Earthline Journal of Mathematical Sciences
E-ISSN: 2581-8147
Volume 15, Number 3, 2025, Pages 381-399
https://doi.org/10.34198/ejms.15325.381399

Enhancing Spatial Autoregressive Models with Bootstrap Techniques: A Methodological Investigation into Bias, Precision, and Sample Size Effects

F. E. Itiveh[1] and C. O. Aronu[2,*]

[1] Department of Statistics, Delta State University of Science and Technology, Ozoro, Delta State, Nigeria

[2] Department of Statistics, Chukwuemeka Odumegwu Ojukwu University, Uli, Anambra State, Nigeria
e-mail: amaro4baya@yahoo.com

Abstract

This study introduces and evaluates two novel bootstrap-enhanced methods: the Bootstrap Simultaneous Autoregressive Lag Model (BSALM) and the Bootstrap Simultaneous Autoregressive Error Model (BSAEM), within the framework of classical Spatial Simultaneous Autoregressive (SAR) models. Using simulated datasets from normal distributions across varying sample sizes ($n = 10$ to 500) and secondary real-world data, the study examines their effectiveness in addressing spatial dependencies. The study's objectives include assessing bias, standard errors, variability, and the influence of sample size on model efficiency. Results demonstrate that both methods significantly reduce bias and variability as sample size increases, highlighting the critical role of adequate data dimensions in spatial analysis. BSALM consistently outperformed BSAEM in bias reduction, while BSAEM proved more adept at capturing complex spatial interdependencies despite exhibiting higher variability. Challenges with smaller datasets revealed increased biases and variability, emphasizing the importance of cautious interpretation in such scenarios. Real-world applications underscored dataset-specific performance variations, with BSALM excelling in bias correction and BSAEM managing intricate spatial structures. By integrating bootstrap techniques into SAR modelling, this study provides practical tools for enhancing predictive accuracy and model validation. While computational demands remain a consideration, these findings offer valuable insights into balancing bias, variability, and efficiency, paving the way for future advancements in spatial econometric analysis.

1. Introduction

In recent years, the Simultaneous Autoregressive (SAR) model has become instrumental across various fields, with research underscoring its capacity to reveal spatial relationships and dependencies in complex data. SAR models have proven essential in ecological studies to explore biodiversity distributions, in finance to examine market dynamics, and in public health to analyze disease spread patterns. Despite its broad applications, ongoing methodological advancements are continuously pushing the boundaries of SAR model capabilities to accommodate increasingly intricate datasets and provide more precise insights (Song et al. [16], Han et al. [8], Yao et al. [18]). A growing interest lies in integrating robust statistical techniques, notably bootstrap methods, into the SAR framework. Bootstrap techniques, celebrated for their flexibility and effectiveness in addressing uncertainty within spatial data, have the potential to enhance the SAR model's analytical robustness. The Bootstrap Method with Monte Carlo Integration (BMMCI) was presented in work by Al Luhayb [1] for one-dimensional integrals. The study used bootstrap samples from known distributions to

Received: January 25, 2025; Accepted: February 22, 2025; Published: February 27, 2025
2020 Mathematics Subject Classification: 62F40.
Keywords and phrases: model efficiency, bias minimization, spatial dependencies, sample size efficiency, predictive accuracy.
*Corresponding author

produce minimally variance-containing confidence ranges and robust standard error estimates. The Bootstrap Method (BM), which provides more accurate confidence intervals and hypothesis testing, was examined by Horowitz [10] as a technique for enhancing distributional approximations in econometrics, particularly in complicated contexts where analytical methods fall short. However, a thorough review of SAR literature reveals a critical gap: despite the effectiveness of bootstrap techniques in other econometric and statistical contexts, they remain underutilized and largely unexplored within SAR modelling. By evaluating the performance of bootstrap-enhanced SAR models, this study seeks to address this gap, offering a comprehensive examination of how these methods affect model bias, precision, and sample size dependencies.

The literature on SAR models spans diverse applications and methodological innovations. For instance, Zhang et al. [20] leveraged SAR models to investigate factors influencing biodiversity, identifying geological and climatic influences on species adaptations, while Kandula and Shaman [13] applied SAR models to optimize COVID-19 vaccine distribution strategies. Studies by Yao et al. [18] on pheasant biodiversity and by Vanli and Alawad [17] on public health monitoring further exemplify the SAR model's adaptability. However, while such studies underscore the SAR model's value in spatial data analysis, few examine the potential enhancements offered by bootstrap techniques. Recent contributions in SAR methodology, such as Gasperoni et al. [7], have introduced advanced model variations to improve stability and robustness in applications, yet an integrated focus on bootstrap within SAR remains absent. Also, the study by Feng et al. [6] introduced a novel Spatiotemporal Quantile-Function-Based Model (SQFBM) to capture individual heterogeneity using correlated random effects and heteroscedastic innovations. It employs a two-stage estimation combining moments and quantile regression, with a hybrid double bootstrap validating its robust application in air quality analysis. This void suggests an overlooked opportunity for improving SAR models, particularly in terms of their predictive accuracy and reliability under conditions of complex spatial dependency.

Therefore, this study aims to close this gap by systematically investigating the incorporation of bootstrap techniques within the SAR model framework. We propose to explore how Bootstrap methods can augment SAR models, particularly concerning bias, precision, and sample size effects. This research is not only methodologically relevant but also timely, given the growing complexity of spatial data in various disciplines, from waste management to ecological modelling. By bridging SAR and bootstrap methods, this study aspires to provide a versatile analytical tool that enhances the accuracy and interpretability of spatial models. The objectives of the study are as follows: to assess the performance of the proposed model in managing spatial dependence using both simulated and empirical data; to compare the biases of the proposed and traditional SAR methods; to analyze precision differences between bootstrap-enhanced and classical SAR Lag and Error models; to investigate the relationship between sample size and bias standard deviation in bootstrap SAR models; and to evaluate the precision and reliability of these enhanced methods. By addressing these objectives, the study seeks to provide a valuable methodological contribution to spatial econometrics, encouraging further interdisciplinary applications and laying the groundwork for future innovations in SAR modelling techniques.

2. Methods

2.1. Method of data collection

This section describes the methods used to collect the data needed to support the research objectives of the study. It highlights the variety and applicability of the datasets used, describing secondary sources. The sources of data used in this study encompass simulated data and secondary data. Simulated data were generated using random normal distribution across various sample sizes ($n = 10$, 15, 20, 25, 30, 40, 50, 100, 200, and 500), providing controlled inputs for statistical modeling and hypothesis testing.

Secondary data were primarily sourced from reputable institutions and databases:

Central Bank Statistical Bulletin 2021: Provides economic indicators and national accounts data.

World Bank National Accounts Data: Offers comprehensive macroeconomic data for cross-country comparisons.

