

Article

Ridge-Type Pretest and Shrinkage Estimation Strategies in Spatial Error Models with an Application to a Real Data Example

Marwan Al-Momani ^{1,*}  and Mohammad Arashi ^{2,3} 

¹ Department of Mathematics, College of Sciences, University of Sharjah, Sharjah P.O. Box 27272, United Arab Emirates

² Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad 9177948974, Iran; arashi@um.ac.ir

³ Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria 0028, South Africa

* Correspondence: malmomani@sharjah.ac.ae

Abstract: Spatial regression models are widely available across several disciplines, such as functional magnetic resonance imaging analysis, econometrics, and house price analysis. In nature, sparsity occurs when a limited number of factors strongly impact overall variation. Sparse covariance structures are common in spatial regression models. The spatial error model is a significant spatial regression model that focuses on the geographical dependence present in the error terms rather than the response variable. This study proposes an effective approach using the pretest and shrinkage ridge estimators for estimating the vector of regression coefficients in the spatial error mode, considering insignificant coefficients and multicollinearity among regressors. The study compares the performance of the proposed estimators with the maximum likelihood estimator and assesses their efficacy using real-world data and bootstrapping techniques for comparison purposes.

Keywords: spatial error model; asymptotic performance; bootstrapping; pretest; ridge estimator; shrinkage

MSC: 62J07; 62H12; 91B72



Citation: Al-Momani, M.; Arashi, M. Ridge-Type Pretest and Shrinkage Estimation Strategies in Spatial Error Models with an Application to a Real Data Example. *Mathematics* **2024**, *12*, 390. <https://doi.org/10.3390/math12030390>

Academic Editor: Sanku Dey

Received: 6 December 2023

Revised: 16 January 2024

Accepted: 23 January 2024

Published: 25 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Data collected over a geographic region may be generally more comparable than data from far away. This phenomenon can be modeled using a covariance structure in conventional statistical models. Spatial regression models, incorporating various spatial dependencies, are increasingly being used in geology, epidemiology, disease monitoring, urban planning, and econometrics.

Time series autoregressive models represent data at time t as linear combinations of the most recent observations. Similarly, in the spatial framework, these models display data from a specific location based on neighboring locations. Data are typically collected from a geographical location known as a site, and proximity is defined by a distance metric.

One of the most used autoregressive models is the spatial error (SE) model, in which a linear regression with a spatially lagged autoregressive error component is used to model the spatial response variable's mean. Ref. [1] investigated the quantile regression estimation for the SE model with potentially variable coefficients. The authors established the proposed estimators' asymptotic properties. Ref. [2] applied the SE model to examine the existence of spatial clustering and correlation between neighboring counties for the data from Egypt's 2006 census. Ref. [3] used the SE model to evaluate the social disorganization theory. Ref. [4] used the combined application of the SE model and spatial lag model based on cross-sectional data from 20 districts in Chengdu. The authors found that the haze had a negative impact on both the selling and rental prices of houses. Ref. [5] proposed

a robust estimation method based on SE models, demonstrating reduced bias, a more stable empirical influence function, and robustness to outliers through simulation. More information about the spatial autoregressive models can be found in [6–8], among others.

In frequentist statistics, sample information is used to establish inferences about unknown parameters, while Bayesian statistics combines sample information with uncertain prior information (UPI) to draw conclusions. Subjective UPI may not always be available. However, model selection procedures like Akaike's information criterion (AIC), Bayesian information criterion (BIC), or model selection techniques can still provide UPI.

An initial trial to estimate the regression parameters using sample information and UPI is called a pretest estimation. The pretest estimator selects significant regression coefficients and chooses between the full model estimator or the revised model estimator, which contains fewer coefficients based on a binary weight. A new modification of the pretest estimator, known as the shrinkage estimator, uses smooth weights between full and sub-model estimators to adjust regression coefficient estimates toward a target value impacted by UPI. Nevertheless, the modified shrinkage estimator suffers sometimes from an over-shrinkage phenomenon. Later, an improved version of this estimator, known as positive shrinkage, controls the over-shrinkage issue.

The concept of using pretest and shrinkage estimating methodologies has received considerable attention from many researchers; for example, ref. [9] introduced an efficient estimation using pretest and shrinkage methods to estimate the regression coefficient vector of the marginal model in the case of multinomial responses. Ref. [10] developed different shrinkage and penalty estimation strategies for the negative binomial regression model when over-fitting and uncertain information exists about the subspace. Ref. [11] proposed shrinkage estimation for the parameter vector of the linear regression model with heteroscedastic errors, and extended their study to the high-dimensional heteroscedastic regression model.

Multicollinearity is a major issue when fitting a multiple linear regression model using the ordinary least squares (OLS) method, which arises when some regressor variables are correlated, especially when the correlation between any two is high. There are several techniques discussed in the literature to reduce the risk of this issue. Ref. [12] introduced the concept of ridge regression as a solution to nonorthogonal problems. They showed that the estimator improves the mean square error of estimation. Ref. [13] introduced a new biased estimator and demonstrated, both theoretically and numerically, the improvements of the new one. Ref. [14] proposed a new version of the Liu estimator for the vector of parameters in a linear regression model based on some prior information.

Using the idea of shrinkage, ref. [15] introduced an improved form of the Liu-type estimator. Analytical and numerical results were used to demonstrate the proposed method's superiority. Ref. [16] suggested the use of the ridge estimator as a suitable approach for handling high-dimensional multicollinearity data. Recently, ref. [17] proposed a novel pretest and shrinkage estimate technique, known as the Liu-type approach, developed for the conditional autoregressive model.

In this article, we propose the ridge-type pretest and shrinkage estimation strategy for the $p \times 1$ regression coefficients vector in the SE model when some prior information is available about the irrelevant coefficients. We will partition the vector β as $(\beta_1, \beta_2)^T$, where β_1 is a $p_1 \times 1$ vector that contains the coefficients of the main effect, and β_2 is a $p_2 \times 1$ vector of irrelevant coefficients, with $p_1 + p_2 = p$. Mainly, we focus on estimating the vector β_1 when the UPI indicates that β_2 is ineffective, which can be achieved by testing a statistical hypothesis of the form $H_0 : \beta_2 = \mathbf{0}$. In some instances, the estimator of the full model may exhibit considerable variability and provide challenges in terms of interpretation. Conversely, the estimator of the sub-model may yield a significantly biased and under-fitted estimate. To tackle this matter, we took into account the pretest, shrinkage, and positive shrinkage ridge estimators for the vector β_1 .

In accordance with our goal, this paper is organized as follows. Section 2 offers an overview of the SE model. A discussion about the maximum likelihood estimators for

the parameters of the SE model is presented in Section 3. In Section 4, we propose the pretest and shrinkage ridge estimators. Asymptotic analysis of the proposed estimators and some theoretical results are presented in Section 5. The estimators are compared numerically using simulated and real data examples in Section 6. Some concluding remarks are presented in Section 7. An appendix containing some proofs is presented at the end of this manuscript.

2. Spatial Error Model

Let $\mathbf{s} = \{s_1, s_2, \dots, s_n\}$ represent a set of (n) spatial sites (frequently known as locations, regions, etc.). Set \mathbf{s} forms what is commonly referred to as a lattice, and the set of nearby sites for s_j , denoted by $K(s_j)$, is defined as follows: $K(s_j) = \{s_i : s_i \text{ is a neighbor of } s_j\}$, $i = 1, 2, \dots, n$. A neighborhood structure can be determined using a predefined adjacency metric. In regular lattices, if two sites just share edges, they are rook-based neighbors; if they also share borders and/or corners, they are queen-based neighbors.

Let $\mathbf{Y}_n(\mathbf{s}) = \{Y(s_1), Y(s_2), \dots, Y(s_n)\}$ be a vector of observations collected at sites $\{s_1, s_2, \dots, s_n\}$, $\mathbf{X}(\mathbf{s}) = (\mathbf{X}(s_1), \mathbf{X}(s_2), \dots, \mathbf{X}(s_n))$ be the $(n \times p)$ matrix of covariates, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ be the $(p \times 1)$ vector of unknown regression parameters, known as the large-scale effect. Following Cressie and Wikle [7], the SE model models the response Y at the j^{th} site s_j as follows:

$$Y(s_j) = \mathbf{X}^T(s_j)\boldsymbol{\beta} + \epsilon(s_j), \quad j = 1, 2, \dots, n, \tag{1}$$

$$\text{with } \epsilon(s_j) = \sum_{i \neq j}^n \lambda_{ji}\epsilon(s_i) + e(s_j), \quad j = 1, 2, \dots, n, \tag{2}$$

where $\mathbf{e}(\mathbf{s}) = (e(s_1), e(s_2), \dots, e(s_n))^T$ is the noise vector that has a Gaussian distribution with a mean of $\mathbf{0}$ and the covariance matrix $\boldsymbol{\Omega} = \text{diag}\{\sigma_j^2, j = 1, 2, \dots, n\}$. Parameters λ_{ji} are used to model the spatial dependencies among the errors $\epsilon_j, j = 1, 2, \dots, n$, with $\lambda_{jj} = 0$. Let $\boldsymbol{\Lambda} = \{\lambda_{j,i}\}_{j,i=1}^n$, and assume that $(\mathbf{I} - \boldsymbol{\Lambda})$ is invertible, where \mathbf{I} is the $(n \times n)$ identity matrix, then by ignoring the spatial indices, the SE in (1) can be rewritten in matrix form, as follows:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \text{with} \quad \boldsymbol{\epsilon} \sim N\left(\mathbf{0}, (\mathbf{I} - \boldsymbol{\Lambda})^{-1}\boldsymbol{\Omega}(\mathbf{I} - \boldsymbol{\Lambda}^T)^{-1}\right) \tag{3}$$

Nature exhibits sparsity in many situations, which means that a small number of factors can account for the majority of the observed variability. In the context of regression analysis, sparsity means a few numbers of the coefficients are significantly different from zero, while the bulk of the coefficients are insignificant and remain zero. Sparsity is frequently used in spatial regression models to imply covariance structures that are easier to compute. Consequently, by setting $\boldsymbol{\Omega} = \sigma^2\mathbf{I}$, and $\boldsymbol{\Lambda} = \rho\mathbf{W}$, where σ^2 is the variance component, ρ is the spatial dependence parameter, and \mathbf{W} is the weight or proximity-known matrix with a main diagonal of zeros, and off-diagonal entices $w_{ji} = 1$ if the location j is a neighbor to location i ; otherwise, $w_{ji} = 0$ for $j \neq i$, the preceding model yields a straightforward and frequently used version. Usually, the weight matrix is normalized as $\mathbf{W}^* = \{\frac{w_{ji}}{w_{j+}}\}_{j,i=1}^n$. So, the SE regression model can be rewritten as follows:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \text{where} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{V}_n) \tag{4}$$

$$\text{and} \quad \mathbf{V}_n = (\mathbf{1} - \rho\mathbf{W}^*)^{-1}(\mathbf{1} - \rho\mathbf{W}^{*T})^{-1} \tag{5}$$

