Preprints.org

Article

Ridge-Type Pretest and Shrinkage Estimation in Spatial Error Model: An Application to Housing Cost Data

[Marwan Al-Momani](#) * and [Mohammad Arashi](#)

Article

Ridge-Type Pretest and Shrinkage Estimation in Spatial Error Model. An Application to Housing Cost Data

Marwan Al-Momani [1,*] and Mohammad Arashi [2]

[1]   Department of Mathematics, College of Sciences, University of sharjah, UAE; malmomani@sharjah.ac.ae
[2]   Department of Statistics, Ferdowsi University of Mashhad, Iran; arashi@um.ac.ir
*   Correspondence: malmomani@sharjah.ac.ae

Abstract: Spatial regression models have garnered significant attention across several disciplines, including functional magnetic resonance imaging analysis, econometrics, home price analysis, and many other domains. The phenomenon of sparsity is often found in nature, when a limited number of factors contribute significantly to the overall variation. Spatial regression models frequently use sparsity to indicate less complex computational and more superficial covariance structures. 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 for estimating the vector of regression coefficients in the spatial error model, taking into consideration of the prior knowledge that some coefficients are insignificant and there is multicollinearity among the regressors. It also introduces pretest and shrinkage ridge estimators for spatial error regression models, evaluating their performance compared to traditional maximum likelihood estimators. It also assesses their efficacy using real-world data and bootstrapping techniques for comparison purposes.

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

---

1. Introduction

Data collected over a geographic region may generally show some dependence, whereby nearby observations are more similar than those made at significant distances. The incorporation of a covariance structure into conventional statistical models allows for the modeling of this phenomenon. Spatial regression models, incorporating various spatial dependencies, are increasingly utilized in various disciplines like geology, epidemiology, disease surveillance, urban planning, and econometrics.

Autoregressive models in the context of time-series reflect the data at time t as a linear combination of the most recent observations. Similarly, in the spatial framework, these models represent the data from a certain spatial place as a function of data from neighboring locations. The collection of data is often associated with a geographical location known as a site, and a metric of distance is used to define the notion of proximity between these sites.

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. [20] investigated the quantile regression estimation for the SE model with potentially variable coefficients. They established the proposed estimators' asymptotic properties. [27] applied the SE model to examine the existence of spatial clustering and correlation between neighboring counties for the data from Egypt's 2006 census. [38] used the SE model in order to evaluate the Social Disorganization Theory. [33] used the combined application of SE model and Spatial Lag Model based on cross-section data of 20 districts in Chengdu. They found that the haze had a negative impact on both the selling and rental prices of houses. [43] 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 [15,18,19,25,42], among others.

In frequentist statistics, we use sample information to establish inferences about unknown parameters, while we columbine the sample information with some uncertain prior information (UPI) which is also known as non-sample information to draw conclusions about the unknown parameters in Bayesian statistics. It is possible that such subjective UPI is not always available. However, procedures for model selection such as Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), or model selection techniques could still be used to provide the UPIs.

One of the Initial trails to incorporate the sample information and the UPI used to estimate the regression parameters by using sample information and the UPI is referred to as pretest estimation. Pretest estimator relies on determining whether some of the regression coefficients are significant, after which the pretest selects either the full model estimator or the revised model estimator, which has fewer number of coefficients. Obviously, the pertest chooses the full or sub model estimators based on a binary weights. A new modification of the pretest estimator that uses a smooth wights between the full and sub model estimators is known as shrinkage estimator, it shrinks the estimator of the regression coefficients in the direction of a target value impacted by the UPI. Nevertheless, the modified shrinkage estimator suffers sometimes from an over shrinkage phenomenon. Later on an improved version of this estimator is proposed that controls the over shrinkage issue, which is known as positive shrinkage estimator.

The concept of using pretest and shrinkage estimating methodologies has received considerable attention by many researchers, for example, [6] introduced an efficient estimation using the pretest and shrinkage methods to estimate the regression coefficients vector of the marginal model in case of multinomial responses. [32] developed different shrinkage and penalty estimation strategies for the negative binomial regression model when an over fitting and uncertainty information exist about the subspace. [35] 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. More details about the pretest and shrinkage can be found in [1,4,21], among other references.

The mulicolinearity is a major issue when fitting a multiple linear regression model using the ordinary least squares method (OLS) that appears when some of the 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. [28] introduced the concept of ridge regression as a solution for nonorthogonal problems.They showed the estimator improve the mean square error of estimation. [30] introduced a new biased estimator and demonstrated theoretically and numerically the improvement of the new one. [31] proposed a new version of Liu estimator for the vector of parameters in a linear regression model based on some prior information.

Using the idea of shrinkage, [7] introduced an improved form of Liu-type estimator. Analytical and numerical results were used to demonstrate the proposed method's superiority. [13] suggested the pretest and shrinkage ridge estimation for the linear regression model and showed the benefit of using the recommended estimators in conjunction with certain penalty estimators. [9] defined Liu-type rank-based estimators for robust regression, analyzed their asymptotic behavior, established biasing parameter superiority criteria, and supported their findings with numerical computations. [14] introduced the pretest and shrinkage approaches based on generalized ridge regression estimation. Later on, [8] suggested the use of the ridge estimator as a suitable approach for handling high-dimensional multicollinear data. Further, [11] proposed an enhanced ridge method for genome regression modeling and used a rank ridge estimator to estimate parameters and make predictions in the presence of multicollinearity and outliers within the data set. Recently, [3] proposed a novel pretest and shrinkage estimate technique, known as the Liu-type approach, developed for the conditional autoregressive model. For more information, we advice the reader to consult the following references [22,23].

In this article, we aim to 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 $\boldsymbol{\beta}$ as $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)^T$, where $\boldsymbol{\beta}_1$ is a $p_1 \times 1$ vector that contains the coefficients of the main effect, and $\boldsymbol{\beta}_2$ is a $p_2 \times 1$ vector of irrelevant coefficients, with $p_1 + p_2 = p$. Mainly, we focus in estimating the vector $\boldsymbol{\beta}_1$ when the UPI indicates that $\boldsymbol{\beta}_2$ is ineffective, which can be achieved by testing a statistical hypothesis of the form $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$. 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 have taken into account the pretest, shrinkage, and positive shrinkage ridge estimators for the vector $\boldsymbol{\beta}_1$.