International Labour Organization (ILOSTAT) Database: Contains global labor market statistics and indicators.

Description of Secondary Datasets

UNER, PGR, Pop (1970-2021): This dataset includes variables such as Unemployment Rate (UNER), Percentage Growth Rate (PGR), and Population Size (Pop) spanning from 1970 to 2021. It consists of 50 observations across 3 dimensions, enabling longitudinal analysis of economic and demographic trends.

R Console Repository Datasets:

Longley's Economic Regression Data (longley): A classic dataset illustrating multicollinearity issues in regression analysis, comprising 16 observations and 7 variables. It serves as a benchmark for assessing statistical methodologies.

Motor Trend Car Road Tests (mtcars): Features data on automotive attributes tested by Motor Trend magazine, facilitating exploratory data analysis and regression modeling with 32 observations and 11 variables.

Swiss Fertility and Socioeconomic Indicators (Swiss Fertility): Offers socio-economic metrics and fertility rates for Swiss provinces circa 1888, aiding historical demographic studies with 47 observations and 6 variables.

Columbus OH Spatial Analysis Dataset (Columbus): Provides demographic and socio-economic data for Columbus, Ohio census tracts, essential for spatial analysis and urban geography, containing 49 observations and 20 variables.

House Sales Prices, Baltimore (Baltimore): Covers housing price data and attributes for Baltimore, MD in 1978, supporting spatial hedonic regression analyses with 211 observations and 17 variables.

Performance of Computer CPUs (cpus): Includes specifications and performance metrics for 209 CPU models, facilitating comparative analysis and modeling with 209 observations and 9 variables.

The Effect of Punishment Regimes on Crime Rates (UScrime): Examines crime rate impacts of sentencing policies across 47 US states in 1960, offering rescaled data for criminological studies.

Data from 93 Cars on Sale in the USA in 1993 (Cars93): Details attributes of 93 car models available in the US in 1993, useful for regression modeling and market analysis with 93 observations and 18 variables.

Violent Crime Rates by US State (USArrests): Provides violent crime statistics per 100,000 people for all 50 US states in 1973, including urban population proportions and police officer densities.

These datasets collectively support diverse analytical approaches, from econometric modeling to spatial analysis and historical demographic studies, ensuring robustness and relevance in addressing the research questions posed in this study.

2.2. Method of data analysis

2.2.1. The proposed bootstrap based simultaneous autoregression (SAR) model

Given the simultaneous autoregression (SAR) model as

$$Y = \rho WY + \beta X + \varepsilon, \tag{1}$$

where:

Y is the response variable with dimension, $n \times 1$, where n is the number of spatial units or observations and represents the observed outcome for each spatial unit.

ρ is the spatial autoregressive coefficient with dimension, scalar and captures the strength of spatial dependence among observations.

W is the spatial weights matrix with dimension, $n \times n$ and encodes spatial relationships among the n units (e.g., adjacency or distance).

WY is the spatial lag of the response with dimension $n \times 1$, and represents the weighted average of neighboring responses, emphasizing spatial autocorrelation.

X is the predictor variable which represents the covariates influencing the response variable with dimension, $n \times p$, where p is the number of predictors.

β is the vector of the regression coefficient which measures the effect of each predictor variable on Y with dimension, $p \times 1$.

ε is the error term which captures residual variation not explained by the model with dimension $n \times 1$. $\varepsilon \sim N(0,\sigma^2 I)$, where: 0 is the mean vector of length n (errors are centered), σ^2 is the variance of the error terms, and I is the $n \times n$ identity matrix, implying no spatial correlation in ε.

The SAR model estimates the parameters ρ and β to capture the spatial dependence in the response variable Y. The model assumes a specific distributional form for the error term ε (e.g., Gaussian). Suppose the response variable Y and X the predictor variable in equation (1) are resampled using the bootstrap method, we shall then rewrite equation (1) as given:

$$Y^B = \rho WY^B + \beta X^B + \varepsilon, \tag{2}$$

where:

$Y^B = B(Y) = \{y_{1b}^*, y_{2b}^*, \cdots, y_{nb}^*\}$ (Bootstrapped response variable, generated from resampling Y, while y_{ib}^* is the *i-th* sampled value from Y for the *b-th* bootstrap iteration, and *n* is the sample size)

$X^B = B(X) = \{X_1^*, X_2^*, \cdots, X_B^*\}$ (The bootstrapped predictor variable bootstrapped predictor variable, generated from resampling X). This is $n \times p$ matrix of the bootstrapped predictor variables (including a column of ones if an intercept is included)

The spatial autoregressive model estimation method such as Spatial Two-Stage Least Squares (S2SLS) will be employed to estimate the unbiased coefficients of the proposed model presented as equation (2).

The Spatial Two-Stage Least Squares (S2SLS) Method

a. In the first stage, the spatial lag variable WY^B can be estimated by regressing Y^B on W:

$$WY^B = \lambda WY^B + \eta, \tag{3}$$

where:

λ is the coefficient for the spatial lag variable and η is the error term; $\eta \sim N(0, \Sigma_\eta)$ under the assumptions of the spatial lag model.

b. Obtain the predicted values of WY^B denoted as $W\hat{Y}^B$.

c. In the second stage, estimate the coefficients β and ρ by regressing Y^B on $W\hat{Y}^B$ and X^B:

$$Y^B = \rho W\hat{Y}^B + \beta X^B + \varepsilon. \tag{4}$$

The derivation of the estimates of the model using the Spatial Two-Stage Least Squares (S2SLS) method was given as:

Recall equation (2)

First Stage:

In the first stage, we shall estimate the spatial lag variable WY^B.

Step 1: Regress Y^B on W:

$$Y^B = \lambda WY^B + \eta, \tag{5}$$

where:

λ is the coefficient for the spatial lag variable and η is the error term.

Second Stage:

In the second stage, we shall estimate the coefficients β and ρ using the predicted values of WY^B from the first stage.

Step 2: Obtain the predicted values of WY^B (denoted as $W\hat{Y}^B$).

Step 3: Regress Y^B on $W\hat{Y}^B$ and X^B:

$$Y^B = \rho W\hat{Y}^B + \beta X^B + \varepsilon. \tag{6}$$

Refer to Appendix A for the proof demonstrating that the S2SLS estimators are unbiased.