3. Maximum Likelihood Estimation

Let $\theta = (\beta, \sigma^2, \rho)$; the maximum likelihood estimator (MLE) of θ may be acquired by the use of a two-step profile-likelihood method; see [6]. At first, we fix ρ and find the MLEs of β, σ^2 as a function of ρ , which are given as follows:

$$\hat{\beta}(\rho) = \left(X^T V_n^{-1} X \right)^{-1} X^T V_n^{-1} Y \tag{6}$$

$$\hat{\sigma}^2(\rho) = \frac{(Y - X\hat{\beta}(\rho))^T V_n^{-1} (Y - X\hat{\beta}(\rho))}{n} \tag{7}$$

Then, we plug $\hat{\beta}$ and $\hat{\sigma}^2$ into the log-likelihood and obtain the MLE of ρ by maximizing the profile's log-likelihood function. Finally, the MLEs of β and σ^2 are computed by replacing ρ with $\hat{\rho}$ in Equations (6) and (7), respectively. Ref. [18] proved that $\hat{\theta}$ is a consistent estimator of θ , and asymptotically has normal distribution. This finding makes it simple to demonstrate that $\hat{\beta}$ is asymptotically normal and consistent. The significance of regression coefficients can often be determined subjectively or through certain model selection techniques in various situations. As a result of this information, the $(p \times 1)$ regression coefficient vector β is divided into two sub-vectors, as $\beta = (\beta_1, \beta_2)$, where β_1 is a $p_1 \times 1$ vector of important coefficients and β_2 is a $p_2 \times 1$ vector of unimportant coefficients with $p_1 + p_2 = p$. Similarly, the matrix of covariates X is also partitioned as $X = (X_1, X_2)$, where X_1 and X_2 consist of the first p_1 and the last p_2 columns of the design matrix X of dimensions $n \times p_1$ and $n \times p_2$, respectively. Consequently, the SE full model in (3) can be rewritten as follows:

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon \tag{8}$$

For the full model in (8), we can obtain the MLEs of (β_1, β_2) using the same technique employed in model (4); see [19]. The MLEs are as follows:

$$\hat{\beta}_1 = \left(X_1^T A_{X_2} X_1 \right)^{-1} X_1^T A_{X_2} Y, \text{ where} \tag{9}$$

$$A_{X_2} = \hat{V}_n^{-1} - \hat{V}_n^{-1} X_2 \left(X_2^T \hat{V}_n^{-1} X_2 \right)^{-1} X_2^T \hat{V}_n^{-1}$$

and $\hat{\beta}_2$ has an identical formula as $\hat{\beta}_1$ by exchanging indices 1 and 2 in the above two equations. The full model estimation may be prone to significant variability and may be difficult to interpret. Our primary goal is to estimate the value of β_1 when the set of regressors included in the partitioned matrix X_2 does not sufficiently explain the variability in the response variable, which can be achieved by formulating a linear hypothesis as follows:

$$H_0 : \beta_2 = 0 \tag{10}$$

Assuming the null hypothesis in (10) is true, the updated model based on this assumption of the model, given (8), becomes

$$Y = X_1\beta_1 + \epsilon. \tag{11}$$

We will refer to the model in (11) as the restricted SE model. Let $\hat{\beta}_1^S$ be the MLE of β_1 of the model in (11), then

$$\hat{\beta}_1^S = \left(X_1^T \hat{V}_n^{-1} X_1 \right)^{-1} X_1^T \hat{V}_n^{-1} Y. \tag{12}$$

Obviously, $\hat{\beta}_1^S$ will have better performance than $\hat{\beta}_1$ if the null hypothesis in (10) is true, while the opposite occurs when β_2 begins to move away from the null space. Yet,

the restricted strategy method can provide the under-fitted and highly biased model. One goal of this research is to overcome the problem of significant bias in spatial error models when multicollinearity is present among the regressors. To address this issue, we suggest using a ridge-type estimate approach for both the full and sub-models to enhance these estimators by incorporating pretest and shrinkage estimating approaches, to reduce biases and improve the overall accuracy of the estimators. To dominate the large bias, we propose the ridge-type estimation strategy of the full and reduced models and then improve the two estimators using the pretest and shrinkage estimation idea.

4. Materials and Methods: Developing Pretest and Shrinkage Ridge Estimation Strategies

In this section, we propose a set of estimators for the SE model parameter vector β_1 in (11). Following [12], the ridge estimator of β for the model given in (4) is defined as

$$\hat{\beta}^{RF} = (X^T \hat{V}_n^{-1} X + kI_p)^{-1} X^T \hat{V}_n^{-1} Y, \tag{13}$$

where $k > 0$ is known as the ridge parameter. Clearly, when $k = 0$, the ridge estimator reduces to the MLE of β , but if $k \rightarrow \infty$, the ridge estimator $\hat{\beta}^{RF} = 0$.

4.1. Full and Reduced Model Ridge Estimators

The unrestricted full model ridge estimator of β_1 , denoted by $\hat{\beta}_1^{UR}$, is defined as follows:

$$\hat{\beta}_1^{UR} = (X_1^T A_{X_2} X_1 + k_f I_{p_1})^{-1} X_1^T A_{X_2} Y, \tag{14}$$

where k_f represents the ridge parameter for the unretracted full model estimator, denoted as $\hat{\beta}_1^{UR}$. Assuming the null hypothesis in (10) is true, the restricted ridge estimator of β_1 for the model in (11), denoted by $\hat{\beta}_1^{RR}$, is given by

$$\hat{\beta}_1^{RR} = (X_1^T \hat{V}_n^{-1} X_1 + k_r I_{p_1})^{-1} X_1^T \hat{V}_n^{-1} Y, \tag{15}$$

where k_r is the ridge parameter for restricted model estimator $\hat{\beta}_1^{RR}$. When the null hypothesis in (10) is accurate or almost accurate (i.e., when β_2 is close to zero), $\hat{\beta}_1^{RR}$ is generally a more effective estimator than $\hat{\beta}_1^{UR}$. Nevertheless, as β_2 deviates from the zero space, $\hat{\beta}_1^{RR}$ becomes inefficient compared with the unrestricted estimator $\hat{\beta}_1^{UR}$. In addition to the gain obtained by employing the ridge estimation idea to the MLE of β_1 , we also aim to find estimators that are functions of $\hat{\beta}_1^{UR}$ and $\hat{\beta}_1^{RR}$, and intend to lessen the dangers connected with any of these two estimators over the majority of the parameter space. The pretest and shrinkage estimators, which will be built in the following subsection, can help with this.

4.2. Pretest, Shrinkage, and Positive Shrinkage Ridge Estimators

In line with testing the null hypothesis in (10), the pretest estimator selects either the full model estimator $\hat{\beta}_1^{UR}$ if H_0 is rejected or the restricted ridge estimator $\hat{\beta}_1^{RR}$ if not. An appropriate test statistic to test the hypothesis in (10) is

$$T_n = \frac{(\hat{\beta}_2^{UR})^T (X_2^T A_{X_1} X_2) (\hat{\beta}_2^{UR})}{s^2},$$

where A_{X_1} is defined in a similar manner, as A_{X_2} , $\hat{\beta}_2^{UR} = (X_2^T A_{X_1} X_2)^{-1} X_2^T A_{X_1} Y$, and $s^2 = (Y - X \hat{\beta}^{RF})^T (Y - X \hat{\beta}^{RF}) / (n - p)$, which is a consistent estimator of σ^2 , and statistic T_n asymptotically follows a chi-square distribution with p_2 degrees of freedom under the null hypothesis. Hence, the pretest estimator, denoted by $\hat{\beta}_1^{PTR}$, is given by

$$\hat{\beta}_1^{PTR} = \hat{\beta}_1^{UR} - (\hat{\beta}_1^{UR} - \hat{\beta}_1^{RR}) I(T_n \leq \chi_{\alpha, p_2}^2), \tag{16}$$

where $I(\cdot)$ is an indicator function, and χ_{α, p_2}^2 is the upper α th quantile of the chi-square distribution with p_2 degrees of freedom. The pretest estimator depends on the level of the significance (α), and selects $\hat{\beta}_1^{UR}$ if the null hypothesis is rejected, and $\hat{\beta}_1^{RR}$ otherwise, based on binary weights. These drawbacks can be improved using smoother weights of the two estimators $\hat{\beta}_1^{UR}$ and $\hat{\beta}_1^{RR}$ instead, which is known as the shrinkage estimator. It is denoted by, $\hat{\beta}_1^{SR}$, and given by

$$\hat{\beta}_1^{SR} = \hat{\beta}_1^{RR} + (\hat{\beta}_1^{UR} - \hat{\beta}_1^{RR})\{1 - (p_2 - 2)T_n^{-1}\}, \quad p_2 \geq 3. \tag{17}$$

The shrinkage estimator may experience an over-shrinkage in which negative coordinates may be produced whenever $(T_n < p_2 - 2)$. The positive shrinkage estimator, a modified version of $\hat{\beta}_1^{SR}$, resolves this issue. It is denoted by $\hat{\beta}_1^{PSR}$, and given by

$$\hat{\beta}_1^{PSR} = \hat{\beta}_1^{RR} + (\hat{\beta}_1^{UR} - \hat{\beta}_1^{RR})\{1 - (p_2 - 2)T_n^{-1}\}^+, \tag{18}$$

where $x^+ = \max(x, 0)$. It is easy to see that all the pronounced shrinkage estimators satisfy the following general form

$$\hat{\beta}_1^{Shrinkage} = \hat{\beta}_1^{UR} - (\hat{\beta}_1^{UR} - \hat{\beta}_1^{RR})g(T_n). \tag{19}$$

Simply, for $\hat{\beta}_1^{PTR}$, $\hat{\beta}_1^{SR}$, and $\hat{\beta}_1^{PSR}$, the corresponding $g(\cdot)$ functions are given by $I(T_n \leq \chi_{\alpha, p_2}^2)$, $(p_2 - 2)T_n^{-1}$, and $(1 - (p_2 - 2)T_n^{-1})I(T_n \leq \chi_{\alpha, p_2}^2)$, respectively.