In accordance with our goal the paper is organized as follows. Section 2 offers an overview of the SE model. The discussion of the maximum likelihood estimators for the parameters of the SE model are discussed in 3. In Section 4, we proposed the pretest and shrinkage ridge estimators. Asymptotic analysis of the proposed estimators and some theoretical results are presented in Section 5. The set of estimators are compared numerically using simulated and real data example in Section 6. Some concluding remarks are given in Section 7. An appendix containing some proofs is given at the end of this manuscript.

2. Spatial Error Model

Let $\boldsymbol{s} = \{s_1, s_2, \ldots, s_n\}$ represents a set of (n) spatial sites (frequently known as locations, regions, etc.). The set $\boldsymbol{s}$ 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: $K(s_j) = \{s_i : s_i \text{ is a neighbor of } s_j\}, i = 1, 2, \ldots, n$. 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 $\boldsymbol{Y}_n(\boldsymbol{s}) = \{Y(s_1), Y(s_2), \ldots, Y(s_n)\}$ be a vector of observations collected at sites $\{s_1, s_2, \ldots, s_n\}$, and $\boldsymbol{X}(\boldsymbol{s}) = (\boldsymbol{X}(s_1), \boldsymbol{X}(s_2), \ldots, \boldsymbol{X}(s_n))$ be the $(n \times p)$ matrix of covariates. Following Cressie and Wikle [19], the SE model models the response $\boldsymbol{Y}$ at the jth site s_j as:

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

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

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$ be the $(p \times 1)$ vector of unknown regression parameters, known as the large-scale effect, $\boldsymbol{e}(\boldsymbol{s}) = (e(s_1), e(s_2), \ldots, e(s_n))^T$ is noise vector that has a Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Omega} = diag\{\sigma_j^2\}_{j=1}^n$. The parameters λ_{ji} are used to model the spatial dependencies among the errors $\epsilon_j, j = 1, 2, \ldots, n$, with $\lambda_{jj} = 0$. Let $\boldsymbol{\Lambda} = \{\lambda_{j,i}\}_{j,i=1}^n$, and assume that $(\boldsymbol{I} - \boldsymbol{\Lambda})$ is invertible, where $\boldsymbol{I}$ is the $(n \times n)$ identity matrix, then by ignoring the spatial indices, the SE in (1) can be rewritten in matrix format as:

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

Nature exhibits sparsity in many situations, which means that a small number of factors can account for the majority of the observed variability. Sparsity is frequently used in spatial regression models to imply covariance structures that are easier to compute. Consequently, by setting $\boldsymbol{\Omega} = \sigma^2 \boldsymbol{I}$, and $\boldsymbol{\Lambda} = \rho \boldsymbol{W}$, where σ^2 is the variance component, ρ is the spatial dependance parameter, and $\boldsymbol{W}$ 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 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 $W^* = \{\frac{w_{ji}}{w_{j+}}\}_{j,i=1}^n$. So, The SE regression model can be rewritten as follows

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

$$\text{and} \quad V_n = (1 - \rho W^*)^{-1} \left(1 - \rho W^{*T}\right)^{-1} \tag{5}$$

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 [18]. At first we fix ρ and find the MLEs of β, σ^2 as a function of ρ, which are given below

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

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

Then, we plug $\hat{\beta}$ and $\hat{\sigma}^2$ into the log-likelihood and obtain the MLE of ρ by maximizing the profile log-likelihood function. Finally, the MLEs of β and σ^2 are computed by replacing ρ by $\hat{\rho}$ in equations (6) and (7), respectively. [34] 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 coefficients 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 are consisting 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:

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

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 [5]. 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 the 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 on estimating β_1 when X_2 does not sufficiently account for the variation 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 $\hat{\boldsymbol{\beta}}_1^S$ be the MLE of $\boldsymbol{\beta}_1$ of the model in (11), then

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

Obviously, $\hat{\boldsymbol{\beta}}_1^S$ will have a better performance than $\hat{\boldsymbol{\beta}}_1$ if the null hypothesis in (10) is true, while the opposite occurs when $\boldsymbol{\beta}_2$ begins to move away from the null space. Yet, the restricted strategy method can provide under-fitted and highly biased model. To dominate the large bias, we propose the ridge-type estimation strategy of the full, and reduced models, 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 parameters vector $\boldsymbol{\beta}_1$ in (11). Following [28], The ridge estimator of $\boldsymbol{\beta}$ for the model given in (4) is defined as

$$\hat{\boldsymbol{\beta}}^{RF} = \left(\boldsymbol{X}^T \hat{\boldsymbol{V}}_n^{-1} \boldsymbol{X} + k\boldsymbol{I}_p \right)^{-1} \boldsymbol{X}^T \hat{\boldsymbol{V}}_n^{-1} \boldsymbol{Y}, \tag{13}$$

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

4.1. Full and Reduced Models Ridge Estimators

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

$$\hat{\boldsymbol{\beta}}_1^{UR} = \left(\boldsymbol{X}_1^T \boldsymbol{A}_{X_2} \boldsymbol{X}_1 + k_f \boldsymbol{I}_{p_1} \right)^{-1} \boldsymbol{X}_1^T \boldsymbol{A}_{X_2} \boldsymbol{Y}, \tag{14}$$

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

$$\hat{\boldsymbol{\beta}}_1^{RR} = \left(\boldsymbol{X}_1^T \hat{\boldsymbol{V}}_n^{-1} \boldsymbol{X}_1 + k_r \boldsymbol{I}_{p_1} \right)^{-1} \boldsymbol{X}_1^T \hat{\boldsymbol{V}}_n^{-1} \boldsymbol{Y}, \tag{15}$$

where k_r is the ridge parameter for restricted model estimator $\hat{\boldsymbol{\beta}}_1^{RR}$. When the null hypothesis in (10) is accurate or almost accurate (i.e., when $\boldsymbol{\beta}_2$ is close to zero), $\hat{\boldsymbol{\beta}}_1^{RR}$ is generally a more effective estimator than $\hat{\boldsymbol{\beta}}_1^{UR}$. Nevertheless, as $\boldsymbol{\beta}_2$ deviates from the zero space, $\hat{\boldsymbol{\beta}}_1^{RR}$ becomes inefficient in comparison with the unrestricted estimator $\hat{\boldsymbol{\beta}}_1^{UR}$. In addition to the gain obtained by employing the idea of ridge estimation to the MLE of $\beta 1$, we also aim to find estimators that are functions of $\hat{\boldsymbol{\beta}}_1^{UR}$ and $\hat{\boldsymbol{\beta}}_1^{RR}$ and intended 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 $\hat{\boldsymbol{\beta}}_1^{UR}$ if H_0 is rejected or the restricted ridge estimator $\hat{\boldsymbol{\beta}}_1^{RR}$ if not. An appropriate test statistics to test the hypothesis in (10) is:

$$T_n = \frac{(\hat{\boldsymbol{\beta}}_2^{UR})^T \left(\boldsymbol{X}_2^T \boldsymbol{A}_{X_1} \boldsymbol{X}_2 \right) (\hat{\boldsymbol{\beta}}_2^{UR})}{s^2},$$

where $\boldsymbol{A}_{X_1}$ is defined in a similar manner as $\boldsymbol{A}_{X_2}$, $\hat{\boldsymbol{\beta}}_2^{UR} = \left(\boldsymbol{X}_2^T \boldsymbol{A}_{X_1} \boldsymbol{X}_2 \right)^{-1} \boldsymbol{X}_2^T \boldsymbol{A}_{X_1} \boldsymbol{Y}$, and $s^2 = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{RF})^T (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{RF})/(n - p)$, which is a consistent estimator of σ^2, and the

statistic T_n follows asymptotically a chi-square distribution with p_2 degrees of freedom under the null hypothesis. Hence, the pretest estimator, denoted by $\hat{\beta}_1^{\text{PTR}}$, is given by

$$\hat{\beta}_1^{\text{PTR}} = \hat{\beta}_1^{\text{UR}} - \left(\hat{\beta}_1^{\text{UR}} - \hat{\beta}_1^{\text{RR}}\right) I\left(T_n \le \chi^2_{\alpha, p_2}\right), \tag{16}$$

where $I(.)$ is an indicator function, and χ^2_{α, p_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 $\hat{\beta}_1^{\text{UR}}$ if the null hypothesis is rejected, and $\hat{\beta}_1^{\text{RR}}$ otherwise based on a binary weights. These drawbacks can be improved using smoother weights of the two estimators $\hat{\beta}_1^{\text{UR}}$ and $\hat{\beta}_1^{\text{RR}}$ instead, which is known as the shrinkage estimator. It is denoted by, $\hat{\beta}_1^{\text{SR}}$, and given by

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

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 $\hat{\beta}_1^{\text{SR}}$, resolves this issue. It is denoted by $\hat{\beta}_1^{\text{PSR}}$, and given by

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

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

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

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

5. Asymptotic Analysis

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

$$H_{(n)} : \boldsymbol{\beta}_{2(n)} = \frac{\boldsymbol{\xi}}{\sqrt{n}}, \quad \boldsymbol{\xi} \in \Re^{p_2}, \text{with } \boldsymbol{\xi} \ne \mathbf{0} \tag{20}$$

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

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

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

$$\Gamma(\boldsymbol{\beta}_1^*, \boldsymbol{M}) = E\left(\boldsymbol{\vartheta}_n^{*T} \boldsymbol{M} \boldsymbol{\vartheta}_n^*\right) = \int \left(\boldsymbol{x}_1^T \boldsymbol{M} \boldsymbol{x}_1\right) dK(\boldsymbol{x}_1) \tag{21}$$

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

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

For the purpose of deriving asymptotic distributional properties, in addition to the first four assumptions of [34], we set the following regularity conditions:

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

$$\boldsymbol{C}_n^{-1} = \begin{pmatrix} \boldsymbol{X}_1^T\hat{\boldsymbol{V}}_n^{-1}\boldsymbol{X}_1 & \boldsymbol{X}_1^T\hat{\boldsymbol{V}}_n^{-1}\boldsymbol{X}_2 \\ \boldsymbol{X}_2^T\hat{\boldsymbol{V}}_n^{-1}\boldsymbol{X}_1 & \boldsymbol{X}_2^T\hat{\boldsymbol{V}}_n^{-1}\boldsymbol{X}_2 \end{pmatrix}^{-1} \quad \text{and}$$

$$\boldsymbol{C}^{-1} = \begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{C}_{11.2}^{-1} & -\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22.1}^{-1} \\ -\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21}\boldsymbol{C}_{11.2}^{-1} & \boldsymbol{C}_{22.1}^{-1} \end{pmatrix},$$

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

In sequel, we call 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 $\hat{\boldsymbol{\beta}}_1^{\text{UR}}$ and the restricted ridge estimator $\hat{\boldsymbol{\beta}}_1^{\text{RR}}$. To this end, we make use of the following lemma. The proof is provided in the Appendix.

Lemma 1. *Assume the NRC. If $k/\sqrt{n} \to k_o \ge 0$, then*

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

where $\xrightarrow{D}$ denotes convergence in distribution. Indeed, Lemma 1 enables us to give some asymptotic distributional results about the estimators $\hat{\boldsymbol{\beta}}_1^{\text{UR}}$ and $\hat{\boldsymbol{\beta}}_1^{\text{RR}}$, presented in the following theorem, which are easy to prove. See [21] for similar results.

Theorem 1. *Let $\boldsymbol{\vartheta}_n^{(1)} = \sqrt{n}\left(\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \boldsymbol{\beta}_1\right)$, $\boldsymbol{\vartheta}_n^{(2)} = \sqrt{n}\left(\hat{\boldsymbol{\beta}}_1^{\text{RR}} - \boldsymbol{\beta}_1\right)$, and $\boldsymbol{\vartheta}_n^{(3)} = \sqrt{n}\left(\hat{\boldsymbol{\beta}}_1^{\text{UR}} - \hat{\boldsymbol{\beta}}_1^{\text{RR}}\right)$. Assume the local alternatives in (20) and NRC. Then, as $n \to \infty$ we have*