The estimated values of ρ and β have were obtained as presented in equations (7) and (8).

$$\hat{\rho} = \frac{(X^B)^{\mathrm{T}} X^B (WY^B)^{\mathrm{T}} Y^B - (WY^B)^{\mathrm{T}} (WY^B)(X^B)^{\mathrm{T}} Y^B}{(WY^B)^{\mathrm{T}} (WY^B)(X^B)^{\mathrm{T}} X^B - [(WY^B)^{\mathrm{T}} (X^B)^{\mathrm{T}} X^B (WY^B)]}, \tag{7}$$

$$\hat{\beta} = \frac{(WY^B)^{\mathrm{T}} (WY^B)(X^B)^{\mathrm{T}} Y^B - (WY^B)^{\mathrm{T}} X^B (WY^B)^{\mathrm{T}} Y^B}{(WY^B)^{\mathrm{T}} (WY^B)(X^B)^{\mathrm{T}} X^B - [(WY^B)^{\mathrm{T}} (X^B)^{\mathrm{T}} X^B (WY^B)]}, \tag{8}$$

where T denotes the transpose of a matrix or vector. The estimation of coefficients using the Ordinary Least Squares (OLS) Method of equations (7) and (8) is presented in Appendix B.

The Bootstrap based Spatial Simultaneous Autoregressive Lag Model (BSALM) can be expressed as equation (2) while the Bootstrap based Spatial Simultaneous Autoregressive Error Model (BSAEM) as

$$Y^B = \lambda WY^B + \beta X^B + \varepsilon, \tag{9}$$

where:

$Y^B = B(Y)$ (The bootstrapped response variable)

$X^B = B(X)$ (The bootstrapped predictor variable)

λ is the spatial autocorrelation parameter for errors.

Where the estimate of the spatial autocorrelation parameter λ is derived as (see Appendix C):

$$\hat{\lambda} = \frac{(Y^B)^{\mathrm{T}} W Y^B - \beta (X^B)^{\mathrm{T}} W Y^B}{\lambda (W Y^B)^{\mathrm{T}} W Y^B}. \tag{10}$$

The bootstrap method proves invaluable in spatial autoregressive (SAR) models for several reasons. Firstly, it addresses limited sample sizes by generating multiple pseudo-samples through resampling, thereby enhancing the precision of parameter estimation. Secondly, SAR models often encounter non-normal error terms and nonlinear relationships, where traditional asymptotic methods may falter; the bootstrap's non-parametric nature bypasses such assumptions, ensuring robust inference. Moreover, it accommodates spatial dependence inherent in SAR models by resampling while preserving spatial structure, enabling efficient inference and uncertainty quantification. Furthermore, the bootstrap simplifies hypothesis testing and confidence interval construction, providing empirical estimates of parameter distributions. Lastly, it aids in model selection by evaluating SAR model performance across bootstrap samples, ensuring reliable and stable results across diverse data subsets. Thus, the bootstrap method stands as a versatile and crucial tool for bolstering inference in SAR models, irrespective of data characteristics or spatial complexities.

Assumptions of the bootstrap based simultaneous autoregression (SAR) model

i. The original dataset D is a representative sample of the population.

ii. The statistic of interest is unbiased and has a sampling distribution that is approximately normal.

iii. The original statistic, computed on the original dataset D, provides an estimate of the population parameter of interest.

iv. Due to sampling variability, the original statistic may deviate from the true population parameter.

v. By generating r resamples from D through bootstrapping, each resample captures different subsets of observations from D.

vi. Each resample, S_r^* contains information about the population, similar to that contained in the original dataset D.

vii. By computing the statistic of interest on each resample, S_r^*, we obtain a distribution of resample-based statistics.

viii. This distribution of resample-based statistics represents the variability in estimating the population parameter.

ix. The distribution of resample-based statistics can provide insights into the sampling variability, robustness, and uncertainty associated with the estimation.

x. If the original statistic is affected by sampling variability, it may not fully capture the true population parameter.

xi. Aggregating the information from multiple resamples provides a more comprehensive understanding of the statistic's behavior and potential improvements in estimation.

xii. By examining the distribution of resample-based statistics, we can assess the precision, confidence intervals, and potential bias associated with the estimation.

In assessing the performance of the resampling statistic (T_p) and other test statistics, bias, efficiency, and variance are pivotal metrics. To assess the performance of the resampling statistic (T_p) and other test statistics based on bias, we can use mathematical techniques to quantify and compare their biases. The bias of an estimator measures the systematic deviation of the expected value of the estimator from the true value of the parameter being estimated (Casella and Berger [3]). Let us denote the true value of the parameter as θ, and the estimator for θ as T. The bias of the estimator T, denoted as $Bias(T)$, is defined as:

$$Bias(T) = E(T) - \theta, \tag{11}$$

where,

$E(T)$ represents the expected value of the estimator T.

To evaluate the performance of the resampling statistic T_p and other test statistics based on bias, we need to compare their biases to zero. A bias of zero indicates that, on average, the estimator provides unbiased estimates of the true parameter (Efron and Tibshirani [5]). To compute the bias of the resampling statistic T_p, we calculate the average value of $T_p$over multiple resamples and subtract the true parameter value:

$$\left|Bias(T_p)\right| = E(T_p) - \theta. \tag{12}$$

Similarly, for other test statistics, we compute their biases using the same formula:

$$|Bias(T_1)| = E(T_1) - \theta$$

$$|Bias(T_2)| = E(T_2) - \theta$$

$$\vdots$$

$$|Bias(T_n)| = E(T_n) - \theta.$$

Once we have obtained the biases of the resampling statistic T_p and other test statistics, we can compare them. A smaller bias, closer to zero, indicates a more desirable estimator as it suggests that the estimator tends to provide unbiased estimates of the true parameter (Hastie et al. [9]).

Efficiency quantifies how accurately an estimator estimates a parameter, with lower variance indicating higher precision (Casella and Berger [3]). Variance, denoted as $Var(T_p)$, measures the spread of estimator values around their expected values. A lower variance signifies a more stable and reliable estimation, making the estimator more efficient. The formula for the variance of the resampling statistic (T_p) is given as:

$$Var(T_p) = E\left[\left(T_p - E(T_p)\right)^2\right], \tag{13}$$

where:

$E(T_p)$ is the expected value of the resampling statistic T_p, and

$\left(T_p - E(T_p)\right)^2$represents the squared deviation of T_p from its expected value.

Similarly, for other test statistics ($T_1, T_2,\ldots,T_n$), their variances are computed using the same general formula:

$$Var(T_i) = E(T_i - E(T_i))^2], \quad i = 1,2,\ldots,n. \tag{14}$$

Efficiency and variance are closely related, as lower variance implies greater precision in estimating the parameter of interest. Comparing the variances of different estimators helps identify the one that provides the most stable and reliable estimates under the given conditions (Efron and Tibshirani [5]). Evaluating both bias and efficiency allows researchers to select estimators that balance accuracy and precision, tailored to the specifics of the estimation problem and the underlying assumptions (Hastie et al. [9]).

3. Results

3.1 Evaluating the performance of the classic SAR Lag and SAR error method against the proposed Bootstrap SAR Lag and Bootstrap SAR Error methods using simulated data from Normal Distribution

Using simulated data from a normal distribution, we compare the effectiveness of the proposed Bootstrap SAR Lag and Bootstrap SAR Error methods against the traditional SAR Lag and SAR Error methods in this section. The mean bias, standard deviation of bias, mean standard error, and standard deviation of standard error for each approach were compared in Tables 1 through 4 where the results are displayed.