5. Asymptotic Analysis

In this section, we will study the asymptotic performances of all estimators based on their asymptotic quadratic risks. Our goal is to investigate the behavior of the set of estimators near the null space, so we consider a sequence of local alternatives given by

$$H_{(n)} : \beta_{2(n)} = \frac{\xi}{\sqrt{n}}, \quad \xi \in \mathfrak{R}^{p_2}, \text{ with } \xi \neq \mathbf{0} \tag{20}$$

Obviously, when $\xi = \mathbf{0}$, the local alternatives in (20) may be simplified to the null hypothesis given in (10). Assuming that $K(x)$ represents the cumulative distribution function of any estimator of β_1 , say $\hat{\beta}_1^*$, then: $K(x) = \lim_{n \rightarrow \infty} P_{H_{(n)}}(\sqrt{n}(\hat{\beta}_1^* - \beta_1))$. Thus, for any $(p_1 \times p_1)$ positive definite matrix M , the weighted quadratic loss function is defined as

$$\begin{aligned} W(\hat{\beta}_1^*, \beta_1) &= n(\hat{\beta}_1^* - \beta_1)^T M(\hat{\beta}_1^* - \beta_1) \\ &= \text{tr} \left[M \left[n(\hat{\beta}_1^* - \beta_1)(\hat{\beta}_1^* - \beta_1)^T \right] \right], \end{aligned}$$

where $\text{tr}(A)$ is the trace of the matrix A . Define $\vartheta_n^* = \sqrt{n}(\hat{\beta}_1^* - \beta_1)$, then if $\vartheta_n^* \xrightarrow{D} \vartheta^*$, where \xrightarrow{D} denotes the convergence in distribution, then the asymptotic (distributional) quadratic risk (ADQR) of $\hat{\beta}_1^*$, denoted by $\Gamma(\hat{\beta}_1^*)$, is given by

$$\Gamma(\hat{\beta}_1^*, M) = E(\vartheta_n^{*T} M \vartheta_n^*) = \int (x_1^T M x_1) dK(x_1) \tag{21}$$

The asymptotic (distributional) bias (ADB) of $\hat{\beta}_1^*$ can be obtained via

$$\text{ADB}(\hat{\beta}_1^*) = E\left(\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\beta}_1^* - \beta_1)\right). \tag{22}$$

To derive asymptotic distributional properties, in addition to the first four assumptions of [18], we set the following regularity conditions:

- (A1) $\max_{1 \leq i \leq n} \frac{1}{n} \mathbf{x}_i^T (\mathbf{X}^T \hat{\mathbf{V}}_n^{-1} \mathbf{X})^{-1} \mathbf{x}_i \rightarrow 0$, as $n \rightarrow \infty$, where \mathbf{x}_i is the i th row of \mathbf{X} .
- (A2) Let $\mathbf{C}_n = \mathbf{X}^T \hat{\mathbf{V}}_n^{-1} \mathbf{X}$. Then, $\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{C}_n = \mathbf{C}$, where \mathbf{C} is the $(p \times p)$ positive definite matrix.
- (A3) Let

$$\begin{aligned} \mathbf{C}_n^{-1} &= \begin{pmatrix} \mathbf{X}_1^T \hat{\mathbf{V}}_n^{-1} \mathbf{X}_1 & \mathbf{X}_1^T \hat{\mathbf{V}}_n^{-1} \mathbf{X}_2 \\ \mathbf{X}_2^T \hat{\mathbf{V}}_n^{-1} \mathbf{X}_1 & \mathbf{X}_2^T \hat{\mathbf{V}}_n^{-1} \mathbf{X}_2 \end{pmatrix}^{-1} \text{ and} \\ \mathbf{C}^{-1} &= \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{C}_{11.2}^{-1} & -\mathbf{C}_{11}^{-1} \mathbf{C}_{12} \mathbf{C}_{22.1}^{-1} \\ -\mathbf{C}_{22}^{-1} \mathbf{C}_{21} \mathbf{C}_{11.2}^{-1} & \mathbf{C}_{22.1}^{-1} \end{pmatrix}, \end{aligned}$$

Then, $\lim_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{C}_n\right)^{-1} = \mathbf{C}^{-1}$, where $\mathbf{C}_{ii,j} = \mathbf{C}_{ii} - \mathbf{C}_{ij} \mathbf{C}_{jj}^{-1} \mathbf{C}_{ji}$ for $i, j = 1, 2$.

In the following, we refer to the above assumptions as the “named regularity condition (NRC)”.

The primary tool to derive expressions of the asymptotic quadratic risks for the proposed estimators is to find the asymptotic distribution of the unrestricted full model ridge estimator $\hat{\beta}_1^{\text{UR}}$ and the restricted ridge estimator $\hat{\beta}_1^{\text{RR}}$. To this end, we make use of the following lemma. The proof is provided in Appendix A.

Lemma 1. Assume the NRC. If $k/\sqrt{n} \rightarrow k_0 \geq 0$, then

$$\sqrt{n}(\hat{\beta}^{\text{RF}} - \beta) \xrightarrow{D} N_p(-k_0 \mathbf{C}^{-1} \beta, \sigma^2 \mathbf{C}^{-1}),$$

where \xrightarrow{D} denotes convergence in distribution. Indeed, Lemma 1 enables us to provide some asymptotic distributional results about the estimators $\hat{\beta}_1^{\text{UR}}$ and $\hat{\beta}_1^{\text{RR}}$, as presented in the following theorem, which are easy to prove.

Theorem 1. Let $\vartheta_n^{(1)} = \sqrt{n}(\hat{\beta}_1^{\text{UR}} - \beta_1)$, $\vartheta_n^{(2)} = \sqrt{n}(\hat{\beta}_1^{\text{RR}} - \beta_1)$, and $\vartheta_n^{(3)} = \sqrt{n}(\hat{\beta}_1^{\text{UR}} - \hat{\beta}_1^{\text{RR}})$. Assume the local alternatives in (20) and NRC. Then, as $n \rightarrow \infty$, we have

1. $\vartheta_n^{(1)} \sim N_{p_1}(-\eta_{11.2}, \sigma^2 \mathbf{C}_{11.2}^{-1})$
2. $\vartheta_n^{(2)} \sim N_{p_1}(\delta - \eta_{11.2}, \sigma^2 \mathbf{C}_{11}^{-1})$
3. $\vartheta_n^{(3)} \sim N_{p_1}(\delta, \sigma^2(\mathbf{C}_{11.2}^{-1} - \mathbf{C}_{11}^{-1}))$
4. $\begin{pmatrix} \vartheta_n^{(1)} \\ \vartheta_n^{(3)} \end{pmatrix} \sim N_{2p_1} \left(\begin{pmatrix} -\eta_{11.2} \\ \delta \end{pmatrix}, \sigma^2 \begin{pmatrix} \mathbf{C}_{11.2}^{-1} & \mathbf{C}_{11.2}^{-1} - \mathbf{C}_{11}^{-1} \\ \mathbf{C}_{11.2}^{-1} - \mathbf{C}_{11}^{-1} & \mathbf{C}_{11.2}^{-1} - \mathbf{C}_{11}^{-1} \end{pmatrix} \right)$
5. $\begin{pmatrix} \vartheta_n^{(2)} \\ \vartheta_n^{(3)} \end{pmatrix} \sim N_{2p_1} \left(\begin{pmatrix} \delta - \eta_{11.2} \\ \delta \end{pmatrix}, \sigma^2 \begin{pmatrix} \mathbf{C}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{11.2}^{-1} - \mathbf{C}_{11}^{-1} \end{pmatrix} \right)$
6. $E[\vartheta_n^{(1)} | \vartheta_n^{(3)}] = -\eta_{11.2} + \vartheta_n^{(3)} - \delta$
7. $Pr(T_n \leq x) = H_q(x; \Delta)$, where $H_q(x; \Delta)$ is the cumulative distribution function of a non-central chi-square distribution with q degrees of freedom and the Δ non-centrality parameter. Where $\eta = (\eta_1^T, \eta_2^T)^T = -k_0 \mathbf{C}^{-1} \beta$, $\eta_{11.2} = \eta_1 - \mathbf{C}_{12} \mathbf{C}_{22}^{-1} ((\beta_2 - \xi) - \eta_2)$, $\delta = \mathbf{C}_{11}^{-1} \mathbf{C}_{12} \xi$.

With Lemma 1 in hand, it is pretty straightforward to reach the asymptotic distributional properties of the shrinkage estimators. Through the subsequent theorems, we will provide the asymptotic bias and weighted quadratic risk functions.

Theorem 2. Under the assumptions of Lemma 1, the asymptotic distributional bias of the shrinkage estimators are given by

1. $\text{ADB}(\hat{\beta}_1^{\text{PTR}}) = -\eta_{11.2} - \delta H_{p_2+2}(\chi_{\alpha, p_2}^2; \Delta)$

2. $ADB(\hat{\beta}_1^{SR}) = -\eta_{11.2} - (p_2 - 2)\delta E(\chi_{p_2+2}^{-2}(\Delta))$
3. $ADB(\hat{\beta}_1^{PSR}) = -\eta_{11.2} - \delta H_{p_2+2}(\chi_{\alpha, p_2}^2; \Delta) + (p_2 - 2)\delta E[\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2)]$,
where

$$E(\chi_q^{-2i}(\Delta)) = \int_{x=0}^{x=\infty} x^{-2i} dH_q(x; \Delta), \quad i = 1, 2.$$

For the proof, refer to Appendix A.

The following results reveal the expressions for the ADQR of the proposed shrinkage estimators.

Theorem 3. Under the assumptions of Lemma 1, the asymptotic distributional quadratic risks of the shrinkage estimators are given by

1.

$$\begin{aligned} \Gamma(\hat{\beta}_1^{PTR}, \mathbf{M}) &= \Gamma(\hat{\beta}_1^{UR}, \mathbf{M}) - 2\eta_{11.2}^T \mathbf{M} \delta H_{p_2+2}(\chi_{\alpha, p_2}^2; \Delta) \\ &\quad - \sigma^2 \text{tr}[\mathbf{M}(\mathbf{C}_{11.2}^{-1} - \mathbf{C}_{11}^{-1})] H_{p_2+2}(\chi_{\alpha, p_2}^2; \Delta) \\ &\quad \delta^T \mathbf{M} \delta [2H_{p_2+2}(\chi_{\alpha, p_2}^2; \Delta) - H_{p_2+4}(\chi_{\alpha, p_2}^2; \Delta)] \end{aligned}$$

2.

$$\begin{aligned} \Gamma(\hat{\beta}_1^{SR}, \mathbf{M}) &= \Gamma(\hat{\beta}_1^{UR}, \mathbf{M}) + 2(p_2 - 2)\eta_{11.2}^T \mathbf{M} \delta E(\chi_{p_2+2}^{-2}(\Delta)) \\ &\quad - (p_2 - 2)\sigma^2 \text{tr}(\mathbf{M} \mathbf{C}_{11}^{-1} \mathbf{C}_{12} \mathbf{C}_{22.1}^{-1} \mathbf{C}_{21} \mathbf{C}_{11}^{-1}) \\ &\quad \{2E(\chi_{p_2+2}^{-2}(\Delta)) - (p_2 - 2)E(\chi_{p_2+2}^{-4}(\Delta))\} \\ &\quad + (p_2 - 2)\delta^T \mathbf{M} \delta \\ &\quad \times \{2E(\chi_{p_2+2}^{-2}(\Delta)) - 2E(\chi_{p_2+4}^{-2}(\Delta)) - (p_2 - 2)E(\chi_{p_2+4}^{-4}(\Delta))\}. \end{aligned}$$

3.