1. $\boldsymbol{\vartheta}_n^{(1)} \sim N_{p_1}(-\boldsymbol{\eta}_{11.2}, \sigma^2\boldsymbol{C}_{11.2}^{-1})$
2. $\boldsymbol{\vartheta}_n^{(2)} \sim N_{p_1}(\boldsymbol{\delta} - \boldsymbol{\eta}_{11.2}, \sigma^2\boldsymbol{C}_{11}^{-1})$
3. $\boldsymbol{\vartheta}_n^{(3)} \sim N_{p_1}(\boldsymbol{\delta}, \sigma^2(\boldsymbol{C}_{11.2}^{-1} - \boldsymbol{C}_{11}^{-1}))$
4. $\begin{pmatrix} \boldsymbol{\vartheta}_n^{(1)} \\ \boldsymbol{\vartheta}_n^{(3)} \end{pmatrix} \sim N_{2p_1}\left(\begin{pmatrix} -\boldsymbol{\eta}_{11.2} \\ \boldsymbol{\delta} \end{pmatrix}, \sigma^2\begin{pmatrix} \boldsymbol{C}_{11.2}^{-1} & \boldsymbol{C}_{11.2}^{-1} - \boldsymbol{C}_{11}^{-1} \\ \boldsymbol{C}_{11.2}^{-1} - \boldsymbol{C}_{11}^{-1} & \boldsymbol{C}_{11.2}^{-1} - \boldsymbol{C}_{11}^{-1} \end{pmatrix}\right)$
5. $\begin{pmatrix} \boldsymbol{\vartheta}_n^{(2)} \\ \boldsymbol{\vartheta}_n^{(3)} \end{pmatrix} \sim N_{2p_1}\left(\begin{pmatrix} \boldsymbol{\delta} - \boldsymbol{\eta}_{11.2} \\ \boldsymbol{\delta} \end{pmatrix}, \sigma^2\begin{pmatrix} \boldsymbol{C}_{11}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_{11.2}^{-1} - \boldsymbol{C}_{11}^{-1} \end{pmatrix}\right)$
6. $E\left[\boldsymbol{\vartheta}_n^{(1)}|\boldsymbol{\vartheta}_n^{(3)}\right] = -\boldsymbol{\eta}_{11.2} + \boldsymbol{\vartheta}_n^{(3)} - \boldsymbol{\delta}$
7. $Pr(T_n \le x) = H_q(x; \Delta)$, where $H_q(x; \Delta)$ is the cumulative distribution function of a non-central chi-square distribution with q degrees of freedom and non-centrality parameter Δ.

where $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^T, \boldsymbol{\eta}_2^T)^T = -k_o\boldsymbol{C}^{-1}\boldsymbol{\beta}$, $\boldsymbol{\eta}_{11.2} = \boldsymbol{\eta}_1 - \boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}((\boldsymbol{\beta}_2 - \boldsymbol{\xi}) - \boldsymbol{\eta}_2)$, $\boldsymbol{\delta} = \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{\xi}$.

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 give 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. $\text{ADB}(\hat{\boldsymbol{\beta}}_1^{\text{PTR}}) = -\boldsymbol{\eta}_{11.2} - \delta H_{p_2+2}(\chi_{\alpha,p_2}^2; \Delta)$
2. $\text{ADB}(\hat{\boldsymbol{\beta}}_1^{\text{SR}}) = -\boldsymbol{\eta}_{11.2} - (p_2 - 2)\delta E(\chi_{p_2+2}^{-2}(\Delta))$
3. $\text{ADB}(\hat{\boldsymbol{\beta}}_1^{\text{PSR}}) = -\boldsymbol{\eta}_{11.2} - \delta H_{p_2+2}(\chi_{\alpha,p_2}^2; \Delta)$

$+ (p_2 - 2)\delta E\left[\chi_{p_2+2}^{-2}(\Delta)I(\chi_{p_2+2}^2(\Delta) \leq p_2 - 2)\right], where$

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

For the proof, refer to Appendix.

The following result reveals the expressions for the ADQR of the proposed shrinkage estimators.

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

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

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

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

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

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

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

For the proof, refer to the Appendix.

6. Numerical Analysis

To demonstrate our theoretical findings, we first use Monte Carlo simulation experiments, then apply the set of proposed estimators to a real data set. The Monte Carlo simulation is used to investigate the performance of the ridge-type set of estimators in comparison to the MLE ($\hat{\boldsymbol{\beta}}_1$) 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 the MLE using Monte Carlo simulation experiments based on their simulated mean squared errors. In each one of theses experiments, we consider an $(N \times N)$ 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 a multicollinearity exits, we generate the design matrix $\boldsymbol{X}$ from multivariate normal distribution with mean $\boldsymbol{0}$, and a variance-covariance matrix with first order autoregressive structure in which $cov(\boldsymbol{X}_i, \boldsymbol{X}_j) = \begin{cases} \rho_x^{|i-j|} & i \neq j \\ 1 & i = j \end{cases}$ and apply it for $\rho_x \in \{0.3, 0.6, 0.9\}$, while the error term $\boldsymbol{\epsilon}$ is generated from another multivariate normal with mean $\boldsymbol{0}$ and a SE variance matrix with $\boldsymbol{V}_n = \sigma^2 (\boldsymbol{I} - \rho \boldsymbol{W}^*)^{-1} (\boldsymbol{I} - \rho \boldsymbol{W}^{*\prime})^{-1}$. We set $\sigma^2 = 1$. For the weight matrix $\boldsymbol{W}^*$, 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 $\boldsymbol{\beta}$ as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ where $\boldsymbol{\beta}_1 = \boldsymbol{1}_{p1}$ is a $p_1 \times 1$ vector of ones, and $\boldsymbol{\beta}_2 = (\Delta, \boldsymbol{0}_{p_2-1})$, $\boldsymbol{0}_{p_2-1}$ is a zero vector of dimension $(p_2 - 1 \times 1)$, and $\Delta = \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|$, where $\|A\|$ is the Euclidian 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 fitted the model in (8) using the spautolm function within the R-package spdep [16], obtain the values of all estimators considered in our study, and computed the simulated mean square error (SMSE) of each estimator as $SMSE(\hat{\boldsymbol{\beta}}_1^*) = \sum_{i=1}^{p_1} \left(\hat{\boldsymbol{\beta}}_{1i}^* - \boldsymbol{\beta}_{1i} \right)^2$. The simulated relative efficiency (SRE) of any estimator, say $\hat{\boldsymbol{\beta}}_1^\circ$, with respect to the MLE ($\hat{\boldsymbol{\beta}}_1$) is calculated as:

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

where $\hat{\boldsymbol{\beta}}_1^\circ$ is any of the estimators $\{\hat{\boldsymbol{\beta}}_1^{UR}, \hat{\boldsymbol{\beta}}_1^{RR}, \hat{\boldsymbol{\beta}}_1^{PTR}, \hat{\boldsymbol{\beta}}_1^{SR}, \hat{\boldsymbol{\beta}}_1^{PSR}\}$. It is evident that when the $SRE(\hat{\boldsymbol{\beta}}_1^\circ)$ 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 $\alpha = 0.05$ for testing the hypothesis in (20). No statistically significant change was seen while altering the spatial dependency parameter. Therefore, we have chosen to simply exhibit the graphs for $\rho = 0.90$. Figures 1–3 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($\hat{\boldsymbol{\beta}}_1^{UR}$) 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, $\hat{\boldsymbol{\beta}}_1^{UR}$ efficiency increases as expected.