The results in Table 1 present the mean bias for Bootstrap SAR Lag and Bootstrap SAR Error methods across varying dimensions of simulated data generated from a normal distribution. Bias measures the systematic deviation between the estimator's expected value and the true parameter. Ideally, lower bias indicates higher accuracy.

For smaller dimensions (e.g., 10 X 3), both methods exhibit relatively higher bias, with Bootstrap SAR Lag at 0.0780 and Bootstrap SAR Error at 0.1020. As the dimensions increase, the biases decrease on average, with the overall mean bias being 0.0457 for Bootstrap SAR Lag and 0.0535 for Bootstrap SAR Error. These findings highlight the importance of sample size, as larger datasets tend to reduce systematic errors, improving estimator performance.

The results suggest that both methods achieve higher accuracy with increasing sample size, though Bootstrap SAR Lag consistently shows slightly lower bias than Bootstrap SAR Error. This makes Bootstrap SAR Lag more efficient in terms of bias minimization. However, for datasets with smaller dimensions, careful interpretation is necessary as both methods may introduce notable systematic errors. Practitioners should consider dataset size and complexity when choosing between these approaches to balance bias and efficiency.

Table 1: Mean value of Bias for Bootstrap SAR Lag and Bootstrap SAR Error using Simulated Data from Normal Distribution

Dimension	Boot.SAR.lag	Boot.SAR.error
10 X 3	0.0780	0.1020
15 X 3	0.0444	0.0460
20 X 3	0.1200	0.1322
25 X 3	0.0260	0.0347
30 X 3	0.0120	0.0187

40 X 3	0.0940	0.1030
50 X 3	0.0170	0.0242
100 X 3	0.0280	0.0312
200 X 3	0.0233	0.0264
500 X 3	0.0144	0.0162
Mean	0.0457	0.0535

The results in Table 2 present the standard deviation of bias for the Bootstrap Spatial Autoregressive Lag (Bootstrap SAR Lag) and Bootstrap Spatial Autoregressive Error (Bootstrap SAR Error) methods. These standard deviations represent the variability of the bias around its mean value, providing a measure of the efficiency of each technique. Variability is higher in smaller dimensions, with standard deviations of 0.1680 for Bootstrap SAR Lag and 0.2250 for Bootstrap SAR Error at 10 × 3 dimensions. As dimensions increase, the standard deviations decrease, suggesting improved consistency and efficiency in larger samples.

Overall, the total standard deviation across dimensions is 0.9837 for Bootstrap SAR Lag and 1.2526 for Bootstrap SAR Error using simulated data from a normal distribution. These findings imply that the Bootstrap SAR Lag method exhibits greater efficiency and stability in estimating biases, particularly in scenarios with increasing sample sizes and dimensions, compared to the Bootstrap SAR Error method.

Table 2: Standard Deviation (SD) measure of Bias for Bootstrap SAR Lag and Bootstrap SAR Error using Simulated Data from Normal Distribution

Dimension	Boot.SAR.lag	Boot.SAR.error
10 X 3	0.1680	0.2250
15 X 3	0.1218	0.2124
20 X 3	0.2227	0.3129
25 X 3	0.0747	0.0754
30 X 3	0.0542	0.0644
40 X 3	0.1727	0.1760
50 X 3	0.0256	0.0442
100 X 3	0.0560	0.0610
200 X 3	0.0544	0.0484
500 X 3	0.0336	0.0329
Total SD	0.9837	1.2526

The results in Table 3 compare the mean standard errors for the Spatial Autoregressive Lag (SAR Lag), Spatial Autoregressive Error (SAR Error), Bootstrap SAR Lag, and Bootstrap SAR Error methods using simulated data from a normal distribution. The Bootstrap methods generally exhibit higher mean standard errors than the classic methods, reflecting their tendency to provide more conservative estimates of variability. This conservativeness is advantageous for mitigating underestimating variability and enhancing reliability in inferential conclusions.As the sample size increases, the mean standard errors for all methods decrease, demonstrating improved stability and precision of the estimates in larger samples. Specifically, the Bootstrap SAR Lag and Bootstrap SAR Error methods show notable reductions in standard errors with increasing dimensions, indicating their efficiency in providing robust estimates under larger sample sizes.

The overall mean standard error across dimensions is 0.2140 for Bootstrap SAR Lag and 0.2305 for Bootstrap SAR Error, compared to 0.1703 for SAR Lag and 0.1705 for SAR Error. While the classic methods show slightly lower standard errors, the efficiency of the Bootstrap methods lies in their ability to balance bias correction and variability, making them suitable for applications requiring robust standard error estimation, particularly in smaller sample scenarios where variability is inherently higher.

Table 3: Mean value of Standard Error for SAR Lag, SAR Error, Bootstrap SAR Lag and Bootstrap SAR Error using Simulated Data from Normal Distribution

Dimension	SAR.lag	SAR.error	Boot.SAR.lag	Boot. SAR.error
10 X 3	0.4263	0.4246	0.6760	0.7358
15 X 3	0.1741	0.1549	0.2082	0.2624
20 X 3	0.1906	0.2063	0.2689	0.2901
25 X 3	0.2065	0.2053	0.2293	0.2388
30 X 3	0.1853	0.1841	0.2069	0.2182
40 X 3	0.1398	0.1498	0.1597	0.1656
50 X 3	0.1541	0.1540	0.1600	0.1627
100 X 3	0.1036	0.1060	0.1099	0.1108
200 X 3	0.0726	0.0707	0.0725	0.0728
500 X 3	0.0496	0.0489	0.0481	0.0482
Mean	0.1703	0.1705	0.2140	0.2305

The results in Table 4 present the standard deviation of the standard error for the Spatial Autoregressive Lag (SAR Lag), Spatial Autoregressive Error (SAR Error), Bootstrap SAR Lag, and Bootstrap SAR Error methods using simulated data from a normal distribution. Consistent with the findings for mean standard errors, the standard deviations of the standard errors are higher for the Bootstrap methods compared to the classic techniques, highlighting their more conservative nature. This increased variability reflects the Bootstrap methods' emphasis on capturing a broader range of uncertainty in the estimates. The standard deviations decrease as sample size increases across all methods, indicating that the variability in the standard error estimates reduces with larger datasets. This trend demonstrates improved stability and precision of the estimates in larger samples, reaffirming the efficiency of all methods under sufficient data conditions.