$$\begin{aligned} \Gamma(\hat{\beta}_1^{PSR}, \mathbf{M}) &= \Gamma(\hat{\beta}_1^{SR}, \mathbf{M}) - 2\eta_{11.2}^T \mathbf{M} \delta A_1 \\ &\quad + (p_2 - 2)\sigma^2 \text{tr}(\mathbf{M} \mathbf{C}_{11}^{-1} \mathbf{C}_{12} \mathbf{C}_{22.1}^{-1} \mathbf{C}_{21} \mathbf{C}_{11}^{-1}) A_2 \\ &\quad - \sigma^2 \text{tr}(\mathbf{M} \mathbf{C}_{11}^{-1} \mathbf{C}_{12} \mathbf{C}_{22.1}^{-1} \mathbf{C}_{21} \mathbf{C}_{11}^{-1}) H_{p_2+2}(\chi_{\alpha, p_2}^2; \Delta) \\ &\quad + \delta^T \mathbf{M} \delta [2H_{p_2+2}(\chi_{\alpha, p_2}^2; \Delta) - H_{p_2+4}(\chi_{\alpha, p_2}^2; \Delta)] \\ &\quad - (p_2 - 2)\delta^T \mathbf{M} \delta A_3, \end{aligned}$$

where \mathbf{M} is a positive definite weight matrix,

$$\Gamma(\hat{\beta}_1^{UR}, \mathbf{M}) = \eta_{11.2}^T \mathbf{M} \eta_{11.2} + \sigma^2 \text{tr}(\mathbf{M} \mathbf{C}_{11.2}^{-1}), \text{ and}$$

$$\begin{aligned} A_1 &= E\{1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta)\} I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2), \\ A_2 &= 2E\{\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2)\} \\ &\quad - (p_2 - 2)E\{\chi_{p_2+2}^{-4}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2)\}, \\ A_3 &= 2E\{\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2)\} \\ &\quad - 2E\{\chi_{p_2+4}^{-2}(\Delta)I(\chi_{p_2+4}^2(\Delta) \leq p_2 - 2)\} \\ &\quad + (p_2 - 2)E\{\chi_{p_2+2}^{-4}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2)\}. \end{aligned}$$

For the proof, refer to Appendix A.

6. Numerical Analysis

To demonstrate our theoretical findings, we first use Monte Carlo simulation experiments and then apply the set of proposed estimators to a real dataset. The Monte Carlo simulation is used to investigate the performance of the ridge-type set of estimators in comparison to the MLE ($\hat{\beta}_1$) given in (9) via the simulated mean square error of each estimator.

6.1. Simulation Experiments

In this section, we compare the set of ridge-type estimators with respect to the MLE using Monte Carlo simulation experiments based on their simulated mean squared errors. In each of these experiments, we consider $(N \times N)$ square lattices using $N = 7, 10$ with the corresponding sample sizes of $n = N^2 = 49, 100$, respectively. To show the performance of the proposed estimators when multicollinearity exists, we generate the design matrix X from the multivariate normal distribution with a mean of $\mathbf{0}$, and a variance-covariance matrix with a first-order autoregressive structure, in which $cov(X_i, X_j) = \begin{cases} \rho_x^{|i-j|} & i \neq j \\ 1 & i = j \end{cases}$ and apply it to $\rho_x \in \{0.3, 0.6, 0.9\}$, while the error term ϵ is generated from another multivariate normal with a mean of $\mathbf{0}$ and an SE variance matrix with $V_n = \sigma^2(I - \rho W^*)^{-1}(I - \rho W^{*'})^{-1}$. We set $\sigma^2 = 1$. For the weight matrix W^* , a queen-based contiguity neighborhood was used. The set of values for ρ is $\{0.3, 0.6, 0.9\}$. We partitioned the vector of coefficients β as $\beta = (\beta_1, \beta_2)$, where $\beta_1 = \mathbf{1}_{p_1}$ is a $p_1 \times 1$ vector of ones, and $\beta_2 = (\Delta, \mathbf{0}_{p_2-1})$, $\mathbf{0}_{p_2-1}$ is a zero vector of dimension $(p_2 - 1 \times 1)$, and $\Delta = \|\beta - \beta_0\|$, where $\|A\|$ is the Euclidean norm of A , and Δ represents the non-centrality parameter. The range of values for Δ is set to be from 0 to 2. Then we fit the model in (8) using the `spautolm` function within the R-package `spdep` [20], obtain the values of all estimators considered in our study, and compute the simulated mean square error (SMSE) of each estimator as $SMSE(\hat{\beta}_1^*) = \sum_{i=1}^{p_1} (\hat{\beta}_{1i}^* - \beta_{1i})^2$. The simulated relative efficiency (SRE) of any estimator, say $\hat{\beta}_1^\circ$, with respect to the MLE ($\hat{\beta}_1$), is calculated as follows:

$$SRE(\hat{\beta}_1^\circ) = \frac{SMSE(\hat{\beta}_1)}{SMSE(\hat{\beta}_1^\circ)}, \tag{23}$$

where $\hat{\beta}_1^\circ$ is any of the estimators $\{\hat{\beta}_1^{UR}, \hat{\beta}_1^{RR}, \hat{\beta}_1^{PTR}, \hat{\beta}_1^{SR}, \hat{\beta}_1^{PSR}\}$. It is evident that when the $SRE(\hat{\beta}_1^\circ)$ is greater than one, it signifies that this estimate outperforms the MLE of the full model, and vice versa. We run the simulation for $(p_1, p_2) \in \{(5, 10), (5, 20), (5, 30)\}$, and use $\alpha = 0.05$ to test the hypothesis in (20). No statistically significant change is seen while altering the spatial dependency parameter. Therefore, we simply exhibit the graphs for $\rho = 0.90$. Figures A1–A3 in Appendix A show the results of the SRE against various values of Δ . The findings support the following conclusions:

- (i) Across all values, the ridge-type full model estimator ($\hat{\beta}_1^{UR}$) consistently outperforms the traditional MLE estimator. Furthermore, as p_2 increases, so does its efficiency for fixed values of ρ and ρ_x . Additionally, when the multicollinearity among the explanatory variables in the design matrix becomes stronger, $\hat{\beta}_1^{UR}$ efficiency increases as expected.
- (ii) The ridge-type sub-model estimator ($\hat{\beta}_1^{RR}$) outperforms all other estimators when $\Delta = 0$. Since the null hypothesis is correct, it is expected. However, once Δ begins to depart from the null space, the estimator’s SRE drops precipitously and approaches zero, making it less effective than the other estimators.
- (iii) The SRE values grow while holding other parameters constant, as the correlation coefficient ρ_x increases among the explanatory factors.
- (iv) As the number of zero coefficients (p_2) increase =, all SRE estimators also increase.
- (v) The ridge-type positive shrinkage estimator ($\hat{\beta}_1^{PSR}$) uniformly prevails over the competing estimators.

6.2. Data Example

In 1970, ref. [21] examined the use of housing market data for census tracts in the Boston Statistical Metropolitan Area. The authors' major objective was to establish a relationship between a set of (15) variables and the median cost of owner-occupied residences in Boston. Ref. [22] offered a corrected version of the dataset along with new spatial data. The dataset is accessible through the R-Package `spdep` version 1.3-1. There are 506 observations in the data, each of which relates to a single census tract. The variables in the data include the tract identification number (TRACT), median owner-occupied housing prices in US dollars (MEDV), corrected median owner-occupied housing prices in US dollars (CMEDV), percentages of residential land zoned for lots larger than 2500 square feet per town (constant for all Boston tracts) (ZN), percentages of non-retail business areas per town (INDUS), average room sizes per home (RM), the percentage of owner-occupied homes built before 1940 (AGE), a dummy variable with two levels, which is 1 if the tract borders the Charles River and 0 otherwise (CHAS), the crime rate per capita (CRIM), the weighted distance to main employment centers (DIS), nitrogen oxide concentration (parts per 10 million) per town (NOX), an accessibility index to radial highway per town (constant for all Boston tracts) (RAD), property tax rate per town (\$10,000) (constant for all Boston tracts) (TAX), percentage of the lower-class population (LSTAT), pupil-teacher ratios per town (constant for all Boston tracts) (PTRATIO), and the variable $1000(b - 0.63)^2$, where b is the proportion of blacks (B). Ref. [23] added the location of each tract in latitude (LAT), and longitude (LON) variables. Assuming an SE model, we can predict the response variable $\log(\text{CMEDV})$ using all available variables, which will be referred to as the full SE model. For these data, a variety of selection techniques were used to determine the sub-model. One sub-model that was used by [19] is the model obtained by the adaptive LASSO algorithm, which will be referred to as our SE sub-model. The two models are summarized in Table 1.

Table 1. Full and sub-model.

Selection Criterion	Model
Full	$\log(\text{CMEDV}) = \log(\text{LSTAT}) + \text{I}(\text{RM}^2) + \text{TAX} + \text{B} + \log(\text{RAD}) + \text{CHAS} + \text{CRIM} + \text{PTRATIO} + \text{AGE} + \text{LAT} + \text{LON} + \log(\text{RAD}) + \text{I}(\text{NOX}^2) + \log(\text{DIS}) + \text{ZN} + \text{INDUS}$
Sub-model	$\log(\text{CMEDV}) = \log(\text{LSTAT}) + \text{I}(\text{RM}^2) + \text{TAX} + \text{B} + \text{CRIM} + \text{PTRATIO}$

Figure 1 displays a colored plot of the correlation coefficients for each variable. The notation (***) indicates high significance with a p -value of less than 0.001. The notation (**) indicates significance at a 1% level, while (*) indicates significance at a 5% level. If none of these symbols are present, it signifies that the correlation coefficient between the two variables is not statistically significant. The CMEDV and a few other factors have a strong linear relationship, as seen in the plot. This plot is useful for examining the strength of linearity between the original response CMEDV and any other variable if it exists. The selected variables by the adaptive LASSO algorithm appear to have strong, medium, and weak relationships with the response variable. Moreover, some variables exhibit collinearity; this issue will show how ridge-type estimators show high performance compared to MLE estimators.