(ii) The Ridge-type sub-model estimator ($\hat{\boldsymbol{\beta}}_1^{RR}$) 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 to 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 estimators SRE also increase.

(v)   The ridge-type positive shrinkage estimator ($\hat{\boldsymbol{\beta}}_1^{PSR}$) uniformly prevails over the competing estimators.
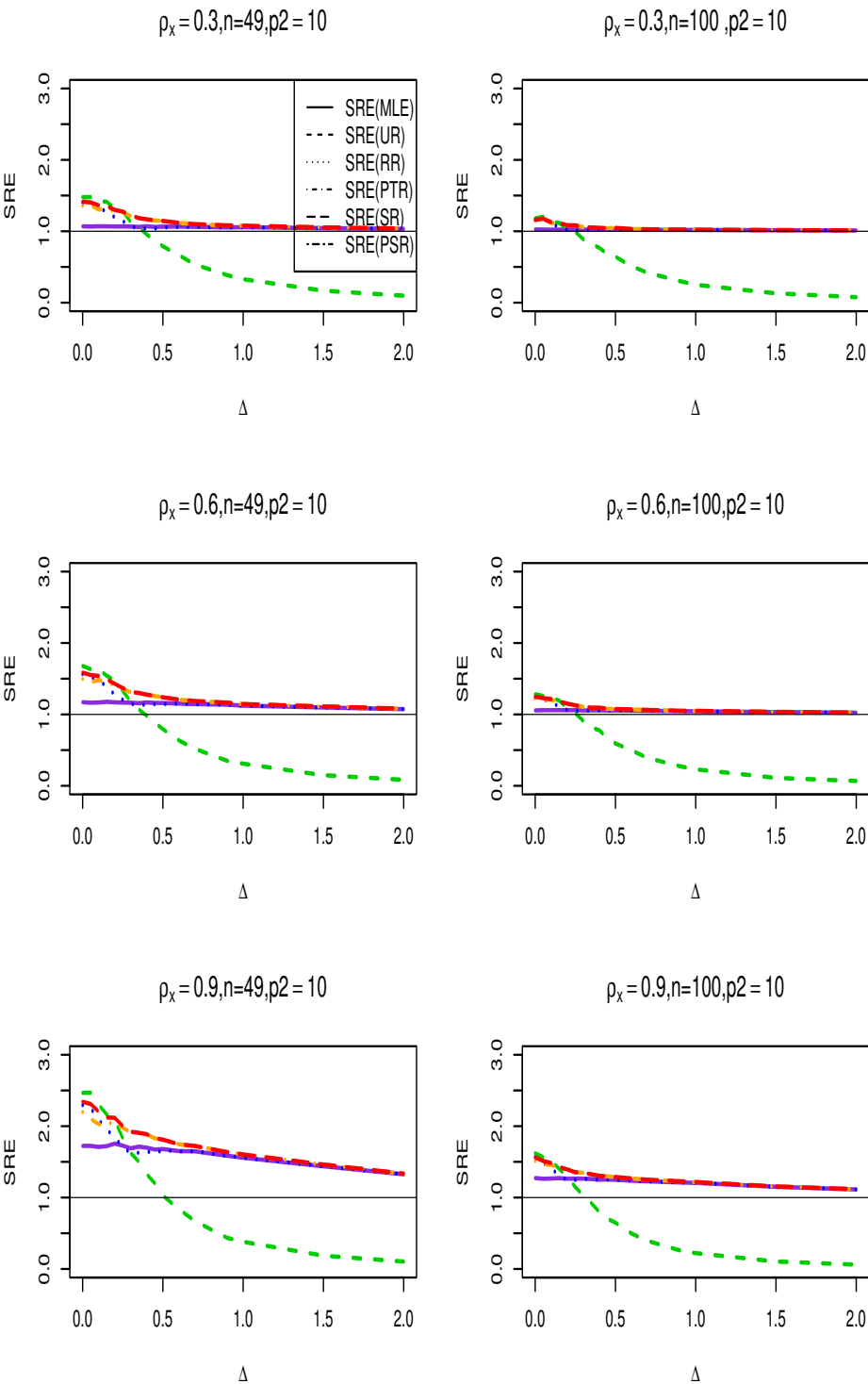
Figure 1. SRE of the suggested estimators with respect to the MLE ($\hat{\beta}_1$) for $n = 49, 100$, $\rho_x \in \{0.3, 0.6, 0.9\}$, $\rho = 0.90$, and $(p_1, p_2) = (5, 10)$.