The total standard deviation across dimensions is 0.4064 for SAR Lag, 0.5096 for SAR Error, 0.6316 for Bootstrap SAR Lag, and 0.8187 for Bootstrap SAR Error. While the Bootstrap methods exhibit higher overall variability, this is a trade-off for their robustness in accounting for uncertainty, making them especially valuable in small-sample scenarios or when the underlying data distribution deviates from ideal assumptions. These results emphasize the Bootstrap methods' utility in providing reliable estimates, albeit with a higher variability compared to traditional approaches.

Table 4: Standard Deviation measure of Standard Error for SAR Lag, SAR Error, Bootstrap SAR Lag and Bootstrap SAR Error using Simulated Data from Normal Distribution

Dimension	SAR.lag	SAR.error	Boot.SAR.lag	Boot.SAR.error
10 X 3	0.1010	0.1213	0.2279	0.1953
15 X 3	0.1196	0.1646	0.1363	0.2548
20 X 3	0.0297	0.0509	0.0447	0.0915
25 X 3	0.0344	0.0377	0.0520	0.0665
30 X 3	0.0418	0.0438	0.0642	0.0822

40 X 3	0.0200	0.0299	0.0373	0.0481
50 X 3	0.0224	0.0222	0.0313	0.0386
100 X 3	0.0158	0.0150	0.0178	0.0205
200 X 3	0.0148	0.0166	0.0138	0.0147
500 X 3	0.0069	0.0076	0.0063	0.0065
Total SD	0.4064	0.5096	0.6316	0.8187

3.2. Evaluating the performance of the classic SAR Lag and SAR error method against the proposed Bootstrap SAR Lag and Bootstrap SAR error methods using Real Life data

This section compares, using real-world data, the performance of the Bootstrap SAR Lag and Bootstrap SAR Error methods to the standard SAR Lag and SAR Error approaches. Tables 5-8 provide a summary of the findings, analyzing mean bias, standard deviation of bias, mean standard error, and standard deviation of standard error across many datasets.

Table 5: Mean value of Bias for Bootstrap SAR Lag and Bootstrap SAR Error using Secondary dataset

S/No.	Name of Dataset	Dimension	Boot.SAR.lag	Boot.SAR.error
1	Longley's Economic Regression Data (longley)	16 X 7	0.5388	0.6661
2	Motor Trend Car Road Tests (mtcars)	32 X 11	0.5531	0.3200
3	Swiss Fertility and Socioeconomic Indicators (Swiss Fertility)	47 X 3	1.6678	0.16122
4	Columbus OH spatial analysis data set (Columbus)	49 X 22	1.5839	0.31172
5	House sales prices, Baltimore (Baltimore)	211 X 17	0.1297	0.02326
6	Performance of Computer CPUs (cpus)	209 X 9	0.9831	0.22866
7	The Effect of Punishment Regimes on Crime Rates (UScrime)	47 X 16	28.2375	81.3975
8	Data from 93 Cars on Sale in the USA in 1993 (Cars93)	93 X 27	0.6787	0.06031
9	Violent Crime Rates by US State (USArrests)	50 X 4	0.2924	0.0340
10	Real life Data	52 X 3	0.11125	0.0924
	Total Mean		34.7763	83.2952

The results in Table 5 present the mean bias for the Bootstrap Spatial Autoregressive Lag (Bootstrap SAR Lag) and Bootstrap Spatial Autoregressive Error (Bootstrap SAR Error) methods across ten real-life datasets.

The bias values vary significantly across datasets, with some exhibiting relatively high biases, such as the UScrime Dataset (28.2375 for Bootstrap SAR Lag, 81.3975 for Bootstrap SAR Error), while others show much lower biases, as observed in the House Sales Prices, Baltimore Dataset (0.1297 for Bootstrap SAR Lag, 0.02326 for Bootstrap SAR Error). Notably, no negative biases are observed across the datasets.

The total mean bias across all datasets is 34.7763 for Bootstrap SAR Lag and 83.2952 for Bootstrap SAR Error. This notable difference highlights the greater sensitivity of the Bootstrap SAR Error method to underlying dataset characteristics, potentially due to its modelling assumptions and the complexity of spatial error structures. However, the disparity in bias between datasets highlights the significance of taking dataset-specific issues into account from an efficiency standpoint. The particular spatial dependencies and heterogeneities found in the data should guide the selection of these approaches, which should strike a balance between the complexity of the spatial modelling framework and the requirement for accuracy.

Table 6: Standard Deviation measure of Bias for Bootstrap SAR Lag and Bootstrap SAR Error using Secondary dataset

S/No.	Name of Dataset	Dimension	Boot.SAR.lag	Boot.SAR.error
1	Longley's Economic Regression Data (longley)	16 X 7	1.0678	1.4012
2	Motor Trend Car Road Tests (mtcars)	32 X 11	1.2118	1.1487
3	Swiss Fertility and Socioeconomic Indicators (Swiss Fertility)	47 X 3	3.5698	0.6148
4	Columbus OH spatial analysis data set (Columbus)	49 X 22	3.3299	0.9129
5	House sales prices, Baltimore (Baltimore)	211 X 17	0.3475	0.2108
6	Performance of Computer CPUs (cpus)	209 X 9	1.9947	0.5289
7	The Effect of Punishment Regimes on Crime Rates (UScrime)	47 X 16	56.466	163.28
8	Data from 93 Cars on Sale in the USA in 1993 (Cars93)	93 X 27	1.449	0.1969
9	Violent Crime Rates by US State (USArrests)	50 X 4	0.6732	0.2005
10	Real life Data	52 X 3	0.2719	0.1012
	Total SD		70.3816	168.5959

The results in Table 6 present the standard deviation of bias for the Bootstrap Spatial Autoregressive Lag (Bootstrap SAR Lag) and Bootstrap Spatial Autoregressive Error (Bootstrap SAR Error) methods across ten real-life datasets. The standard deviation is highest for The Effect of Punishment Regimes on Crime Rates dataset (56.466 for Bootstrap SAR Lag and 163.28 for Bootstrap SAR Error), indicating exceptionally high variability in the bias for this dataset. In contrast, datasets like House Sales Prices, Baltimore exhibit relatively low standard deviations (0.3475 for Bootstrap SAR Lag and 0.2108 for Bootstrap SAR Error), reflecting more stable bias values.

The total standard deviation across all datasets is 70.3816 for Bootstrap SAR Lag and 168.5959 for Bootstrap SAR Error. This significant difference underscores the higher variability in bias for the Bootstrap SAR Error method, likely due to the complexities of spatial error structures it captures. These findings suggest that while both Bootstrap methods introduce variability in bias, Bootstrap SAR Lag generally demonstrates more consistent performance across datasets compared to Bootstrap SAR Error. The higher standard deviations for Bootstrap SAR Error indicate that its bias is more sensitive to dataset-specific factors, particularly in cases with complex or heterogeneous spatial structures. When choosing a method, practitioners should weigh the trade-offs between stability and flexibility from the standpoint of efficiency. Because of its reduced overall variability, Bootstrap SAR Lag can be more appropriate for applications that need reliable bias performance. On the other hand, even while it increases variability, Bootstrap SAR Error might offer a more accurate depiction of geographical interdependence, which makes it useful for datasets with especially intricate spatial error structures.