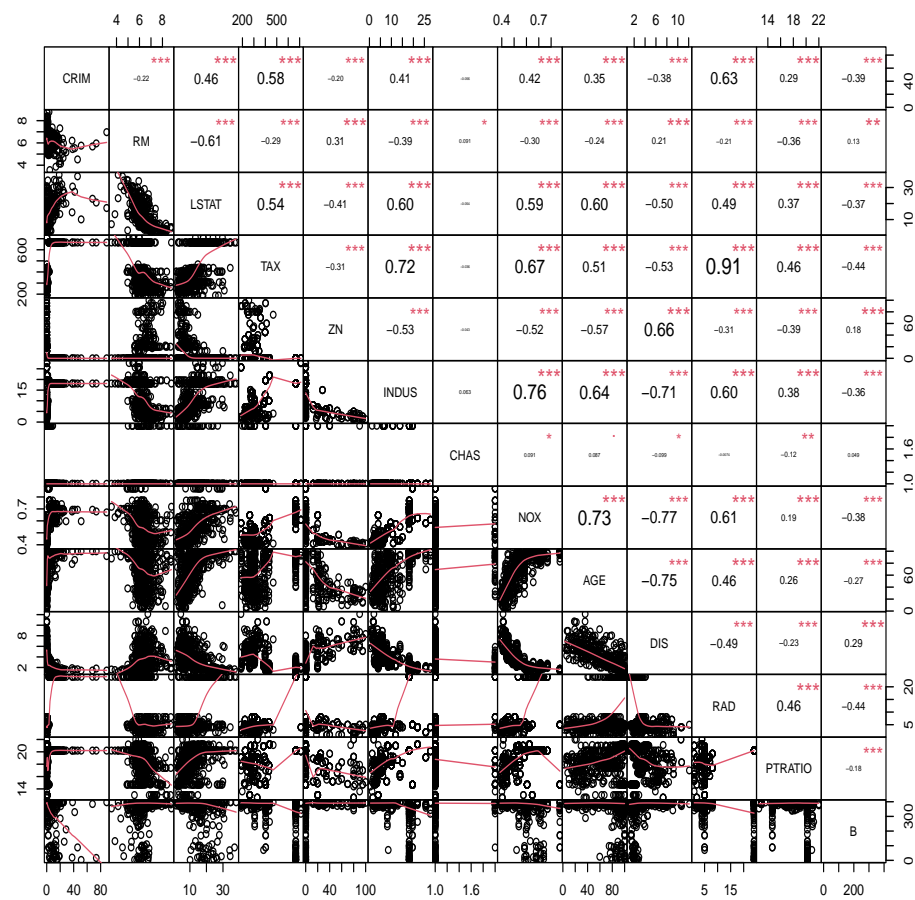


Figure 1. Correlation matrix for the Boston housing data.

Table 2 presents the estimated ridge-type values of the proposed estimators.

Table 2. Estimated values.

Coefficient	$\hat{\beta}_1^{UR}$	$\hat{\beta}_1^{RR}$	$\hat{\beta}_1^{PTR}$	$\hat{\beta}_1^{SR}$	$\hat{\beta}_1^{PSR}$
Intercept	3.8310	3.6490	3.8310	3.7882	3.7882
log(LSTAT)	-0.2635	-0.2872	-0.2635	-0.2691	-0.2691
I(RM ²)	0.0081	0.0077	0.0081	0.0080	0.0080
TAX	-0.0005	-0.0003	-0.0005	-0.0005	-0.0005
B	0.0006	0.0005	0.0006	0.0006	0.0006
CRIM	-0.0053	-0.0047	-0.0053	-0.0051	-0.0051
PTRATIO	-0.0168	-0.0157	-0.0168	-0.0165	-0.0165

To assess the effectiveness of the suggested estimators, we will use two different methods, aiming to provide a reliable and valid evaluation of our results. The first method employs the bootstrapping methodology, whereas the second one involves validation using out-of-sample data.

The bootstrapping technique suggested by [24], computes the mean squared prediction error (MSPE) for any estimator as follows:

1. Fit the SE full and sub-models as they appear in Table 1 using the `spauto1m` function and obtain the MLEs of β_1 , σ^2 , the spatial dependence parameter ρ , and the covariance matrix V_n .
2. As the columns of matrices $(X_1^T A_{X_2} X_1)^{-1}$ and $(X_1^T \hat{V}_n^{-1} X_1)^{-1}$ are not orthogonal, and the sample size is large, we follow [25] to estimate the tuning ridge parameters for the

two estimators, $\hat{\beta}_1^{UR}$ and $\hat{\beta}_1^{RR}$, which are, respectively, given by

$$k_f = \frac{\hat{\sigma}^2 tr(X_1^T A_{X_2} X_1)^{-1}}{(\hat{\beta}_1)^T (X_1^T A_{X_2} X_1)^{-1} \hat{\beta}_1}, \text{ and } k_r = \frac{\hat{\sigma}^2 tr(X_1^T \hat{V}_n^{-1} X_1)^{-1}}{(\hat{\beta}_1^S)^T (X_1^T \hat{V}_n^{-1} X_1)^{-1} \hat{\beta}_1^S}.$$

3. Use the Cholesky decomposition method to express the matrix \hat{V}_n in a decomposed form as $\hat{V}_n = \mathbf{U}\mathbf{U}^T$, where \mathbf{U} is an $(n \times n)$ lower triangular matrix.
4. Let $\hat{\epsilon} = \mathbf{U}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})$, where $\hat{\epsilon} = (\hat{\epsilon}_1, \hat{\epsilon}_2, \dots, \hat{\epsilon}_n)$; we define the centered residual as $\epsilon_i^c = \hat{\epsilon}_i - \frac{1}{n} \sum_{j=1}^n \hat{\epsilon}_j$, and then select with the replacement a sample of size (n) $(\epsilon_1^c, \epsilon_2^c, \dots, \epsilon_n^c)$ to obtain $\epsilon^* = (\epsilon_1^*, \epsilon_2^*, \dots, \epsilon_n^*)$.
5. Calculate the bootstrapping response value as $\mathbf{Y}^* = \mathbf{X}\hat{\beta} + \mathbf{U}^{-1}\epsilon^*$, and then use it to fit the full and sub-models and obtain bootstrapping estimated values of all estimators.
6. Calculate the predicted value of the response variable using each estimator as follows:

$$y_{ki}^{\hat{*}} = X_1 \hat{\beta}_1^* + \hat{\rho}^* \sum_{j=1}^n W_{ij}^* (y_{kj}^{\hat{*}} - X_j \hat{\beta}_1^*),$$
 where $\hat{\beta}_1^*$ represents any of the estimators in the set $\{\hat{\beta}_1, \hat{\beta}_1^{UR}, \hat{\beta}_1^{RR}, \hat{\beta}_1^{PTR}, \hat{\beta}_1^{SR}, \hat{\beta}_1^{PSR}\}$.
7. For the k^{th} bootstrapping sample, we calculate the square root of the mean square prediction error (MSPE) as

$$MSPE_k(\hat{\beta}_1^*) = \sqrt{\frac{\sum_{i=1}^n (y_i^{\hat{*}} - y_i)^2}{n}}, k = 1, 2, \dots, B, \tag{24}$$

where B is the number of bootstrapping samples.

8. Calculate the relative efficiency (RE) of any estimator with respect to the MLE $\hat{\beta}_1$ as follows:

$$RE(\hat{\beta}_1^{\bullet}) = \frac{MSPE(\hat{\beta}_1)}{MSPE(\hat{\beta}_1^{\bullet})}, \tag{25}$$

where $\hat{\beta}_1^{\bullet}$ is any of the ridge-type proposed estimators. We apply the bootstrapping technique $B = 2000$ times.

Table 3 summarizes the results of the relative efficiencies, where a relative efficiency value exceeding one indicates the superior performance of the estimator in the denominator.

Table 3. RE of the proposed estimators.

Estimator	$\hat{\beta}_1^{UR}$	$\hat{\beta}_1^{RR}$	$\hat{\beta}_1^{PTR}$	$\hat{\beta}_1^{SR}$	$\hat{\beta}_1^{PSR}$
RE	1.0198	2.9468	2.8624	2.3070	2.3287

The second approach is based on out-of-sample data. In general, when using out-of-sample data for non-spatial regression models, it is assumed that the errors are independent. However, the errors in the SE regression model are not independent. Nonetheless, by employing a transformation, we may overcome this challenge. We suggest modifying the SE model to ensure the errors are independent while keeping σ^2 and ρ constant. A related transformation technique in spatial models can be found in [6]. Note that the $(n \times n)$ covariance matrix V_n is positive definite, so V_n can be rewritten as $V_n = A^T A$, where A is an upper triangular matrix with positive entries on the diagonal; see ([26], p. 338). By multiplying the model in (4) by $(A^{-1})^T$, we obtain

$$\mathbf{Y}^* = \mathbf{X}^* \boldsymbol{\beta} + \epsilon^*, \tag{26}$$

where $\mathbf{Y}^* = (A^{-1})^T \mathbf{Y}$, $\mathbf{X}^* = (A^{-1})^T \mathbf{X}$, and $\epsilon^* = (A^{-1})^T \epsilon$, with $\epsilon^* \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. Practically, we obtain the MLEs of σ^2 , and ρ , and use these estimates to find the estimated matrix \hat{A} . The steps of using out-of-sample data are as follows:

1. Create a data frame containing the columns of \mathbf{X}^* and \mathbf{Y}^* .

2. Divide the data frame into training and testing subsets. The testing data subset is known as the out-of-sample dataset.
3. Using the training dataset, we follow the same procedure discussed in Section 3. That is, we divide the training data into two subsets, as $X_{train}^* = (X_{1,train}^*, X_{2,train}^*)$, fit the full and sub-models, and obtain the array of estimators, which are denoted by $\hat{\beta}_{1,train}^*, \hat{\beta}_{1,train}^{UR}, \hat{\beta}_{1,train}^{RR}, \hat{\beta}_{1,train}^{PTR}, \hat{\beta}_{1,train}^{SR}, \hat{\beta}_{1,train}^{PSR}$.
4. Divide the testing data into two subsets, as $X_{test}^* = (X_{1,test}^*, X_{2,test}^*)$, and calculate the predicted response values as follows:

$$\hat{Y}^* = X_{1,test}^* \hat{\beta}_{1,train}^*$$

where $\hat{\beta}_{1,train}^*$ is any of the estimators obtained in step (2).

5. Compute the average of the MSPE using Equation (24), replacing \hat{y}_i^* with \hat{y}_i^* , and y_i by $y_{i,test}^*$.

We divide the dataset into 80% for the training set and 20% for the testing set, respectively; we repeat steps (2–5) 2000 times, and then obtain the relative efficiency as in Equation (25). Table 4 shows the relative efficiency results based on out-of-sample data.

Table 4. The RE of the proposed estimators using out-of-sample data.

Estimator	$\hat{\beta}_1^{UR}$	$\hat{\beta}_1^{RR}$	$\hat{\beta}_1^{PTR}$	$\hat{\beta}_1^{SR}$	$\hat{\beta}_1^{PSR}$
RE	1.0447	5.1399	1.2816	1.5649	1.5649

Tables 3 and 4 illustrate the better performance of the sub-model ridge-type estimator ($\hat{\beta}_1^{RR}$) compared to all other estimators. It is then followed by the pretest estimator ($\hat{\beta}_1^{PTR}$), demonstrating the correctness of the sub-model that was selected. Also, the ridge-positive shrinkage estimator performs better than the shrinkage one. Furthermore, all ridge-type estimators outperformed the MLE of β_1 .