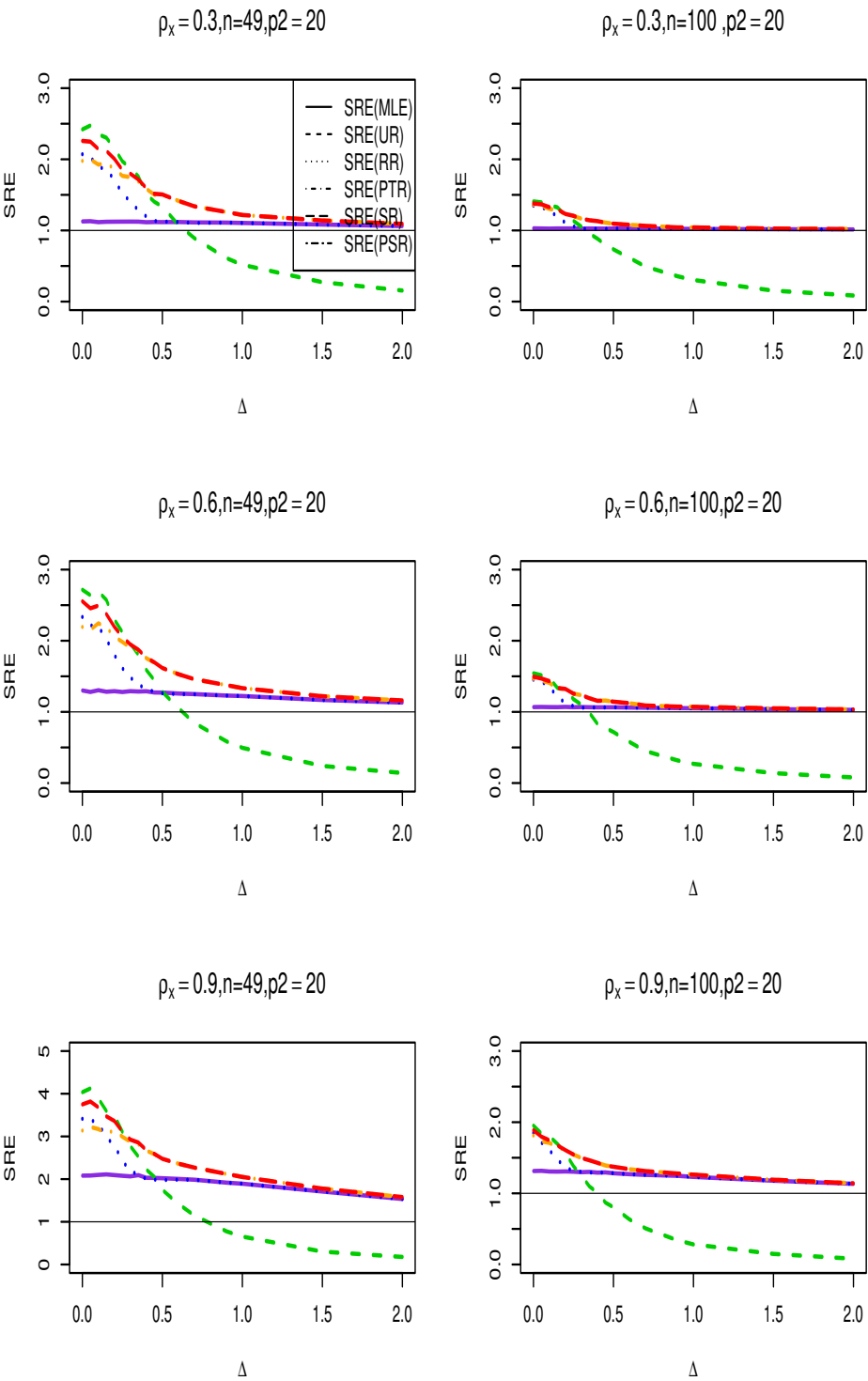
Figure 2. SRE of the suggested estimators with respect to the MLE ($\hat{\beta}_1$) for $n = 49, 100$, $\rho_x \in \{0.3, 0.6, 0.9\}$, $\rho = 0.90$, and $(p_1, p_2) = (5, 20)$.
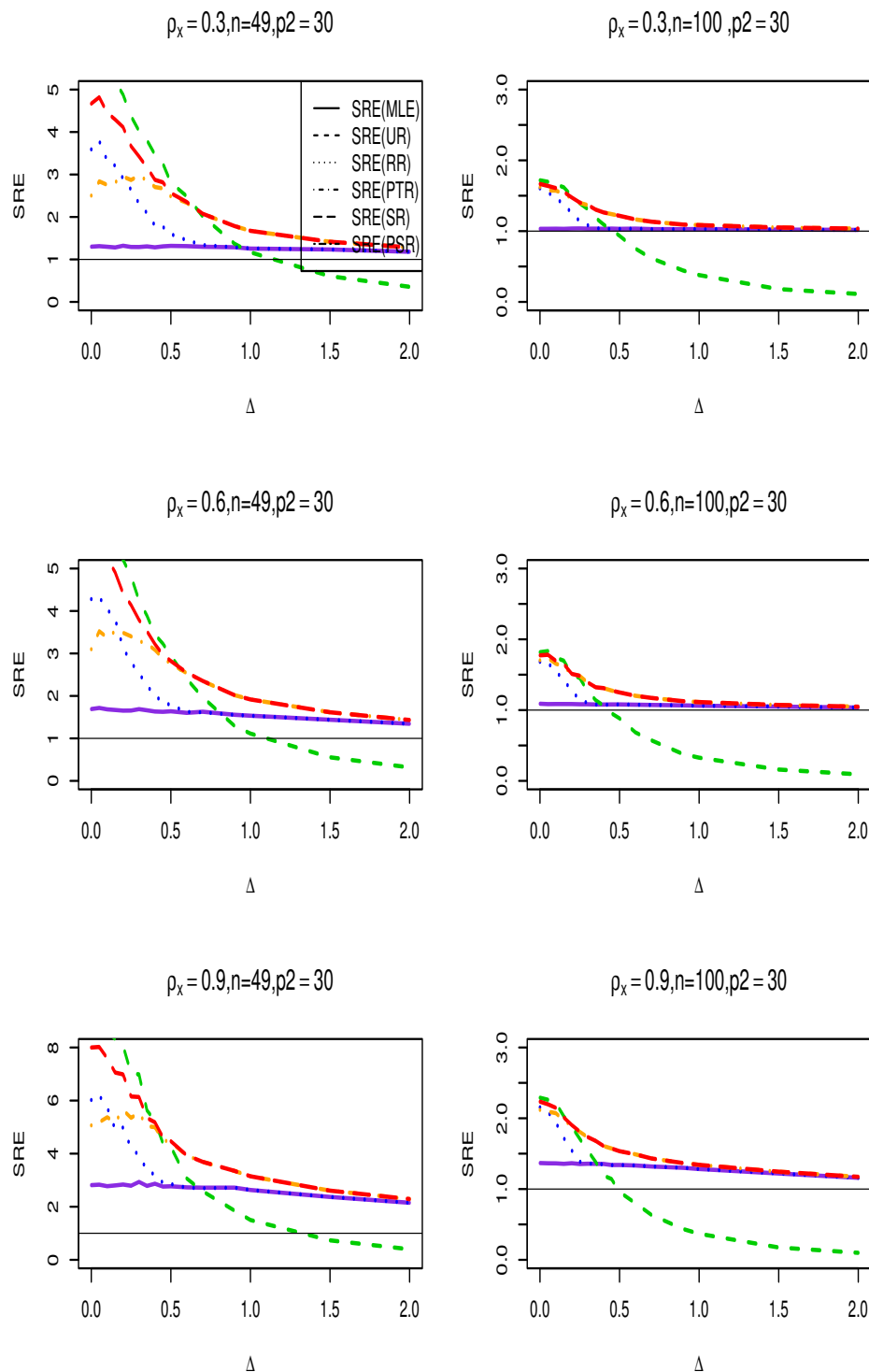
Figure 3. SRE of the suggested estimators with respect to the MLE ($\hat{\beta}_1$) for $n = 49, 100$, $\rho_x \in \{0.3, 0.6, 0.9\}$, $\rho = 0.90$, and $(p_1, p_2) = (5, 30)$.