These findings also demonstrate how crucial it is to comprehend the fundamental features of any dataset since they have a direct impact on the effectiveness and dependability of the Bootstrap techniques. To choose amongst these approaches in practice and optimise performance while taking into consideration the unpredictability present in real-world spatial data, thorough dataset-specific diagnostic evaluations are crucial.

Table 7: Mean value of Standard Error for SAR Lag, SAR Error, Bootstrap SAR Lag and Bootstrap SAR Error using Secondary dataset

S/No.	Name of Dataset	Dimension	Boot.SAR.lag	SAR.lag	Boot.SAR.error	SAR.error
1	Longley's Economic Regression Data (longley)	16 X 7	16.4191	12.7573	16.2942	10.455
2	Motor Trend Car Road Tests (mtcars)	32 X 11	1.15417	1.15901	0.87022	0.8596
3	(Swiss Fertility)	47 X 3	1.36919	1.29115	0.89179	0.8025
4	(Columbus)	49 X 22	1.89271	1.76224	1.33905	1.3046
5	(Baltimore)	211 X 17	2.22163	1.69176	1.79655	1.2394
6	(cpus)	209 X 9	1.59088	1.37811	1.42762	1.1493
7	(UScrime)	47 X 16	255.383	232.987	256.045	227.26
8	(Cars93)	93 X 27	1.61124	1.54227	1.47603	1.3793
9	(USArrests)	50 X 4	0.52176	0.57983	0.48151	0.4657
10	Real life Data	52 X 3	0.69478	0.55366	0.5473	0.4513
	Total Mean		282.8585	255.7023	281.1693	245.3667

The results in Table 7 compare the mean standard errors for SAR Lag, SAR Error, Bootstrap SAR Lag, and Bootstrap SAR Error methods using real-life datasets. Bootstrap methods exhibit consistently higher mean standard errors than the classic methods, indicating a tendency for more conservative error estimation. The mean standard error varies significantly across datasets, such as UScrime (255.383 for Bootstrap SAR Lag and 256.045 for Bootstrap SAR Error) and Baltimore (2.22163 for Bootstrap SAR Lag and 1.79655 for Bootstrap SAR Error).

The higher mean standard errors of Bootstrap methods suggest their robustness in addressing uncertainty, particularly in datasets with complex spatial dependencies like UScrime. However, the more conservative estimates might compromise efficiency in datasets with simpler spatial structures, such as USArrests. Classic methods, while less conservative, may yield underestimates in cases involving substantial variability. Practitioners must balance the trade-off between conservatism and precision, using Bootstrap methods for datasets with high spatial or structural complexity while favouring classic methods for simpler scenarios to enhance efficiency.

Table 8: Standard Deviation measure of Standard Error for SAR Lag, SAR Error, Bootstrap SAR Lag and Bootstrap SAR Error using Secondary dataset

S/No.	Name of Dataset	Dimension	Boot.SAR.lag	SAR.lag	Boot.SAR.error	SAR.error
1	Longley's Economic Regression Data (longley)	16 X 7	32.416	25.1871	31.837	20.386
2	Motor Trend Car Road Tests (mtcars)	32 X 11	1.839	1.78025	1.1604	1.1119
3	(Swiss Fertility)	47 X 3	2.5537	2.40371	1.5864	1.4162
4	(Columbus)	49 X 22	3.2743	3.16639	2.1316	2.2312
5	(Baltimore)	211 X 17	2.4852	2.04803	1.9567	1.3442
6	(cpus)	209 X 9	3.1225	2.71403	2.7549	2.221
7	(UScrime)	47 X 16	508.35	463.713	509.57	452.28
8	(Cars93)	93 X 27	2.5255	2.32359	2.2487	1.9841
9	(USArrests)	50 X 4	0.9318	1.04958	0.8059	0.7962
10	Real life Data	52 X 3	0.9878	0.79793	0.6603	0.5489
	Total Standard Deviation		558.4858	505.1836	554.7119	484.31197

The results in Table 8 present the standard deviation of the standard errors for SAR Lag, SAR Error, Bootstrap SAR Lag, and Bootstrap SAR Error methods across real-life datasets. The UScrime dataset shows the highest standard deviation for all methods, indicating substantial variability in the error estimates for this dataset. Overall, the total standard deviations are as follows: 558.4858 for SAR Lag, 505.1836 for SAR Error, 554.7119 for Bootstrap SAR Lag, and 484.31197 for Bootstrap SAR Error. The lower overall standard deviations for the Bootstrap SAR Error method (484.31197) suggest that it provides more consistent error estimates compared to the Bootstrap SAR Lag (554.7119) and traditional SAR Lag (558.4858). However, traditional SAR Error (505.1836) also exhibits reasonable consistency, but it may underestimate the variability in certain datasets with complex spatial structures.

These results highlight that Bootstrap SAR methods, despite showing slightly higher variability in some datasets, can offer improved error estimation reliability due to their ability to account for more complex spatial dependencies. Practitioners should carefully assess these methods based on the specific variability and spatial characteristics of their datasets to optimize efficiency and reliability in statistical modelling.

4. Conclusion

The study aimed to develop and evaluate two innovative bootstrap methods: Bootstrap Simultaneous Autoregressive Lag Model (BSALM) and Bootstrap Simultaneous Autoregressive Error Model (BSAEM) within the classical spatial simultaneous autoregressive (SAR) framework. Using simulated normal distribution data and real-world datasets, the study examined the methods' performance in terms of bias, standard errors, and variability across varying sample dimensions. The findings underscore the critical interplay between sample size and methodological efficiency in spatial modelling.

The results highlight that both Bootstrap methods reduce bias and variability as sample size increases, consistent with findings by Yao et al. [18] and Vanli and Alawad [17], which emphasize the role of larger datasets in improving SAR model precision and reliability. For smaller datasets, Bootstrap SAR Lag showed lower mean bias (0.0780) compared to Bootstrap SAR Error (0.1020), making it more efficient in bias minimization. However, both methods demonstrated higher biases in smaller dimensions, necessitating caution in interpretation when using limited data. Bootstrap methods exhibited higher mean standard errors than classical SAR approaches, reflecting their conservative nature in error estimation. This aligns with the literature on bootstrap techniques, which suggests that conservatism enhances robustness in accounting for variability. For instance, the Bootstrap SAR Lag method demonstrated an overall mean standard error of 0.2140, slightly lower than the Bootstrap SAR Error (0.2305), reinforcing its reliability in bias correction and variability management.