7. Conclusions

This paper discusses the pretest, shrinkage, and positive shrinkage ridge-type estimators of the parameter vector (β) for the SE model when there is a previous suspicion that certain coefficients are insignificant, and multicollinearity exists between two or more regressor variables. To obtain the proposed set of estimators for the main effect vector of coefficients (β_1), we test the hypothesis $H_0 : \beta_2 = 0$. The proposed estimators were compared analytically via their asymptotic distributional quadratic risks, and numerically through simulation experiments and a real data example.

Our results showed that there is no significant effect of the spatial dependence parameters (ρ), while the performance of the ridge estimators increases as the correlation among the regressor variables increases. Moreover, the performance of the ridge estimators is always better than the MLE. In addition, the estimator ($\hat{\beta}_1^{RR}$) dominates all estimators under the null hypothesis $H_0 : \beta_2 = 0$, or when near the null space, and delivers higher efficiency than the other estimators. However, the proposed positive shrinkage ridge estimators ($\hat{\beta}_1^{PSR}$) perform better than the MLE in all seniors. Further, we apply the set of estimators to a real data example, and use bootstrapping and validation based on out-of-sample data techniques to evaluate their performance based on the relative efficacy of the square root of the mean squared prediction error.

The ridge-type pretest and shrinkage estimators significantly reduced the MSPE. These estimators handled spatial error models well by collecting and minimizing prediction variation. The lower MSPE shows that the proposed estimation strategy makes more accurate and trustworthy forecasts than MLE. This suggests that adding these estimators to the model improves predicted accuracy and model performance. This part of the residual analysis shows that ridge-type estimators are effective at addressing prediction errors in spatial error modeling.

The idea of a ridge-type pretest and shrinkage estimation strategy applies to a wide range of spatial regression models. Regarding continuous data, the method can be used for different models, such as the conditional autoregressive model, simultaneous autoregressive model, and spatial autoregressive moving average model. For discrete spatial data types, it can be applied to generalized linear models with a conditional autoregressive covariance structure, auto-logistic models for binary data, and auto-Poisson and negative binomial models for count data, among others. Lee, L. F [27] considered the estimation of the spatial model, which includes a spatial lag of the dependent variable and spatially autoregressive disturbances and provides the most effective spatial two-stage least squares (2SLS) estimator, which are instrumental variable estimators and optimal in the asymptotic sense. Such a method may be used and benefit from the pretest and shrinkage estimation strategy to improve the estimators of several spatial regression models. Liu and Yang [28] discussed the impact of spatial dependence on the convergence rate of quasi-maximum likelihood (QML) estimators and guided how to rectify the finite sample bias in the spatial error dependence model. Based on our findings, it is expected that employing the ridge-type pretest and shrinkage estimation for the spatial dependence model will be a beneficial addition, and provide better results in terms of the estimators’ biases.

Author Contributions: M.A.-M. initiated the research, designed the study, proposed estimators, established a methodology, did the numerical studies, including the simulation and the data example, analyzed findings, wrote a manuscript, underwent critical revision, and revised the final version. M.A. meticulously stages the research from design to final approval, including methodology, especially the theory, and the writing of the original manuscript, reviewing and editing, ensuring accuracy, clarity, and coherence of findings through an iterative process. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset is accessible through the R-Package “spdep”.

Acknowledgments: We express our heartfelt gratitude to the four anonymous reviewers for their valuable feedback, which prompted us to include several details in the work and enhance its presentation.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proofs of the Main Results

Proof of Lemma 1. For the proof, we follow the approach of Yuzbasi et al. [29], with a slight modification. Let $W \sim N_p(0, \sigma^2 C)$ and define

$$\begin{aligned}
 V_n(\mathbf{u}) &= \sum_{i=1}^n [(\epsilon_i - \mathbf{u}^T \mathbf{x}_i / \sqrt{n})^2 - \epsilon_i^2] + k \sum_{j=1}^p [(\beta_j + u_j / \sqrt{n})^2 - \beta_j^2] \\
 V(\mathbf{u}) &= -2\mathbf{u}^T \mathbf{W} + \mathbf{u}^T \mathbf{C} \mathbf{u} + 2k_0 \mathbf{u}^T \boldsymbol{\beta},
 \end{aligned}$$

where $\mathbf{u} = (u_1, \dots, u_p)^T$. Following [30],

$$\sum_{i=1}^n [(\epsilon_i - \mathbf{u}^T \mathbf{x}_i / \sqrt{n})^2 - \epsilon_i^2] \xrightarrow{D} -2\mathbf{u}^T \mathbf{W} + \mathbf{u}^T \mathbf{C} \mathbf{u}$$

with finite-dimensional convergence holding trivially. Also

$$k \sum_{j=1}^p [(\beta_j + u_j / \sqrt{n})^2 - \beta_j^2] \xrightarrow{D} k_0 \sum_{j=1}^p u_j \beta_j.$$

Thus $V_n(\mathbf{u}) \xrightarrow{D} V(\mathbf{u})$ with the finite-dimensional convergence holding trivially. Since $V_n(\mathbf{u})$ is convex and $V(\mathbf{u})$ has a unique minimum, it follows that

$$\begin{aligned} \underset{D}{\operatorname{argmin}} V_n(\mathbf{u}) &= \sqrt{n}(\hat{\boldsymbol{\beta}}^{\text{RF}} - \boldsymbol{\beta}) \\ \underset{D}{\operatorname{argmin}} V(\mathbf{u}) &= \mathbf{C}^{-1}(\mathbf{W} - k_0\boldsymbol{\beta}) \sim N_p(-k_0\mathbf{C}^{-1}\boldsymbol{\beta}, \sigma^2\mathbf{C}^{-1}). \end{aligned}$$

It concludes

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{\text{RF}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(-k_0\mathbf{C}^{-1}\boldsymbol{\beta}, \sigma^2\mathbf{C}^{-1}).$$

□

Proof of Theorem 2. Because all of the pronounced estimators are special cases $\hat{\boldsymbol{\beta}}_1^{\text{Shrinkage}}$, we give the bias of this estimator here. Then the proof follows by applying relevant $g(\cdot)$ function in each estimator. Hence, we have

$$\text{ADB}(\hat{\boldsymbol{\beta}}_1^{\text{Shrinkage}}) = \text{ADB}(\hat{\boldsymbol{\beta}}_1^{\text{UR}}) - \lim_{n \rightarrow \infty} \sqrt{n}E \left[\left(\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \hat{\boldsymbol{\beta}}_1^{\text{RR}} \right) g(T_n) \right].$$

Using part one of Lemma 1, $\text{ADB}(\hat{\boldsymbol{\beta}}_1^{\text{UR}}) = -\boldsymbol{\eta}_{11.2}$. Further using part three of Lemma 1 along with Theorem 1 in Appendix B of [31], we have

$$\lim_{n \rightarrow \infty} \sqrt{n}E \left[\left(\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \hat{\boldsymbol{\beta}}_1^{\text{RR}} \right) g(T_n) \right] = \delta E \left[g(\chi_{p_2+2}^2(\Delta)) \right].$$

Therefore, the asymptotic bias of the general shrinkage estimator is given by

$$\text{ADB}(\hat{\boldsymbol{\beta}}_1^{\text{Shrinkage}}) = -\boldsymbol{\eta}_{11.2} - \delta E \left[g(\chi_{p_2+2}^2(\Delta)) \right].$$

The proof is complete considering the expressions for $E \left[g(\chi_{p_2+2}^2(\Delta)) \right]$ given in Table A1.

□

Table A1. Expressions for the corresponding $g(\cdot)$ functions in the proposed shrinkage estimators.

Shrinkage Estimator	$g(\cdot)$ Function	$E \left[g(\chi_{p_2+2}^2(\Delta)) \right]$
$\hat{\boldsymbol{\beta}}_1^{\text{PTR}}$	$I(T_n \leq \chi_{\alpha, p_2}^2)$	$H_{p_2+2}(\chi_{\alpha, p_2}^2; \Delta)$
$\hat{\boldsymbol{\beta}}_1^{\text{SR}}$	$(p_2 - 2)T_n^{-1}$	$(p_2 - 2)E(\chi_{p_2+2}^{-2}(\Delta))$
$\hat{\boldsymbol{\beta}}_1^{\text{PSR}}$	$(1 - (p_2 - 2)T_n^{-1})I(T_n \leq \chi_{\alpha, p_2}^2)$	$H_{p_2+2}(\chi_{\alpha, p_2}^2; \Delta) - (p_2 - 2)$
		$E \left[\chi_{p_2+2}^{-2}(\Delta) I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2) \right]$

Proof of Theorem 3. Similar to the proof of Theorem 2, we provide the ADQR of the shrinkage estimator $\hat{\boldsymbol{\beta}}_1^{\text{Shrinkage}}$ here. Then the proof follows by applying relevant $g(\cdot)$ function in each estimator. Hence, we have

$$\begin{aligned} \Gamma(\hat{\boldsymbol{\beta}}_1^{\text{Shrinkage}}, \mathbf{M}) &= \Gamma(\hat{\boldsymbol{\beta}}_1^{\text{UR}}, \mathbf{M}) \\ &\quad - 2 \lim_{n \rightarrow \infty} nE \left[\left(\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \boldsymbol{\beta} \right)^T \mathbf{M} \left(\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \hat{\boldsymbol{\beta}}_1^{\text{RR}} \right) g(T_n) \right] \\ &\quad + \lim_{n \rightarrow \infty} nE \left[\left(\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \hat{\boldsymbol{\beta}}_1^{\text{RR}} \right)^T \mathbf{M} \left(\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \hat{\boldsymbol{\beta}}_1^{\text{RR}} \right) g^2(T_n) \right] \end{aligned}$$

From Lemma 1, we have

$$\begin{aligned} \Gamma(\hat{\boldsymbol{\beta}}_1^{\text{UR}}, \mathbf{M}) &= \operatorname{tr} \left\{ \mathbf{M} \left[\operatorname{cov}(\hat{\boldsymbol{\beta}}_1^{\text{UR}}) \right] \right\} \\ &= \operatorname{tr} \left\{ \mathbf{M} \left[\lim_{n \rightarrow \infty} nE \left(\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \boldsymbol{\beta}_1 \right) \left(\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \boldsymbol{\beta}_1 \right)^T \right] \right\} \end{aligned}$$

$$\begin{aligned}
 &= \operatorname{tr} \left\{ \mathbf{M} \left[\lim_{n \rightarrow \infty} E(\boldsymbol{\vartheta}_n^{(1)} \boldsymbol{\vartheta}_n^{(1)T}) \right] \right\} \\
 &= \operatorname{tr} \left\{ \mathbf{M} \left[\lim_{n \rightarrow \infty} \operatorname{cov}(\boldsymbol{\vartheta}_n^{(1)}) + E(\boldsymbol{\vartheta}_n^{(1)}) E(\boldsymbol{\vartheta}_n^{(1)})^T \right] \right\} \\
 &= \operatorname{tr} \left\{ \mathbf{M} \left[\sigma^2 \mathbf{C}_{11.2}^{-1} + \boldsymbol{\eta}_{11.2} \boldsymbol{\eta}_{11.2}^T \right] \right\} \\
 &= \boldsymbol{\eta}_{11.2}^T \mathbf{M} \boldsymbol{\eta}_{11.2} + \sigma^2 \operatorname{tr}(\mathbf{M} \mathbf{C}_{11.2}^{-1}).
 \end{aligned}$$