6.2. Data Example

In 1970, [26] examined the use of housing market data for census tracts in the Boston Statistical Metropolitan Area. Their major objective was to establish a relationship between a set of (15) variables and the median cost of owner-occupied residences in Boston. [24] offered a corrected version of the

doi:10.20944/preprints202312.0644.v1

13 of 20

data set along with new spatial data. The data set is accessible through the R-Package `spdep`. 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 that is 1 if the tract borders the Charles River and 0 otherwise (`CHAS`), crime rate per capita (`CRIM`), weighted distance to main employment centers (`DIS`), nitrogen oxides concentration (parts per 10 million) per twon (`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`). [37] added the location of each tract in latitude (`LAT`), and longitude (`LON`) variables.

Assuming a SE model, we can predict the response variable `log(CMEDV)` using all available variables, that will be referred to as full SE model. For these data, a variety of selection techniques were used to determine the submodel. One submodel that was used by [5] is the model obtained by adaptive LASSO algorithm, which will be referred as our SE submodel. The two models are summarized in Table 1.

Table 1. Full and Submodel.

Selection Criterion	Model

Full	`log(CMEDV)    = log(LSTAT)+I(RM^2) + TAX`
	`+B +log(RAD) + CHAS  +CRIM + PTRATIO`
	`+ AGE+ LAT+ LON+log(RAD)+ I(NOX^2)`
	`+ log(DIS) + ZN+ INDUS`

Submodel	`log(CMEDV)    = log(LSTAT)+ I(RM^2)+ TAX+ B+ CRIM` `+ PTRATIO`

Figure 4 displays a coloured plot of the correlation coefficients for each variable. When a strong linear relationship is present, the color seems dark; when a weak linear relationship is present, the color shifts to light or may even vanish. 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 exists. The selected variables by adaptive LASSO algorithm appear to have a strong, medium and weak relationship with the response variable. Moreover, some variables exhibit collinearity, and this issue will show how really the ridge-type estimators will show up the high performance in comparison with the MLE estimator.
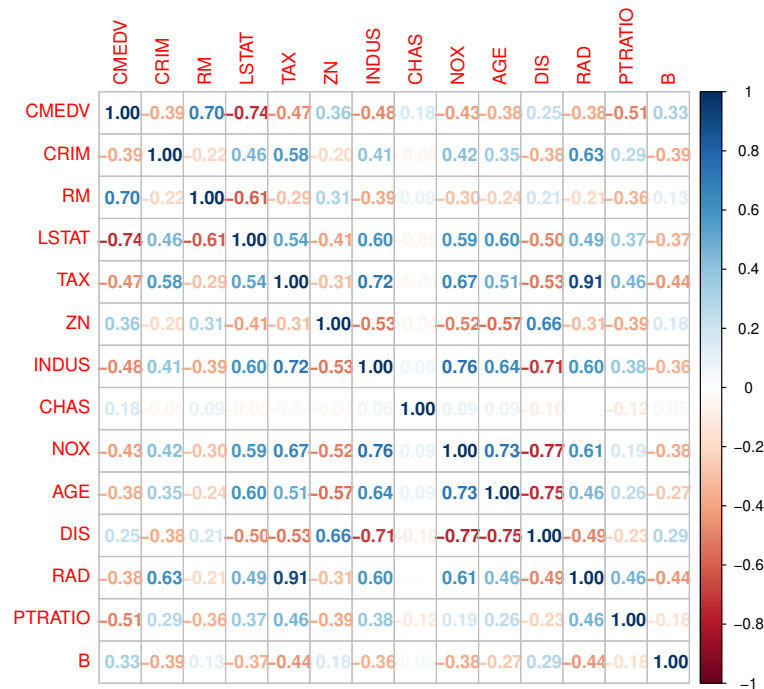
Figure 4. Correlation Matrix for the Boston Housing Data.

To assess the effectiveness of the suggested estimators we employed a bootstrapping technique suggested by [41], computing the mean squared prediction error (MSPE) for any estimator as follows:

1. Fit a SE full and sub models as appear in Table 1 using the `spautolm` function and get the MLEs of β_1, σ^2, the spatial dependance parameter ρ, the covariance matrix V_n.

2. As the columns of the two 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 followed Philip S. *et al* [17] to estimate the tuning ridge parameters for the two estimators $\hat{\beta}_1^{UR}$ and $\hat{\beta}_1^{RR}$ which are, respectively, given by:
$$k_f = \frac{\hat{\sigma}^2 tr(X1^T A_{X_2} X1)^{-1}}{(\hat{\beta}_1)^T (X1^T A_{X_2} X1)^{-1} \hat{\beta}_1}, \text{ and } k_r = \frac{\hat{\sigma}^2 tr(X1^T \hat{V}_n^{-1} X1)^{-1}}{(\hat{\beta}_1^S)^T (X1^T \hat{V}_n^{-1} X1)^{-1} \hat{\beta}_1^S}.$$

3. Use the Cholesky decomposition method in order to express the matrix $\hat{V}_n$ in a decomposed form as $\hat{V}_n = U U^T$, where U is an $(n \times n)$ lower triangular matrix.

4. Let $\hat{\epsilon} = U^{-1} \left(Y - X\hat{\beta} \right)$, where $\hat{\epsilon} = (\hat{\epsilon}_1, \hat{\epsilon}_2, \ldots, \hat{\epsilon}_n)$, and define the centered residual as $\epsilon_i^c = \hat{\epsilon}_i - \frac{1}{n} \sum_{j=1}^{n} \hat{\epsilon}_j$, then select with replacement a sample of size(n) form $(\epsilon_1^c, \epsilon_2^c, \ldots, \epsilon_n^c)$ to get $\epsilon^\star = (\epsilon_1^\star, \epsilon_2^\star, \ldots, \epsilon_n^\star)$.