Real-world datasets revealed substantial variability in bias and standard errors, influenced by spatial complexities. For example, the UScrime dataset exhibited high mean biases (28.2375 for Bootstrap SAR Lag and 81.3975 for Bootstrap SAR Error), while datasets like Baltimore had significantly lower biases (0.1297 and 0.02326, respectively). Variability in error estimates was also dataset-specific, with Bootstrap SAR Error showing a relatively lower overall standard deviation of standard errors (484.31197) compared to Bootstrap SAR Lag (554.7119), suggesting its suitability for scenarios involving complex spatial dependencies.

Theoretical advantages of Bootstrap methods provide further justification for their observed performance. These methods are particularly adept at addressing model misspecifications, a common challenge in spatial econometrics. By resampling directly from the data, bootstrap techniques capture inherent complexities and uncertainties more effectively than traditional approaches. Their non-parametric nature allows for flexibility in addressing violations of standard assumptions, such as normality and homoscedasticity. This robustness enhances both predictive accuracy and model validation, as evidenced by their superior performance in this study.

The study bridges a methodological gap by integrating Bootstrap methods into SAR modelling, providing robust tools for predictive accuracy and model validation. Practitioners are advised to select methods based on sensitivity to spatial interdependence, dataset size, and variability. Bootstrap SAR Lag may be preferred for consistent performance across datasets, while Bootstrap SAR Error better captures intricate spatial structures despite higher variability. Limitations include potential biases in secondary data due to collection heterogeneities and the computational intensity of bootstrap resampling, which challenges scalability for large datasets. Future research should explore optimized sample sizes and alternative bootstrapping techniques to balance computational feasibility and methodological adaptability.

In conclusion, this study contributes significantly to spatial autoregressive modelling by offering insights into bias, variability, and methodological efficiency, reinforcing the necessity of robust statistical approaches for reliable spatial analysis. These findings align with prior studies, including Yao et al. [18] and Vanli and Alawad [17], and provide a foundation for future advancements in spatial econometric methods.

References

[1] Al Luhayb, A. S. M. (2023). The bootstrap method for Monte Carlo integration inference. *Journal of King Saud University - Science, 35*(6), 102768. https://doi.org/10.1016/j.jksus.2023.102768

[2] Arisanti, R., Utami, E. R., Muslim, A., & Hayati, M. (2023). The relationship between economic growth and e-commerce at the beginning of COVID-19 pandemic in East Java. *Decision Science Letters, 12*(1), 149-162. https://doi.org/10.5267/j.dsl.2022.9.002

[3] Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Duxbury Press.

[4] Dantas, A., & Fonseca, C. R. (2023). Global biogeographical patterns of ants and their abiotic determinants. *Perspectives in Ecology and Conservation, 21*(3), 237-246. https://doi.org/10.1016/j.pecon.2023.07.003

[5] Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall. https://doi.org/10.1007/978-1-4899-4541-9

[6] Feng, X., Li, W., & Zhu, Q. (2024). Estimation and bootstrapping under spatiotemporal models with unobserved heterogeneity. *Journal of Econometrics, 238*(1), 105559. https://doi.org/10.1016/j.jeconom.2023.105559

[7] Gasperoni, F., Luati, A., Paci, L., & D'Innocenzo, E. (2023). Score-driven modeling of spatio-temporal data. *Journal of the American Statistical Association, 118*(542), 1066-1077. https://doi.org/10.1080/01621459.2021.1970571

[8] Han, J., Liu, G., Bai, W., Zou, Q., Cao, Y., Zhou, C., & Williams, G. M. (2022). Relationship between human disturbance and habitat use by the endangered François' langur (Trachypithecus francoisi) in Mayanghe Nature Reserve, China. *Pakistan Journal of Zoology, 54*(1), 191-198. https://doi.org/10.17582/journal.pjz/20200704020744

[9] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

[10] Horowitz, J. L. (2019). Bootstrap methods in econometrics. *Annual Review of Economics, 11,* 193-224. https://doi.org/10.1146/annurev-economics-080218-025651

[11] Jaya, I. G. N. M., & Chadidjah, A. (2021). Spatial autoregressive in ecological studies: A comparison of the SAR and CAR models. *Engineering Letters, 29*(1), 207-212.

[12] Jha, S. K., & Singh, N. V. (2023). A skew-normal spatial simultaneous autoregressive model and its implementation. *Sankhya A, 85*(1), 306-323. https://doi.org/10.1007/s13171-021-00246-3

[13] Kandula, S., & Shaman, J. (2021). Investigating associations between COVID-19 mortality and population-level health and socioeconomic indicators in the United States: A modeling study. *PLOS Medicine, 18*(7), Article e1003693. https://doi.org/10.1371/journal.pmed.1003693

[14] Li, N., Sun, J. T., Yin, Y., Hong, X. Y., & Xue, X. F. (2023). Global patterns and drivers of herbivorous eriophyoid mite species diversity. *Journal of Biogeography, 50*(2), 330-340. https://doi.org/10.1111/jbi.14535

[15] Pratschke, J., & Abbiati, G. (2023). "Like with like" or "do like?" Modeling peer effects in the classroom. *Social Science Quarterly, 104*(3), 265-280. https://doi.org/10.1111/ssqu.13257

[16] Song, Y., Liang, X., Zhu, Y., & Lin, L. (2021). Robust variable selection with exponential squared loss for the spatial autoregressive model. *Computational Statistics and Data Analysis, 155*, 107094. https://doi.org/10.1016/j.csda.2020.107094

[17] Vanli, O. A., & Alawad, N. (2023). Spatiotemporal health surveillance accounting for risk factors and spatial correlation. *Quality and Reliability Engineering International, 39*(6), 2258-2276. https://doi.org/10.1002/qre.3335

[18] Yao, H., Wang, P., Wang, N., McGowan, P. J. K., Si, X., Li, J., & Xu, J. (2022). Functional and phylogenetic structures of pheasants in China. *Avian Research, 13*, Article 100041. https://doi.org/10.1016/j.avrs.2022.100041

[19] Zhang, L., Cheng, D., Xue, W., & Yang, T. (2022). Combined change point estimation in threshold quantile autoregressive models. *Scientia Sinica Mathematica, 52*(1), 63-84. https://doi.org/10.1360/N012019-00167

[20] Zhang, Y., Chen, J., & Sun, H. (2021). Alpine speciation and morphological innovations: Revelations from a species-rich genus in the Northern Hemisphere. *AoB PLANTS, 13*(3), plab018. https://doi.org/10.1093/aobpla/plab018

Appendix

A. Proof demonstrating that the S2SLS estimators are unbiased

To prove that the S2SLS estimators are unbiased, we need to show that the estimators are consistent and have a finite variance.

