Ashham, M., H. K. Sharaf, K. Salman, and S. Salman. "Simulation of heat transfer in a heat exchanger tube with inclined vortex rings inserts." *International Journal of Applied Engineering* 12, no. 20 (2017): 9605-9613.

[14] Li, Tizheng, and Yue Guo. "Penalized profile quasi-maximum likelihood method of partially linear spatial autoregressive model." *Journal of Statistical Computation and Simulation* 90, no. 15 (2020): 2705-2740. https://doi.org/10.1080/00949655.2020.1788561

[15] Hussain, Wafaa Jaafar, Hussam Abdel Razzaq Rasheed, and Ahmed Abd Ali Akkar. "The Best Estimate of The Reliability Function of The Series and parallel systems with Lomax Distribution Based on Type II Censoring Data." *Journal of Administration and Economics* 123 (2020).

[16] Hunt, Richard, Shelton Peiris, and Neville Weber. "Estimation methods for stationary Gegenbauer processes." *Statistical Papers* 63, no. 6 (2022): 1707-1741. https://doi.org/10.1007/s00362-022-01290-3

[17] Lung, Robert, and Nick Polydorides. "Imaging of atmospheric dispersion processes with Differential Absorption Lidar." *arXiv preprint arXiv:2304.06225* (2023).

[18] Düker, Marie-Christine, Robert Lund, and Vladas Pipiras. "High-dimensional latent Gaussian count time series: Concentration results for autocovariances and applications." *arXiv preprint arXiv:2301.00491* (2023).

[19] Nadarajah, Kanchana, Gael M. Martin, and Donald S. Poskitt. "Optimal bias correction of the log-periodogram estimator of the fractional parameter: A jackknife approach." *Journal of Statistical Planning and Inference* 211 (2021): 41-79. https://doi.org/10.1016/j.jspi.2020.04.010

[20] Chen, Zhiyong, and Jianbao Chen. "Bayesian analysis of partially linear additive spatial autoregressive models with free-knot splines." *Symmetry* 13, no. 9 (2021): 1635. https://doi.org/10.3390/sym13091635

[21] Harke, Franz H., Miryam S. Merk, and Philipp Otto. "Estimation of Asymmetric Spatial Autoregressive Dependence on Irregular Lattices." *Symmetry* 14, no. 7 (2022): 1474. https://doi.org/10.3390/sym14071474

[22] Sun, Hu, Zuofeng Shang, and Yang Chen. "Matrix Autoregressive Model with Vector Time Series Covariates for Spatio-Temporal Data." *arXiv preprint arXiv:2305.15671* (2023).
[23] Carron, Andrea, Elena Arcari, Martin Wermelinger, Lukas Hewing, Marco Hutter, and Melanie N. Zeilinger. "Data-driven model predictive control for trajectory tracking with a robotic arm." *IEEE Robotics and Automation Letters* 4, no. 4 (2019): 3758-3765. https://doi.org/10.1109/LRA.2019.2929987