SPATIAL QUANTILE AUTOREGRESSIVE MODEL WITH APPLICATION TO POVERTY RATES IN THE DISTRICTS OF IRAQ

Sawsan Qassim Hadi* *, *Omar Abdulmohsin Ali

Department of Statistics, College of Administration and Economics, University of Baghdad
Iraq, Baghdad
****Corresponding author*** :sawsan.k@coadec.uobaghdad.edu.iq

Keywords: SPATIAL DEPENDENCE, QUANTILE REGRESSION, SPATIAL LAG, WEIGHT MATRIX, SPATIAL QUANTILE AUTOREGRESSIVE.

Abstract

This research aims to provide insight into the Spatial Autoregressive Quantile Regression Model (SARQR), which is more general than the Spatial Autoregressive Model (SAR) and Quantile Regression Model (QR) by integrating aspects of both. Since Bayesian approaches may produce accurate estimates of parameters and overcome the problems that standard estimating techniques face, hence, in this model, they were used to estimate the parameters. Bayesian inference was carried out using Markov Chain Monte Carlo (MCMC). The application was devoted to a dataset of poverty rates across districts in Iraq. Considering the poverty rate as the dependent variable with eight explanatory variables. The analysis confirmed spatial dependence among regions, as indicated by the estimated values of the spatial correlation parameter (ρ) across different scenarios. The highest value of the parameter ρ was 0.43 at σ=2 and the quantile 0.1. It made clear that poverty rates are heavily influenced by spatial dependence and that failing to consider this could result in the loss of important information regarding the phenomenon and eventually impair the accuracy of statistical index estimation.This affects the accuracy of poverty rate predictions. This enhancement offers suggestions for methods of reducing poverty.

1 Introduction

A statistical technique for examining and analyzing the relationship between the response variable (Y) and one or more explanatory variables (X) is regression analysis. The linear regression model's parameters can be estimated using various statistical methods, the most important and widely used being the ordinary least squares (OLS) method. OLS estimates are characterized by being the best linearly unbiased estimates. However, spatial data analysis can be biased due to spatially correlated observations. Spatial econometrics focuses on spatial effects and variables distributed based on location rather than time. In contrast, traditional econometrics focuses on the interactions between explanatory variables and their effect on the response variable. If there is spatial heterogeneity in regression coefficients, the regression becomes less capable of explaining the actual data phenomenon. In some cases, testing for spatial effects by including outliers in the data can lead to a method's failure to address these spatial effects. Therefore, what is often done is the removal of outliers. In reality, removing outliers might be a wrong step because, sometimes, outliers can provide information that other data cannot. By not including outliers in the model, the analysis results can be biased or may not reflect the actual phenomenon.

There are several gaps in the literature related to the SARQR model and Bayesian estimation methods. First, because SARQR models are a mixture of two models—spatial autoregression and quantile regression—there is a dearth of theoretical knowledge regarding them. It is yet unknown what processes enable these models to analyze spatial data more effectively than conventional models. Second, there isn't a thorough examination of various Bayesian estimation methods in the context of spatial autoregressive models in the literature currently in publication; therefore, it's crucial to research how the selection of various estimate strategies influences the analysis's outcomes. Furthermore, researchers' and practitioners' capacity to properly employ SARQR models is hampered by the lack of analytical tools (Page 3 of 9—AI Writing Submission) that are accessible to assist their usage. Finally, further study in this area is required to guide the outcomes of these models for policymakers across multiple disciplines, as the literature currently lacks a thorough grasp of assessing discrepancies between distinct quantiles and the influence of geographical variables on them. Researchers in 1971 presented a study on estimating spatial econometrics, focusing on spatial dependence in estimating cross-sectional models. They also proposed a balanced spatial variance matrix for disturbances and linked the results to more traditional estimation procedures. [11]

Bivand and Klaasen published a small volume titled "Spatial Econometrics," which is considered foundational research on the topic of spatial econometrics and its distinctive methodology. They were the first to discover the geographic dependence among regression residuals [27]. The quantile regression, was first proposed in 1978 to model the effect of explanatory variables on the entire distribution of the response variable. Quantile regression methods are primarily applied to continuous response variables; however, spatial dependence

https://doi.org/10.1049/icp.2025.0085

has not been taken into account. Ignoring spatial dependence can lead to inconsistent and biased estimates. One study applied quantile regression to a spatial lag model for land price modeling, comparing confidence intervals with those from a parametric spatial lag model, highlighting the importance of spatial self-dependence in land price models [14]. They found that spatial correlation improved efficiency compared to standard estimates in a study by Yang, which used rainfall data from Illinois and simulation data to predict conditions using a linear quantile regression model [34]. Additionally, a study assessed the effectiveness of macro-control policies on urban housing prices. The results showed that restrictions on home purchases effectively limited speculative demand, but the effective reduction of prices remains a challenge in high-priced areas [37].