From Lemma 1

$$\lim_{n \rightarrow \infty} nE \left[(\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \boldsymbol{\beta})^T \mathbf{M} (\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \hat{\boldsymbol{\beta}}_1^{\text{RR}}) g(T_n) \right] = \operatorname{tr} \left\{ \mathbf{M} \left[\lim_{n \rightarrow \infty} E(\boldsymbol{\vartheta}_n^{(3)} \boldsymbol{\vartheta}_n^{(1)T} g(T_n)) \right] \right\}.$$

Using double expectation, parts three and six of Lemma 1, and Theorems 1 & 3 in Appendix B of [31], we have

$$\begin{aligned}
 \lim_{n \rightarrow \infty} E(\boldsymbol{\vartheta}_n^{(3)} \boldsymbol{\vartheta}_n^{(1)T} g(T_n)) &= \lim_{n \rightarrow \infty} E \left[E(\boldsymbol{\vartheta}_n^{(3)} \boldsymbol{\vartheta}_n^{(1)T} g(T_n) | \boldsymbol{\vartheta}_n^{(3)}) \right] \\
 &= \lim_{n \rightarrow \infty} E \left[\boldsymbol{\vartheta}_n^{(3)} E(\boldsymbol{\vartheta}_n^{(1)T} g(T_n) | \boldsymbol{\vartheta}_n^{(3)}) \right] \\
 &= \lim_{n \rightarrow \infty} E \left[\boldsymbol{\vartheta}_n^{(3)} [-\boldsymbol{\eta}_{11.2} + \boldsymbol{\vartheta}_n^{(3)} - \boldsymbol{\delta}]^T g(T_n) | \boldsymbol{\vartheta}_n^{(3)} \right] \\
 &= -\lim_{n \rightarrow \infty} E \left[\boldsymbol{\vartheta}_n^{(3)} \boldsymbol{\eta}_{11.2}^T g(T_n) \right] + \lim_{n \rightarrow \infty} E \left[\boldsymbol{\vartheta}_n^{(3)} (\boldsymbol{\vartheta}_n^{(3)} - \boldsymbol{\delta})^T g(T_n) \right] \\
 &= -\lim_{n \rightarrow \infty} E \left[\boldsymbol{\vartheta}_n^{(3)} g(T_n) \right] \boldsymbol{\eta}_{11.2}^T + \lim_{n \rightarrow \infty} E \left[\boldsymbol{\vartheta}_n^{(3)} \boldsymbol{\vartheta}_n^{(3)T} g(T_n) \right] \\
 &\quad - \lim_{n \rightarrow \infty} E \left[\boldsymbol{\vartheta}_n^{(3)} g(T_n) \right] \boldsymbol{\delta}^T \\
 &= -\boldsymbol{\delta} \boldsymbol{\eta}_{11.2}^T E \left[g(\chi_{p_2+2}^2(\Delta)) \right] + \sigma^2 (\mathbf{C}_{11.2}^{-1} - \mathbf{C}_{11}^{-1}) E \left[g(\chi_{p_2+2}^2(\Delta)) \right] \\
 &\quad + \boldsymbol{\delta} \boldsymbol{\delta}^T E \left[g(\chi_{p_2+4}^2(\Delta)) \right] - \boldsymbol{\delta} \boldsymbol{\delta}^T E \left[g(\chi_{p_2+2}^2(\Delta)) \right]
 \end{aligned}$$

Thus, it yields

$$\begin{aligned}
 \lim_{n \rightarrow \infty} nE \left[(\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \boldsymbol{\beta})^T \mathbf{M} (\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \hat{\boldsymbol{\beta}}_1^{\text{RR}}) g(T_n) \right] &= -\boldsymbol{\eta}_{11.2}^T \mathbf{M} \boldsymbol{\delta} \\
 &\quad E \left[g(\chi_{p_2+2}^2(\Delta)) \right] \\
 &\quad + \sigma^2 \operatorname{tr}[\mathbf{M} (\mathbf{C}_{11.2}^{-1} - \mathbf{C}_{11}^{-1})] \\
 &\quad E \left[g(\chi_{p_2+2}^2(\Delta)) \right] \\
 &\quad + \boldsymbol{\delta}^T \mathbf{M} \boldsymbol{\delta} E \left[g(\chi_{p_2+4}^2(\Delta)) \right] \\
 &\quad - \boldsymbol{\delta}^T \mathbf{M} \boldsymbol{\delta} E \left[g(\chi_{p_2+2}^2(\Delta)) \right]
 \end{aligned}$$

In a similar manner, we have

$$\begin{aligned}
 \lim_{n \rightarrow \infty} nE \left[(\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \hat{\boldsymbol{\beta}}_1^{\text{RR}})^T \mathbf{M} (\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \hat{\boldsymbol{\beta}}_1^{\text{RR}}) g^2(T_n) \right] &= \operatorname{tr} \left\{ \mathbf{M} \left[\lim_{n \rightarrow \infty} E(\boldsymbol{\vartheta}_n^{(3)} \boldsymbol{\vartheta}_n^{(3)T} g(T_n)) \right] \right\} \\
 &= \sigma^2 \operatorname{tr}[\mathbf{M} (\mathbf{C}_{11.2}^{-1} - \mathbf{C}_{11}^{-1})] \\
 &\quad E \left[g^2(\chi_{p_2+2}^2(\Delta)) \right] \\
 &\quad + \boldsymbol{\delta}^T \mathbf{M} \boldsymbol{\delta} E \left[g^2(\chi_{p_2+4}^2(\Delta)) \right]
 \end{aligned}$$

Gathering all required expressions, we finally have

$$\begin{aligned}
 \Gamma(\hat{\beta}_1^{\text{Shrinkage}}, M) &= \eta_{11.2}^T M \eta_{11.2} + \sigma^2 \text{tr}(M C_{11.2}^{-1}) \\
 &\quad - 2 \left\{ -\eta_{11.2}^T M \delta E \left[g(\chi_{p_2+2}^2(\Delta)) \right] \right. \\
 &\quad + \sigma^2 \text{tr} \left[M (C_{11.2}^{-1} - C_{11}^{-1}) \right] E \left[g(\chi_{p_2+2}^2(\Delta)) \right] \\
 &\quad \left. + \delta^T M \delta E \left[g(\chi_{p_2+4}^2(\Delta)) \right] - \delta^T M \delta E \left[g(\chi_{p_2+2}^2(\Delta)) \right] \right\} \\
 &\quad + \sigma^2 \text{tr} \left[M (C_{11.2}^{-1} - C_{11}^{-1}) \right] E \left[g^2(\chi_{p_2+2}^2(\Delta)) \right] \\
 &\quad + \delta^T M \delta E \left[g^2(\chi_{p_2+4}^2(\Delta)) \right] \\
 &= \eta_{11.2}^T M \eta_{11.2} + \sigma^2 \text{tr}(M C_{11.2}^{-1}) + 2 \eta_{11.2}^T M \delta E \left[g(\chi_{p_2+2}^2(\Delta)) \right] \\
 &\quad + \sigma^2 \text{tr} \left[M (C_{11.2}^{-1} - C_{11}^{-1}) \right] \left\{ -2 E \left[g(\chi_{p_2+2}^2(\Delta)) \right] + E \left[g^2(\chi_{p_2+2}^2(\Delta)) \right] \right\} \\
 &\quad + \delta^T M \delta \\
 &\quad \times \left\{ -2 E \left[g(\chi_{p_2+4}^2(\Delta)) \right] + 2 E \left[g(\chi_{p_2+2}^2(\Delta)) \right] + E \left[g^2(\chi_{p_2+2}^2(\Delta)) \right] \right\}.
 \end{aligned}$$

The proof is complete using Table A1. \square

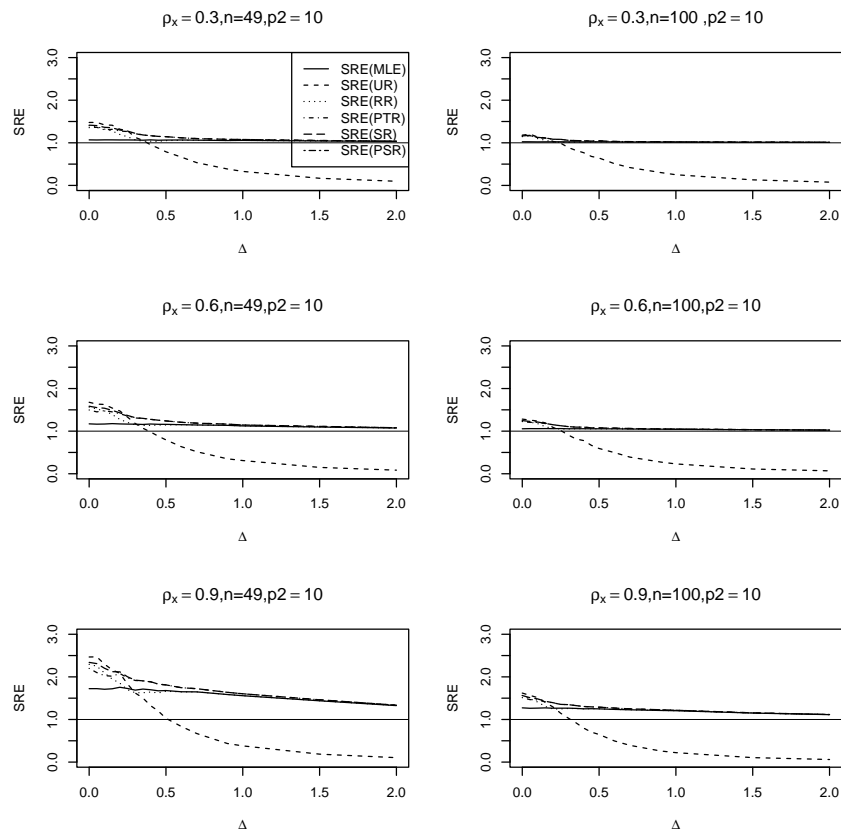


Figure A1. SRE of the suggested estimators with respect to the MLE ($\hat{\beta}_1$) for $n = 49, 100$, $\rho_x \in \{0.3, 0.6, 0.9\}$, $\rho = 0.90$, and $(p_1, p_2) = (5, 10)$.