Derivation:

Step 1: First Stage Estimation

By regressing Y^B on W, we have equation (5) written as:

$$Y^B = \lambda W Y^B + \eta$$

Rewriting the equation (4), we get:

$$(1 - \lambda W) Y^B = \eta$$

Simplifying, we obtain:

$$Y^B = (1 - \lambda W)^{-1} \eta \tag{5.1}$$

Step 2: Obtaining $W\hat{Y}$

The predicted values of WY^B denoted as $W\hat{Y}^B$, are given by:

$$W\hat{Y}^B = (1 - \lambda W)^{-1} \eta \tag{5.2}$$

Step 3: Second Stage Estimation

By regressing Y^Bon $W\hat{Y}^B$and X^B, we have:

$$Y^B = \rho W\hat{Y}^B + \beta X^B + \varepsilon \tag{6}$$

B. Estimation of Coefficients using Ordinary Least Squares (OLS) Method

To estimate the coefficients ρ and β, we use the ordinary least squares (OLS) estimation, which provides consistent and unbiased estimators under the classical assumptions of OLS. The steps to theoretically estimate the coefficients ρ and β using the Ordinary Least Squares (OLS) estimation method can be expressed as:

Given equation (6) for the model:

The objective of OLS estimation is to minimize the sum of squared residuals (SSE) between the observed values of Y^B and the predicted values based on the model.

$$SSE = \varepsilon^\tau \varepsilon = \left(Y^B - \rho W\hat{Y}^B - \beta X^B\right)^\tau \left(Y^B - \rho W\hat{Y}^B - \beta X^B\right) \tag{6.1}$$

To minimize the SSE, we compute partial derivatives with respect to ρ and β, and set them to zero.

$$\frac{\partial SSE}{\partial \rho} = \left(Y^B - \rho WY^B - X^B\beta\right)^\tau (WY^B) = 0 \tag{6.2}$$

$$\frac{\partial SSE}{\partial \beta} = \left(Y^B - \rho WY^B - X^B\beta\right)^\tau (X^B) = 0 \tag{6.3}$$

At this point, we can then solve simultaneously to obtain the values of ρ and β that satisfy the conditions. Let us recall that equations (6.2) and (6.3) can be written as a system of equations:

$$\begin{bmatrix} (WY^B)^{\mathrm{T}}(WY^B) & (WY^B)^{\mathrm{T}}X^B \\ (X^B)^{\mathrm{T}}(WY^B) & (X^B)^{\mathrm{T}}X^B \end{bmatrix} \begin{bmatrix} \rho \\ \beta \end{bmatrix} = \begin{bmatrix} (WY^B)^{\mathrm{T}}Y^B \\ (X^B)^{\mathrm{T}}Y^B \end{bmatrix}$$

Suppose we let:

A=$\begin{bmatrix} (WY^B)^{\mathrm{T}}(WY^B) & (WY^B)^{\mathrm{T}}X^B \\ (X^B)^{\mathrm{T}}(WY^B) & (X^B)^{\mathrm{T}}X^B \end{bmatrix}$ (the coefficient matrix),

$X = \begin{bmatrix} \rho \\ \beta \end{bmatrix}$ (the unknown coefficients),

$Q = \begin{bmatrix} (WY^{\mathrm{B}})^{\mathrm{T}}Y^B \\ (X^B)^{\mathrm{T}}Y^B \end{bmatrix}$ (the response vector).

Then, the system becomes:

$$AX = Q \tag{6.4}$$

Solution for ρ and β can be obtained using matrix algebra as given:

$$X = A^{-1}Q$$

Explicitly, the estimates for ρ and β are:

Then calculate the inverse of matrix [A] and multiply it by vector [B] to obtain the estimates for ρ and β.

$$\hat{\rho} = \frac{(X^B)^{\mathrm{T}}X^B(WY^B)^{\mathrm{T}}Y^B - (WY^B)^{\mathrm{T}}(WY^B)(X^B)^{\mathrm{T}}Y^B}{(WY^B)^{\mathrm{T}}(WY^B)(X^B)^{\mathrm{T}}X^B - [(WY^B)^{\mathrm{T}}(X^B)^{\mathrm{T}}X^B(WY^B)]} \tag{7}$$

$$\hat{\beta} = \frac{(WY^B)^{\mathrm{T}}(WY^B)(X^B)^{\mathrm{T}}Y^B - (WY^B)^{\mathrm{T}}X^B(WY^B)^{\mathrm{T}}Y^B}{(WY^B)^{\mathrm{T}}(WY^B)(X^B)^{\mathrm{T}}X^B - [(WY^B)^{\mathrm{T}}(X^B)^{\mathrm{T}}X^B(WY^B)]} \tag{8}$$

The equations (7) and (8) provide the estimates for the coefficients ρ and β using the Ordinary Least Squares (OLS) estimation method.

C. Derivation of the Spatial Autocorrelation Parameter (λ) in the Bootstrap-Based Spatial Error Model

To derive the estimate of the spatial autocorrelation parameter λ in equation (9), we will use the structure of the Bootstrap-based Spatial Simultaneous Autoregressive Error Model (BSAEM):

$$Y^B = \lambda WY^B + \beta X^B + \epsilon \tag{9}$$

Rearranging equation (9), we isolate ϵ:

$$\epsilon = Y^B - \lambda WY^B - \beta X^B \tag{9.1}$$

The objective is to estimate λ to minimise the sum of squared residuals (SSE) for the errors. The residual sum of squares is given by:

$$SSE = \epsilon^{\mathrm{T}}\epsilon = \left(Y^B - \lambda WY^B - \beta X^B\right)^{\mathrm{T}}\left(Y^B - \lambda WY^B - \beta X^B\right)$$

To find the estimate of λ, take the partial derivative of SSE concerning λ and set it to zero:

$$\frac{\partial SSE}{\partial \lambda} = 0$$

Differentiate with respect to λ:

$$\frac{\partial SSE}{\partial \lambda} = -2\left(Y^B - \lambda WY^B - \beta X^B\right)^{\mathrm{T}} WY^B$$

Set the derivative to zero:

$$\left(Y^B - \lambda WY^B - \beta X^B\right)^{\mathrm{T}} WY^B = 0$$

Expanding the term:

$$(Y^B)^{\mathrm{T}}WY^B - \lambda(WY^B)^{\mathrm{T}}WY^B - \beta(X^B)^{\mathrm{T}}WY^B = 0 \tag{9.2}$$

Rearranging for λ:

$$\hat{\lambda} = \frac{(Y^B)^{\mathrm{T}}WY^B - \beta(X^B)^{\mathrm{T}}WY^B}{\lambda(WY^B)^{\mathrm{T}}WY^B} \tag{10}$$

Recall that β is already estimated (e.g., using OLS as shown in equations (8)), and substitute its value into the above expression to obtain the final estimate for λ.This estimates the spatial autocorrelation parameter λ in the Bootstrap-based Spatial Simultaneous Autoregressive Error Model (BSAEM).