There are some relevant literature reviews related to the research topic that have been studied by certain researchers, including the study by Ngwira, A. (2019). To estimate the model, they employed the Integrated Nested Laplace Approximations method. The results indicated that the different ages of mothers significantly affect birth weight rates, with clear spatial effects [26]. Wardhani and Yanti (2021) conducted a study on spatial autoregressive quantile regression to analyze malnutrition data for toddlers in Bandung City in 2018. They used the spatial autoregressive quantile regression method and instrumental variable quantile regression to estimate the factors influencing malnutrition in children under five. The results showed different parameter values and variables for each quantile [33]. This study addressed Dai et al. (2022) partially linear spatial autoregressive models using the quantile regression technique and potentially variable coefficients. The changing coefficients were approximated using the B-spline. For the coefficients, tests of rank scores were developed for the invariance and constancy of changing hypotheses [10]. The study by Lee and Huang (2024) uses quantile regression to assess social vulnerability to local economic and public health outcomes during the COVID-19 pandemic. It reveals that social vulnerability varies within large areas with varying pandemic consequences [18].

The issue of spatial dependence in the analysis of economic and social data is a vital concern that requires significant attention from researchers, as this problem directly affects the accuracy of the estimates and models used. In many studies, the impact of spatial dependence is often overlooked, leading to unreliable and inaccurate results [20]. The SARQR model is an effective tool for addressing this issue, as it allows models to account for the effects of spatial variables on the response variable. However, many gaps remain in the literature regarding the estimation of this model using Bayesian methods [35]. While traditional regression methods have shown some success, they may lack the flexibility needed to handle complex spatial data that involve non-linear relationships or heterogeneous distributions [21]. Therefore, researching the use of Bayesian estimation methodologies for spatial autoregressive quantile models is an urgent necessity, as such studies can contribute to enhancing the understanding of spatial dynamics and provide more accurate. where observations are taken from a set of locations or positions. The research becomes more challenging due to the complexity of the potential spatial dependence between different locations. Therefore, it is necessary to consider the spatial varying coefficient model instead of the ordinary spatial regression model.

The purpose of this study is to explore the impact of the spatial autoregressive quantile model and the Bayesian estimation method on spatial data analysis, focusing on addressing the knowledge gaps in the current literature. The study aims to analyze the model's effectiveness by assessing how to integrate spatial autoregression with quantile regression to achieve more accurate and reliable estimates. Additionally, the study seeks to explore the effect of using Bayesian estimation techniques to improve the accuracy and effectiveness of the models and to understand how different estimation choices affect the results. Furthermore, the study will conduct practical applications on real data to analyze social and economic challenges, contributing to enhancing the understanding of the relationship between spatial variables and their impact on outcomes. By providing valuable insights for policymakers on how to use these models to analyze complex issues, the study hopes to support more informed decision-making. The research hypothesis posits that the Bayesian estimation method can accurately and effectively estimate the parameters of the SARQR model. Accordingly, the study aims to estimate the parameters of the spatial quantile regression model using spatial analysis methodology to obtain accurate and realistic results and predictions.

The research is organized into sections: Section 2 spatial analysis, weight matrix, QR, SAR, and SARQR models, parameter estimation method, and comparison criteria. Section 3 presents SARQR model results and parameter estimates, while Section 4 summarizes observed results and offers further guidance.

2. Methodology

2.1 Spatial analysis

Spatial analysis, often known as spatial data analysis, is a method used to measure spatial relationships between phenomena based on location measurements. It helps understand the dispersion of phenomena on Earth's surface and predict their future behavior. It assumes each phenomenon has a specific range and spread, and aims to uncover mutual spatial relationships between components and multiple types of phenomena to construct a clear spatial model [30]. The type of point pattern distribution random, uniform, or clustered determines the methodology of spatial analysis. Serial correlation is another name for spatial autocorrelation [2]. Statistical theory and geographic methods significantly aid in understanding the potential for geographical and temporal developments, events influenced by location over time [1].

2.2 The Spatial Weight Matrix

The spatial autocorrelation matrix is a square matrix used to represent spatial relationships, denoted by the symbol (W). It has dimensions of (nxn) and is built based on adjacencies. [12]. The choice of spatial weight matrices is crucial for determining spatial effects and determining spatial correlation. The results of any spatial analysis depend on the matrix used, so selecting the appropriate weight matrix is essential. Determining spatial correlation is a fundamental requirement before initiating standard spatial econometric analysis or exploratory analysis of spatial data [29]. The matrix's general formula can be expressed as follows:[28]

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

there are many types of spatial weight matrices, including :

2.2.1 Binary Contiguity Weights Matrix:

Its dimensions are (n×n). If i and j are adjacent, then wij=1; if i and j are not adjacent, then wij=0, where i and j are locations, as shown in the formula below [5].

$$W_{ij} = \begin{Bmatrix} 1 & \text{i neighbor} \\ 0 & \text{otherwise} \end{Bmatrix} \quad (2)$$

Here are different types of contiguities, as listed below:[20]

a) Rook Contiguity: is a common contingency where two adjacent areas share a common border.
b) Bishop Contiguity: Contiguity occurs when two areas share a common point.
c) Queen Contiguity: This matrix combines elements from the Rook and Bishop contiguity matrix.
d) Linear Contiguity: The has elements indicating adjacent areas and zero elements indicating only one or no adjacency in a row.