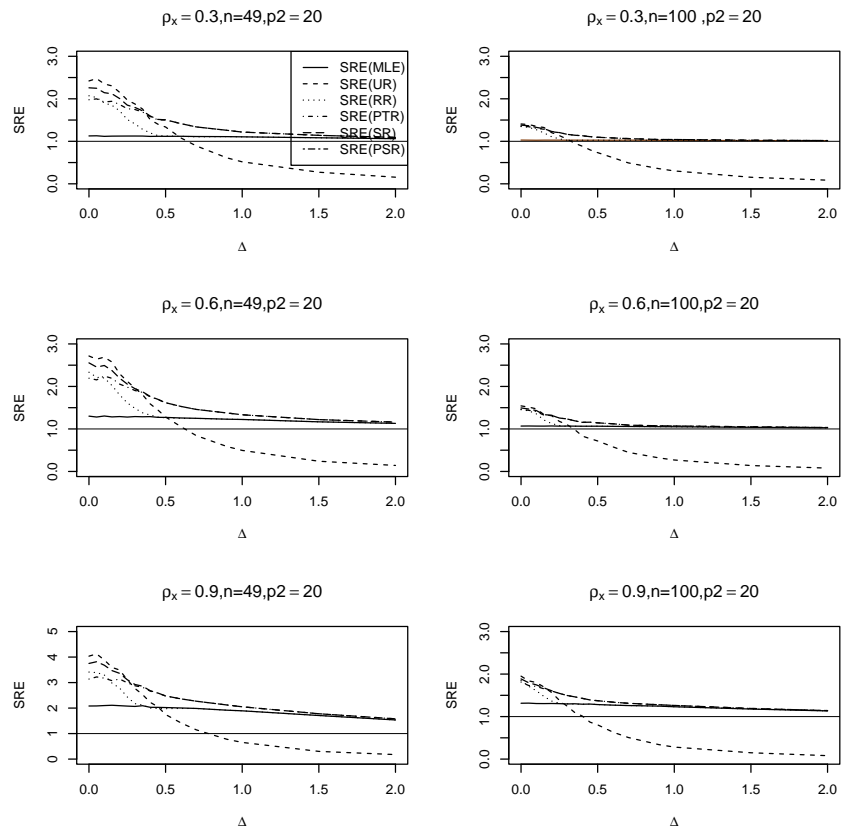


Figure A2. SRE of the suggested estimators with respect to the MLE ($\hat{\beta}_1$) for $n = 49, 100$, $\rho_x \in \{0.3, 0.6, 0.9\}$, $\rho = 0.90$, and $(p_1, p_2) = (5, 20)$.

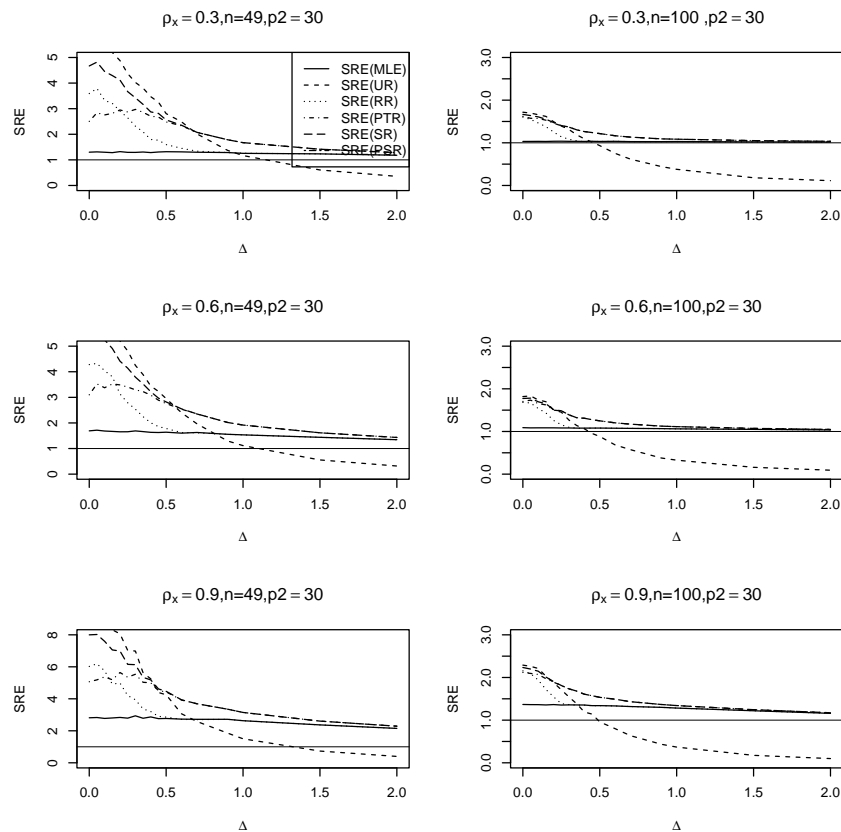


Figure A3. SRE of the suggested estimators with respect to the MLE ($\hat{\beta}_1$) for $n = 49, 100$, $\rho_x \in \{0.3, 0.6, 0.9\}$, $\rho = 0.90$, and $(p_1, p_2) = (5, 30)$.

References

1. Dai, X.; Li, E.; Tian, M. Quantile regression for varying coefficient spatial error models. *Commun. Stat.—Theory Methods* **2019**, *50*, 2382–2397. [[CrossRef](#)]
2. Higazi, S.F.; Abdel-Hady, D.H.; Al-Oulfi, S.A. Application of spatial regression models to income poverty ratios in Middle Delta contiguous counties in Egypt. *Pak. J. Stat. Oper. Res.* **2013**, *9*, 93. [[CrossRef](#)]
3. Piscitelli, A. Spatial Regression of Juvenile Delinquency: Revisiting Shaw and McKay. *Int. J. Crim. Justice Sci.* **2019**, *14*, 132–147.
4. Liu, R.; Yu, C.; Liu, C.; Jiang, J.; Xu, J. Impacts of haze on housing prices: An empirical analysis based on data from Chengdu (China). *Int. J. Environ. Res. Public Health* **2018**, *15*, 1161. [[CrossRef](#)] [[PubMed](#)]
5. Yildirim, V.; Mert, K.Y. Robust estimation approach for spatial error model. *J. Stat. Comput. Simul.* **2020**, *90*, 1618–1638. [[CrossRef](#)]
6. Cressie, N. *Statistics for Spatial Data*; John Wiley & Sons: Nashville, TN, USA, 1993.
7. Cressie, N.; Wikle, C.K. *Statistics for Spatio-Temporal Data*; Wiley-Blackwell: Chichester, UK, 2011.
8. Haining, R. *Spatial Data Analysis: Theory and Practice*; Cambridge University Press: Cambridge, UK, 2003.
9. Al-Momani, M.; Riaz, M.; Saleh, M.F. Pretest and shrinkage estimation of the regression parameter vector of the marginal model with multinomial responses. *Stat. Pap.* **2022**, *64*, 2101–2117. [[CrossRef](#)]
10. Lisawadi, S.; Ahmed, S.E.; Reangsephet, O. Post estimation and prediction strategies in negative binomial regression model. *Int. J. Model. Simul.* **2020**, *41*, 463–477. [[CrossRef](#)]
11. Nkurunziza, S.; Al-Momani, M.; Lin, E.Y. Shrinkage and lasso strategies in high-dimensional heteroscedastic models. *Commun. Stat.—Theory Methods* **2016**, *45*, 4454–4470. [[CrossRef](#)]
12. Hoerl, A.E.; Kennard, R.W. A new Liu-type estimator in linear regression model. *Technometrics* **1970**, *12*, 55–67. [[CrossRef](#)]
13. Kejian, L. A new class of biased estimate in linear regression. *Commun. Stat.—Theory Methods* **1993**, *22*, 393–402. [[CrossRef](#)]
14. Li, Y.; Yang, H. A new Liu-type estimator in linear regression model. *Stat. Pap.* **2010**, *53*, 427–437. [[CrossRef](#)]
15. Arashi, M.; Kibria, B.M.G.; Norouzirad, M.; Nadarajah, S. Improved preliminary test and Stein-rule Liu estimators for the ill-conditioned elliptical linear regression model. *J. Multivar. Anal.* **2014**, *126*, 53–74. [[CrossRef](#)]
16. Arashi, M.; Norouzirad, M.; Roozbeh, M.; Khan, N.M. A high-dimensional counterpart for the ridge estimator in multicollinear situations. *Mathematics* **2021**, *9*, 3057. [[CrossRef](#)]
17. Al-Momani, M. Liu-type pretest and shrinkage estimation for the conditional autoregressive model. *PLoS ONE* **2023**, *18*, e0283339. [[CrossRef](#)] [[PubMed](#)]
18. Mardia, K.V.; Marshall, R.J. Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **1984**, *71*, 135–146. [[CrossRef](#)]
19. Al-Momani, M.; Hussein, A.A.; Ahmed, S.E. Penalty and related estimation strategies in the spatial error model. *Stat. Neerl.* **2016**, *71*, 4–30. [[CrossRef](#)]
20. Bivand, R. R packages for Analyzing Spatial Data: A comparative case study with Areal Data. *Geogr. Anal.* **2022**, *54*, 488–518. [[CrossRef](#)]
21. Harrison, D.; Rubinfeld, D.L. Hedonic housing prices and the demand for Clean Air. *J. Environ. Econ. Manag.* **1978**, *5*, 81–102. [[CrossRef](#)]
22. Gilley, O.W.; Pace, R.K. On the Harrison and Rubinfeld data. *J. Environ. Econ. Manag.* **1996**, *31*, 403–405. [[CrossRef](#)]
23. Pace, R.K.; Gilley, O.W. Using the Spatial Configuration of the Data to Improve Estimation. *J. Real Estate Financ. Econ.* **1997**, *14*, 333–340. [[CrossRef](#)]
24. Solow, A.R. Bootstrapping correlated data. *J. Int. Assoc. Math. Geol.* **1985**, *17*, 769–775. [[CrossRef](#)]
25. Boonstra, P.S.; Mukherjee, B.; Taylor, J.M. A small-sample choice of the tuning parameter in ridge regression. *Stat. Sin.* **2015**, *23*, 1185. [[CrossRef](#)] [[PubMed](#)]
26. Seber, G.A.F. *Spatial Data Analysis: Theory and Practice*. In *A Matrix Handbook for Statisticians*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
27. Lee, L. Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances. *Econom. Rev.* **2003**, *22*, 307–335. [[CrossRef](#)]
28. Liu, S.F.; Yang, Z. Asymptotic Distribution and Finite Sample Bias Correction of QML Estimators for Spatial Error Dependence Model. *Econometrics* **2015**, *3*, 376–411. [[CrossRef](#)]
29. Yuzbasi, B.; Arashi, M.; Ahmed, S.E. Shrinkage estimation strategies in generalized ridge regression models under low/high-dimension regime. *Int. Stat. Rev.* **2020**, *88*, 229–251. [[CrossRef](#)]
30. Fu, W.; Knight, K. Asymptotics for lasso-type estimators. *Ann. Stat.* **2000**, *28*, 1356–1378. [[CrossRef](#)]
31. Judge, G.G.; Bock, M.E. *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*; North-Holland Pub. Co.: Amsterdam, The Netherlands, 1978.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.