2.2.2 Row-Standardized Weights Matrix:

This matrix is called the adjusted matrix, where the sum of each row equals one. It is calculated based on the Binary Contiguity Weights Matrix. The formula is shown below[7][32].

$$W_{ij}^{std} = \begin{Bmatrix} \frac{w_{ij}}{\sum_i w_{ij}} & \text{i neighbor j} & 0 < w_{ij}^{std} \le 1 \\ & 0 \quad \text{otherwise} & \end{Bmatrix} \quad (3)$$

2.3 Quantile Regression

Referred to as a robust and alternative regression method to ordinary least squares (OLS), most studies indicate that robust regression is insensitive to outliers and heteroscedasticity, making it capable of accommodating residuals that do not follow a normal distribution, which are common in many applications[6][13]. Instead of being restricted to estimating the conditional expectation (E(Y|X)) as in ordinary mean regression, quantile regression allows for a more thorough analysis of the relationship between the response variable and the explanatory variables by estimating various conditional quantiles $(Q_\tau(Y/X), 0 < \tau < 1)$ of the response variable distribution[3][4][22]. It has been applied in various fields such as econometrics, finance, medical studies, agriculture, and others. The mathematical formula for the quantile regression model is shown below:

$$Q_\tau(Y/X) = X\beta_\tau \quad (4)$$

Where:
Y: A vector of observations of the dependent variable.
X: A matrix of observations of the explanatory variables.
$\beta\tau$: Represents the parameter vector at quantile (τ).
τ: The quantile level, where $0 < \tau < 1$.
To obtain $\beta\tau$ according to the following formula: [36]

$$\min_{\beta_\tau} \sum_{i=1}^{n} \varphi_\tau(y_i - x_i\beta_\tau) \quad (5)$$

where $\varphi_\tau(.)$ is the check function. To select the loss function, it can be expressed as follows:[23][24]

$$\varphi_\tau(\varepsilon) = \begin{Bmatrix} \tau\varepsilon & \text{if} & \varepsilon \ge 0 \\ -(1-\tau)\varepsilon & \text{if} & \varepsilon < 0 \end{Bmatrix}, \varepsilon = y_i - x_i\beta_\tau \quad (6)$$

Alternatively, it can be phrased as follows:

$$\varphi_\tau(\varepsilon) = \frac{|\varepsilon| + (2\tau - 1)\varepsilon}{2} = \varepsilon(\tau - I(\varepsilon < 0)) \quad (7)$$

2.4 Spatial Autoregressive model

is a model that uses cross-sectional data to combine simple regression with spatial lag on the dependant variable. Spatial lag is symbolized by SAR. The specification of the spatial lag is characterized by including a new variable located on the equation's right side. When analyzing spatially related data, self-correlation and spatial dependence must be taken into account to avoid biased estimation. Generally, spatial models are important for researchers due to their ability to calculate location-specific effects. Also known as the Mixed Spatial Autoregressive Model or Spatial Lag Model, this model represents a special case of the General Spatial Autoregressive Model (SAC) proposed by Anselin. Mathematically, it can be expressed as follows:[19][21]

$$Y = \rho WY + X\beta + \varepsilon \;, \varepsilon \sim N(0,\sigma 2\ In) \quad (8)$$

Where:
ρ: is the parameter representing spatial effects or it's also called the spatial autoregressive parameter.
W: A ((n×n)) contiguity matrix.
ε: vector of random errors.
I_n: Represents an (n×n) identity matrix.
The value of the spatial autoregressive parameter falls within the range $-1 < \rho < +1$.

2.5 Spatial Quantile Autoregressive Model

This model, represented by the notation SARQR, combines two models: quantile regression and spatial autoregressive regression. It is distinguished by its capacity to resolve the issue of data heterogeneity and overcome outliers [31]. Regression parameters and a spatial lag parameter in the SARQR model are dependent on certain quantile values. This model's mathematical formula is as follows:

$$Y = \rho_\tau WY + X\beta_\tau + \varepsilon \quad (9)$$

Where:
ρ_τ: Represents the spatial autoregressive parameter at quantile (τ).

2.6 The Bayesian method for estimating parameters of the SARQR model.

Involves use of posterior probability function, which is crucial for providing information about unknown features in the model. Therefore, determining the posterior probability function is a critical step in initiating the estimation process. In the case of SARQR, which does not assume a specific error distribution in its analysis, specifying the posterior probability function becomes challenging. This limitation historically restrained the use of Bayesian methods in SARQR estimation until 2001, when Yu and Moyeed proposed using an asymmetric Laplace distribution to represent error bounds in linear QR models, regardless of the actual data distribution, reflecting the typical behavior in quantitative regression analysis using Bayesian methods[35].

Despite the significance of this assumption in Bayesian estimation, difficulties arose in deriving subsequent distributions. However, Yu and Moyeed suggested the potential use of Markov Chain Monte Carlo (MCMC) methods to overcome these challenges, which are crucial even in complex scenarios for obtaining posterior distributions. The probability density function of the asymmetric Laplace distribution incorporating the measurement parameter can be expressed as follows:[35]

$$f_\tau(\varepsilon_i) = \tau(1-\tau)\,\sigma \exp\{-\sigma\,\varphi_\tau(\varepsilon_i)\} \quad (10)$$

And the posterior probability function of the (10) function becomes as follows:

$$f_\tau(\varepsilon_i) = \tau^n(1-\tau)^n\sigma^n \exp\{-\textstyle\sum_{i=1}^{n}\sigma\,\varphi_\tau(\varepsilon_i)\} \quad (11)$$

Where:

ε_i: represents the random error term (residuals), which has a bounded distribution.

$$F(\varepsilon_i)\int_{-\infty}^{0} f_\tau(\varepsilon_i)d\varepsilon_i = \tau \ \text{ or } \ F_\varepsilon^{-1}(\tau) = 0 \quad (12)$$

The distribution of the dependent variable will be an asymmetric Laplace distribution, according to the probability density function in equation below:

$$f_\tau(y_i) = \tau^n(1-\tau)^n\sigma^n \exp\{-\textstyle\sum_{i=1}^{n}\sigma\,\varphi_\tau(y_i-\rho_\tau wy - X\beta_\tau)\} \quad (13)$$

The minimization of the equation $L(y,x)=\sum_{i=1}^{n}\varphi_\tau(y_i-\rho_\tau wy - X\beta_\tau)$ used in the conventional method for parameter estimation of the model is equivalent to maximizing the likelihood function (13).

Using the (13) function directly is very challenging for obtaining the Bayesian model of posterior distributions, and the algorithm tends to be inefficient. Therefore, it is noteworthy that many researchers have used the mixed representation of the asymmetric Laplace distribution[16]. when paired with Bayesian methods. This approach is of great importance in facilitating computational operations in posterior distributions for parameter estimation. One of the most famous transformations (ALD) was developed by researchers Kozumi and Kobayashi in 2011, who successfully demonstrated that the asymmetric Laplace distribution can be represented as a mixture of normal and exponential distributions, as detailed below[15][17].

Assuming that (z) is a random variable distributed according to the standard normal distribution, meaning

$$f(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right) \qquad -\infty < z < \infty \quad (14)$$

And assuming that (e) is a random variable following an exponential distribution with parameter σ, meaning

$$f(e_i\backslash\sigma) = \frac{1}{\sigma}\exp\left(-\frac{e_i}{\sigma}\right) \qquad e_i \ge 0 \quad (15)$$

e_i and z_i are independent.

Therefore, the random variable (ε) follows asymmetric Laplace distribution has the probability density function as follows:

$$\varepsilon = k_1e_i + \sqrt{k_2\sigma\, e_i}\,z_i \quad (16)$$

From the above result, the regression model can be as:

$$\left.\begin{array}{l} y_i = \rho_\tau wy + x'\beta_\tau + k_1e_i + \sqrt{k_2\sigma\, e_i}\,z_i \qquad i = 1,2,\dots,n \\ e_i \sim \exp\left(\frac{1}{\sigma}\right) \\ z_i \sim N(0,1) \\ k1 = \frac{1-2\tau}{\tau(1-\tau)} \qquad\qquad k2 = \frac{2}{\tau(1-\tau)} \end{array}\right\} \quad (17)$$

Based on the mixed representation of the error distribution in equation (17), the conditional distribution of the response variable (y) given the error variable (e) follows a normal distribution with mean $(\rho\sum w_{ij}y_j + x_i'\beta + k_1e_i)$ and variance $(k_2\sigma e_i)$, as shown in the equation below:

$$p(y\backslash\beta,\rho,\sigma,e,x)\,\alpha\,\frac{1}{\sqrt{2\pi k_2\sigma e_i}}\,|I-\rho w|\, e^{\left[\frac{-(y_i-\rho\sum w_{ij}y_j - x_i'\beta - k_1e_i)^2}{2k_2\sigma e_i}\right]} \quad (18)$$

a) Prior distributions: For conducting Bayesian analysis, it is essential to specify the prior distributions, which play a crucial role in shaping the future estimation method of model parameters. It is assumed that the parameter β is independently distributed according to a normal distribution $(N(0,\delta_k^2))$, with the function as follows:

$$\pi(\beta_k\backslash\,\delta_k^2) \propto (\delta_k^2)^{-\frac{1}{2}}\ e^{\left[-\frac{\beta_k^2}{2\delta_k^2}\right]} \qquad k = 1,\dots,p \quad (19)$$

The inverse gamma distribution for σ is specified using a method akin to that of researchers Kozumi and Kobayashi (2011). which is shown below with the shape parameter a_σ and scale parameter b_σ:

$$\pi(\sigma) \propto \sigma^{-(a_\sigma+1)}\ e^{\left[-\frac{b_\sigma}{\sigma}\right]} \quad (20)$$

In this study, the prior distribution for the parameter ρ was assumed to be uniform, ranging between (ρ_{min},ρ_{max}), as:

$$\pi(\rho) \sim U(\rho_{min},\rho_{max}) \quad (21)$$

As for the prior distribution of the parameter (δ_k^2), which includes (δ), it is specified as an inverse gamma distribution, depicted as follows:

$$\pi(\delta_k^2) \propto (\delta_k^2)^{-(a_\delta+1)}\ e^{\left[-\frac{b_\delta}{\delta_k^2}\right]} \quad (22)$$

b) Full conditional distributions: In Gibbs sampling, where these distributions are obtained from the posterior distribution, obtaining entire conditional distributions is a fundamental prerequisite. The full conditional distributions for the model, which have been derived, are illustrated below:

1) Sample βk from conditional distribution p(βk\σ,ρ,e,δk,y,x),

$$p(\beta_k\backslash\sigma,\rho,e,\delta_k,y,x) \sim N(\beta_k^*,\gamma_k^2) \quad (23)$$

where: $\beta_k^* = \gamma_k^2\left[\frac{1}{k_2\sigma}\sum_{i=1}^{n}\frac{y_i-\mu_i^*-k_1e_i}{e_i}\,x_{ik}\right]$

$$\gamma_k^2 = \left[\frac{1}{\delta_k^2}+\frac{1}{k_2\sigma}\sum_{i=1}^{n}\frac{x_{ik}^2}{e_i}\right]^{-1} \quad \mu_i^* = \rho\sum_{j=1}^{n}w_{ij}y_j + \sum_{j=1,j\ne k}^{p}x_{ij}\beta_j$$

2) Sample σ from conditional distribution $p(\sigma\backslash\beta,\rho,e,\delta,y,x)$,

$$p(\sigma\backslash\beta,\rho,e,\delta,y,x) \propto p(y\backslash\beta,\rho,\sigma,e,\delta,x) * \pi(e_i\backslash\sigma) * \pi(\sigma)$$

$$= \sigma^{-\left(a_\sigma+\frac{3n}{2}+1\right)}\ e^{\left[-\frac{b_\sigma+\sum_{i=1}^{n}\frac{(y_i-\rho\sum_{j=1}^{n}w_{ij}y_j-x_i'\beta-k_1e_i)^2}{2k_2e_i}+e_i}{\sigma}\right]}$$

$$p(\sigma\backslash\beta,\rho,e,\delta,y,x) \sim \text{InvGamma}(a_\sigma^*, b_\sigma^*) \quad (24)$$

where: $a_\sigma^* = a_\sigma + \frac{3n}{2}$

$$b_\sigma^* = b_\sigma + \sum_{i=1}^{n} \frac{\left(y_i - \rho \sum_{j=1}^{n} w_{ij} y_j - x_i' \beta - k_1 e_i\right)^2}{2 k_2 e_i} + e_i$$

3) Sample ρ from conditional distribution $p(\rho \backslash \beta, \sigma, e, \delta, y, x)$,
$p(\rho \backslash \beta, \sigma, e, \delta, y, x) \propto p(y \backslash \beta, \rho, \sigma, e, \delta, x) * \pi(\rho)$

$$= |I - \rho w| e^{\left[\frac{-1}{2k_2\sigma} \sum_{i=1}^{n} \frac{\left(y_i - \rho \sum w_{ij} y_j - x_i' \beta - k_1 e_i\right)^2}{e_i}\right]} I\{\rho min<\rho<\rho max \quad (25)$$

4) Sample δ_k^2 from conditional distribution $p(\delta_k^2 \backslash \beta)$,
$p(\delta_k^2 \backslash \beta_k) \propto p(\beta_k) * p(\delta_k^2)$

$$= (\delta_k^2)^{-\left(a_\delta + \frac{1}{2} + 1\right)} e^{-\left[\frac{b_\delta + \frac{\beta_k^2}{2}}{\delta_k^2}\right]}$$

$$p(\delta_k^2 \backslash \beta_k) \sim InvGamma(a_\delta^*, b_\delta^*) \quad (26)$$

where: $a_\delta^* = a_\delta + \frac{1}{2}$; $b_\delta^* = b_\delta + \frac{\beta_k^2}{2}$

5) Sample e_i from conditional distribution $p(e_i \backslash \beta, \sigma, \rho, y, x)$,
$p(e_i \backslash \beta, \sigma, \rho, y, x) \propto p(y \backslash \beta, \rho, \sigma, e, \delta, x) * \pi(e_i)$

$$= e_i^{\frac{1}{2}-1} e^{\left[-\frac{1}{2}\left(\frac{e_i^{-1}\left(y_i - \rho \sum_{j=1}^{n} w_{ij} y_j - x_i' \beta\right)^2}{k_2 \sigma} + \left(\frac{k_1^2}{k_2 \sigma} + \frac{2}{\sigma}\right) e_i\right)\right]}$$

$$p(e_i \backslash \beta, \sigma, \rho, y, x) \sim GIG\left(\frac{1}{2}, m^2, n^2\right) \quad (27)$$

where: $m^2 = \frac{\left(y_i - \rho \sum_{j=1}^{n} w_{ij} y_j - x_i' \beta\right)^2}{k_2 \sigma}$; $n^2 = \frac{k_1^2}{k_2 \sigma} + \frac{2}{\sigma}$

2.7 Comparison criteria

Choosing a specific method among several options is crucial in data analysis statistical criteria are used, such as:

2.7.1 Mean Absolute Percentage Error:

It is one of the most common criteria for prediction accuracy. It is calculated as the average of the absolute error divided by the actual value, summed over all observations (n). A lower value indicates better performance. The specific formula for calculating MAPE is as follows:[25]

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (28)$$

2.7.2 Root Mean Squares Error:

It is the product of the square root of the total squared errors and (n−k−1) divided. The best model is one with the lowest root mean squared error (RMSE) value. This criterion's particular calculation formula is [25].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-k-1}} \quad (29)$$

2.7.3 Coefficient of determination:

The coefficient of determination, known as R-squared or R2, is a statistical measure that represents the percentage of the dependent variable's variance that the independent variable (X) in a regression model explains. Its value lies between 0 and 1, where a value close to 1 indicates that the regression model explains a large portion of the variance in Y. The model with the highest value is considered the best fit, which is calculated using the following formula:[8] [9]

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (30)$$

3 Results

3.1 Data

The data was obtained from the Central Statistical Organization of the Ministry of Planning. The poverty rate is the variable that is dependent (Y) in this research, and it was analyzed alongside 8 explanatory variables: illiteracy rate (X1), unemployment rate (X2), average family size (X3), net non-enrollment ratio in primary (X4), incidence of death, illness, or injury to family members (X5), job loss or unemployment (X6), loss of food ration share (X7), and economic activity (X8). The research surveyed 88 districts, excluding the Kurdistan Region.

3.2 Tables

In this section, to uncover the objective and significance of the research, real data representing poverty rates in the districts of Iraq were analyzed. The estimation of SARQR model parameters was conducted using Bayesian methods, assuming three different values of σ (2, 5, 10), along with different values of five quantiles (0.1, 0.3, 0.5, 0.7, 0.9). The practical application results were obtained using the R programming language, as illustrated in Tables 1 to 5.

Table 1 Estimated parameter values for the SARQR model at quantile 0.1

Parameters	σ=2	σ=5	σ=10
β1	0.23493	0.2576	0.26551
β2	0.88482	0.88412	0.88447
β3	-0.2945	-0.237	-0.2113
β4	0.91751	0.91132	0.90714
β5	-14.94	-9.4484	-8.9282
β6	-4.5924	-2.0608	0.63829
β7	24.0837	22.306	21.8577
β8	-0.5209	-0.5301	-0.5405
ρ	0.43119	0.41084	0.39738

Table 2 Estimated parameter values for the SARQR model at quantile 0.3

Parameters	σ=2	σ=5	σ=10
β1	0.26681	0.27251	0.27396
β2	0.87858	0.8834	0.87435
β3	-0.1974	-0.1971	-0.1827
β4	0.90949	0.90509	0.90639
β5	-5.2695	-4.6857	-3.836
β6	2.12957	2.26518	3.26883
β7	20.2727	20.5085	20.1101
β8	-0.5433	-0.5424	-0.546
ρ	0.39643	0.39459	0.38922

Table 3 Estimated parameter values for the SARQR model at quantile 0.5

Parameters	σ=2	σ=5	σ=10

β1	0.27574	0.27083	0.27457
β2	0.87942	0.88202	0.87976
β3	-0.1711	-0.177	-0.1677
β4	0.9041	0.91005	0.90994
β5	-2.6093	-2.5928	-2.6743
β6	3.8598	3.75516	3.92831
β7	19.5526	19.7708	19.9578
β8	-0.5494	-0.5491	-0.5501
ρ	0.38876	0.38774	0.38213

Table 4 Estimated parameter values for the SARQR model at quantile 0.7

Parameters	σ=2	σ=5	σ=10
β1	0.29255	0.28264	0.29164
β2	0.87839	0.87895	0.87818
β3	-0.1403	-0.1492	-0.1466
β4	0.89886	0.90526	0.8985
β5	-0.7141	-1.4546	-1.6364
β6	5.52015	4.31218	4.61534
β7	19.1104	19.5361	20.3431
β8	-0.5551	-0.5503	-0.5562
ρ	0.37477	0.37919	0.37629

Table 5 Estimated parameter values for the SARQR model at quantile 0.9

Parameters	σ=2	σ=5	σ=10
β1	0.31977	0.30063	0.3224
β2	0.87574	0.87884	0.88101
β3	-0.0359	-0.0862	-0.0912
β4	0.89845	0.90945	0.88999
β5	7.89722	2.78845	0.71152
β6	12.4241	9.61718	6.81905
β7	15.8239	16.7347	18.4641
β8	-0.5782	-0.5685	-0.5599
ρ	0.33309	0.34864	0.35257

The estimated value of the spatial correlation parameter (ρ), when seen in all scenarios, varies and indicates spatial dependence among the districts. The highest value was recorded at quantile 0.1 and standard deviation 2, while the smallest value was at quantile 0.9 and standard deviation 2. The comparisons were made using metrics like coefficient of determination, root mean squared error, and mean absolute percentage error, as per table 6.

Table 6 Comparison metrics results at quantile 0.1

τ		σ=2	σ=5	σ=10
0.1	R^2	0.5016	0.5276	0.5289
	RMSE	13.4471	13.0917	13.0728
	MAPE	0.56276	0.5933	0.62115
0.3	R^2	0.5238	0.5199	0.5149
	RMSE	13.1438	13.1971	13.2665
	MAPE	0.64995	0.66282	0.67874
0.5	R^2	0.5097	0.5080	0.5034
	RMSE	13.3371	13.3603	13.4226
	MAPE	0.69192	0.69487	0.70532
0.7	R^2	0.4813	0.4928	0.4926
	RMSE	13.7178	13.5652	13.5679
	MAPE	0.7484	0.72797	0.72701
0.9	R^2	0.2918	0.3934	0.4294
	RMSE	16.0297	14.8347	14.3879
	MAPE	0.97381	0.86783	0.82168

Based on the tables above, it can be observed that the highest values for the metrics (R^2, RMSE, and MAPE) were recorded at quantile 0.1 and a standard deviation of 10. Conversely, the lowest values for all three metrics were noted at quantile 0.9 and a standard deviation of 10. The best metrics were obtained when (σ) equaled 2 at quantiles 0.3 and 0.5.

4 Conclusion

The research results indicate that using the Spatial Autoregressive Quantile Regression Model (SARQR) with the Bayesian estimation method produces accurate estimates of poverty parameters, taking into account the spatial effects between different regions. By comparing the model using various criteria Mean Absolute Percentage Error and Root Mean Squares Error and the value of R-squared, it was found that the Bayesian method provides precise estimates that effectively reflect the spatial variation in poverty data. Spatial correlation was observed when analyzing the data, with the highest value across different scenarios being $\rho = 0.43$ at $\sigma = 2$ and quantile 0.1, indicating the effectiveness of accounting for geographical dependence between regions in the model. Based on these conclusions, it can be stated that the Spatial Autoregressive Quantile Regression Model (SARQR) with Bayesian estimation and its evaluation across different performance criteria and quantiles provides an effective analytical tool for studying poverty. It is recommended that policymakers improve poverty alleviation strategies by targeting the most needy regions based on accurate spatial and quantile analyses.

5 References

[1] Al Ramahi, F., Al Bahadly, Z.: 'The spatial analysis for Bassia eriophora (Schrad.) Asch. plant distributed in all Iraq by using RS & GIS Techniques', Baghdad Science Journal, 2020, 17(1), pp.126-135.

[2] Al-ramahi, F., Shnain, A., Ali, A.:'The Modern Techniques in Spatial Analysis to Isolate, Quarantine the Affected Areas and Prevent the Spread of COVID-19 Epidemic', Iraqi Journal of Science, 2022, 63(9), pp.4102-4117.

[3] Al-Tai, A., Al-Kazaz, Q.: 'Semi parametric Estimators for Quantile Model via LASSO and SCAD with Missing Data', Journal of Economics and Administrative Sciences, 2022, 28(133), pp. 82-96.

[4] Alwan, E., Ali, O. : 'About Semi-parametric Methodology for Fuzzy Quantile Regression Model Estimation: A Review', Journal of Economics and Administrative Sciences, 2023, 29(138), pp. 127-146.

[5] Awoke, S., Getaneh, F., Derebe, M.:'Spatial patterns and determinants of low utilization of delivery care service and

postnatal check-up within 2 months following birth in Ethiopia:Bivariate analysis', Plos one,2024,19(1), pp.1-21.

[6] Benoit, D., Alhamzawi, R., Yu, K.: 'Bayesian lasso binary quantile regression', Computational Statistics, 2013, 28, pp. 2861-2873.

[7] Cameron. : 'Spatial Econometrics Introduction & Reading List'. PhD thesis, Texas A&M University, 2014.

[9] Chatterjee, S., and Hadi, A.S.: 'Regression analysis by example' (John Wiley & Sons, 2015).

[9] Chicco,D.,Warrens, M.J., Jurman,G.: 'The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation', Peerj computer science, 2021, 7:e623, pp. 1-24.

[10] Dai, X., Li, S., Tian, M.: 'Quantile regression for partially linear varying coefficient spatial autoregressive models', Communications in Statistics-Simulation andComputation, 2022, pp. 1-16.

[11] Fisher, W. :'Econometric estimation with spatial dependence', Regional and Urban Economics, 1971, 1(1), pp. 19-40.

[12] Gumprecht, D., 'Spatial Methods in Econometrics: An Application to R&D Spillovers' (Department of Statistics and Mathematics, 2005), pp. 1-16.

[13] Koenker, R., Bassett Jr, G. :'Regression quantiles', Econometrica: journal of the Econometric Society, 1978, 46(1), pp. 33-50.

[14] Kostov, P. :'A spatial quantile regression hedonic model of agricultural land prices', Spatial Economic Analysis, 2009, 4(1), pp. 53-72.

[15] Kottas, A., Krnjajić, M. :'Bayesian semiparametric modelling in quantile regression', Scandinavian Journal of Statistics, 2009, 36(2), pp. 297-319.

[16] Kotz, S., Kozubowski, T., Podgórski, K. :'The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance' (2001 No. 183). Springer Science & Business Media.

[17] Kozumi, H., Kobayashi, G. :'Gibbs sampling methods for Bayesian quantile regression', Journal of statistical computation and simulation, 2011, 81(11), pp. 1565-1578.

[18] Lee, J., Huang, Y. :'Social Vulnerability and COVID-19 Pandemic Outcomes: Evidence from Spatial Quantile Regression', Review of Regional Studies, 2024, 4(1), pp.27-52.

[19] Lee, J., Kim, J., Shin, J., et al. :'Analysis of wildfires and their extremes via spatial quantile autoregressive model', Extremes, 2023, 26(2), pp.353-379.

[20] LeSage James, P. :'Spatial Econometrics' (Department of Economics University of Toledo, CIRCULATED FOR REVIEW December, 1998).

[21]LeSage,J.:'The theory and practice of spatial econometrics' University of Toledo, Ohio, 1999, 28(11), pp. 1-39.

[22] Majid, H., Al-Bayati, M. :'Bayesian Tobit Quantile Regression Model Using Four Level Prior Distributions', Journal of Economics and Administrative Sciences, 2018, 24(105), pp. 487-513.

[23] Majid, H., Al-Bayati, M. :'Bayesian Tobit Quantile Regression Model Using Double Adaptive Elastic Net and Adaptive Ridge Regression', Journal of Economics and Administrative Sciences, 2018, 24(107), pp. 521-537.

[24] Marasinghe, D.: 'Quantile regression for climate data.' Master thesis, Clemson University, 2014.

[25] Mohammed, J., Aziz, M., Tawfeeq, A. :'COMPARE BETWEEN ORDINARY LEAST SQUARE AND MAXIMUM LIKELIHOOD METHODS FOR ESTIMATE PARAMETER OF FUZZY SPATIAL LAG MODEL', World Bulletin of Management & Law, 2022, 6, ISSN: 2749-3601, pp. 79-86.

[26] Ngwira, A. :'Spatial quantile regression with application to high and low child birth weight in Malawi', BMC Public Health, 2019, 19, pp. 1-11.

[27] Paelinck, P. & Klaassen, H. [with the assistance of J.-P. Ancot, A. C. P. Verster, Sj. Wagenaar]. :'Spatial Econometrics', Farnborough, Eng.: Saxon House, 1979.

[28] Permai, S., Mukhaiyar, U., PP, N. S., et al. :'Spatial weighting approach in numerical method for disaggregation of MDGs indicators'. IOP Conf. Seri. Materials Science and Engineering, March 2018, pp. 1-7.

[29] Rusche, K. :'Quality of life in the regions: an exploratory spatial data analysis for West German labor markets', Jahrbuch für Regionalwissenschaft, 2010, 30(1), pp. 1-22.

[30] Shawq, A., Globe, I., Hussaini, M. :'Comparison of Some Spatial Regression Models Using Simulation', Journal of Techniques, 2023, 5(2), pp. 197-205.

[31] Tribhuwaneswari, A., Hapsery, A., Rahayu, W. :'Spatial autoregressive quantile regression as a tool for modelling human development index factors in 2020 East Java'. AIP Conf. Proc. October 2022, 2668(1) pp. 1-11.

[32] Tybl, A. :'An overview of spatial econometrics', 2016, arXiv preprint arXiv:1605.03486.

[33] Wardhani,A.,Yanti,T. :'Pemodelan Spatial Autoregressive Quantile Regression (SARQR)pada Data Gizi Buruk Balita di Kota Bandung', Prosiding Statistika, 2021, pp. 606-612.

[34] Yang, Y., He, X. :'Quantile regression for spatially correlated data: An empirical likelihood approach', Statistica Sinica, 2015, pp. 261-274.

[35] Yu, K., Moyeed, R.:'Bayesian quantile regression', Statistics & Probability Letters, 2001, 54(4), pp. 437-447.

[36] Yu,K., Lu,Z., Stander,J.:'Quantile regression: applications and current research areas', Journal of the Royal Statistical Society Series D:The Statistician, 2003, 52(3), pp. 331-350.

[37] Zhang, H., Wang, X. :'Effectiveness of macro-regulation policies on housing prices: A spatial quantile regression approach',Housing,Theory&Society,2016,33(1), pp.23-40.