5. Calculate the bootstrapping response value as $Y^\star = X\hat{\beta} + U^{-1}\epsilon^\star$, 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: $\hat{y}_{ki}^\star = X_1\hat{\beta}_1^* + \hat{\rho}^\star \sum_{j=1}^{n} W_{ij}^* \left(\hat{y}_{kj}^\star - X_j\hat{\beta}_1^* \right)$, 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 kth bootstrapping sample, calculate the square root of the mean square prediction error (MSPE) as

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

where B is the number of bootstrapping samples.

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

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

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

Table 2 summarizes the results of the relative efficiencies, where a number greater than one of the relative efficiency implies the superior performance of the estimator in the denominator.

Table 2. 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 table illustrates the better performance of the submodel 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 submodel that was selected. Also, the ridge positive shrinkage estimator performs better than the shrinkage one. Furthermore, all ridge-type estimators outperformed the MLE of $\beta 1$.

7. Conclusion

This paper discusses the pretest, shrinkage, and positive shrinkage ridge-type estimators of the parameter vector ($\boldsymbol{\beta}$) for SE model when there is a previous suspicion that certain coefficients are insignificant, and the muticolinearity presents between two or more regressor variables. To obtain the proposed set of estimators for the main effect vector of coefficients ($\boldsymbol{\beta}_1$), we test the hypothesis $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$. The proposed estimators were compared analytically via their asymptotic distributional quadratic risks, and numerically by 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 when the correlation among the regressor variables increase. Moreover, the performance of the ridge estimators always better than the MLE. In addition, the estimator ($\hat{\beta}_1^{RR}$) dominates all estimators under the null hypothesis $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ or when near the null space and delivers higher efficiency than the other estimators. However, the proposed positive shrinkage ridge estimators ($\hat{\beta}_1^{PSR}$) performs better than the MLE in all seniors. Further, we apply the set of estimators to a real data example, and used a bootstrapping technique to evaluate their performance based on the relative efficacy of square root of the mean squared prediction error.

Appendix A

In this section, we give proofs of the main results.

Proof of Lemma 1: For the proof, we follow the approach of Yuzbasi *et al* [44], with a slight modification. Let $W \sim N_p(0, \sigma^2 C)$ and define

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

where $\boldsymbol{u} = (u_1, \dots, u_p)^T$. Following [36],

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

with finite-dimensional convergence holding trivially. Also

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

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

$$
argmin V_n(\boldsymbol{u}) = \sqrt{n}(\hat{\boldsymbol{\beta}}^{\mathrm{RF}} - \boldsymbol{\beta})
$$

$$
\xrightarrow{D}
$$

$$
argmin V(\boldsymbol{u}) = \boldsymbol{C}^{-1}(\boldsymbol{W} - k_o \boldsymbol{\beta}) \sim N_p(-k_o \boldsymbol{C}^{-1} \boldsymbol{\beta}, \sigma^2 \boldsymbol{C}^{-1}).
$$

It concludes

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

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

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

Using part one of Lemma 1, $\mathrm{ADB}(\hat{\boldsymbol{\beta}}_1^{\mathrm{UR}}) = -\boldsymbol{\eta}_{11.2}$. Further using part three of Lemma 1 along with Theorem 1 in Appendix B of [29], we get

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

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

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

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

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

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

From Lemma 1, we have

$$
\begin{aligned}
\Gamma(\hat{\beta}_1^{\text{UR}}, M) &= tr\left\{M\left[cov(\hat{\beta}_1^{\text{UR}})\right]\right\} \\
&= tr\left\{M\left[\lim_{n\to\infty} nE(\hat{\beta}_1^{\text{UR}}-\beta_1)(\hat{\beta}_1^{\text{UR}}-\beta_1)^T\right]\right\} \\
&= tr\left\{M\left[\lim_{n\to\infty} E(\vartheta_n^{(1)}\vartheta_n^{(1)T})\right]\right\} \\
&= tr\left\{M\left[\lim_{n\to\infty} cov(\vartheta_n^{(1)}) + E(\vartheta_n^{(1)})E(\vartheta_n^{(1)})^T\right]\right\} \\
&= tr\left\{M\left[\sigma^2 C_{11.2}^{-1} + \eta_{11.2}\eta_{11.2}^T\right]\right\} \\
&= \eta_{11.2}^T M\eta_{11.2} + \sigma^2 tr(MC_{11.2}^{-1}).
\end{aligned}
$$

From Lemma 1

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

Using double expectation, parts three and six of Lemma 1, and Theorems 1 & 3 in Appendix B of [29], we get

$$
\begin{aligned}
\lim_{n\to\infty} E(\vartheta_n^{(3)}\vartheta_n^{(1)T}g(T_n)) &= \lim_{n\to\infty} E\left[E(\vartheta_n^{(3)}\vartheta_n^{(1)T}g(T_n))|\vartheta_n^{(3)}\right] \\
&= \lim_{n\to\infty} E\left[\vartheta_n^{(3)}E(\vartheta_n^{(1)T}g(T_n))|\vartheta_n^{(3)}\right] \\
&= \lim_{n\to\infty} E\left[\vartheta_n^{(3)}[-\eta_{11.2} + \vartheta_n^{(3)} - \delta]^T g(T_n)|\vartheta_n^{(3)}\right] \\
&= -\lim_{n\to\infty} E\left[\vartheta_n^{(3)}\eta_{11.2}^T g(T_n)\right] + \lim_{n\to\infty} E\left[\vartheta_n^{(3)}(\vartheta_n^{(3)} - \delta)^T g(T_n)\right] \\
&= -\lim_{n\to\infty} E\left[\vartheta_n^{(3)}g(T_n)\right]\eta_{11.2}^T + \lim_{n\to\infty} E\left[\vartheta_n^{(3)}\vartheta_n^{(3)T}g(T_n)\right] \\
&\quad - \lim_{n\to\infty} E\left[\vartheta_n^{(3)}g(T_n)\right]\delta^T \\
&= -\delta\eta_{11.2}^T E\left[g(\chi^2_{p_2+2}(\Delta))\right] + \sigma^2(C_{11.2}^{-1} - C_{11}^{-1})E\left[g(\chi^2_{p_2+2}(\Delta))\right] \\
&\quad + \delta\delta^T E\left[g(\chi^2_{p_2+4}(\Delta))\right] - \delta\delta^T E\left[g(\chi^2_{p_2+2}(\Delta))\right]
\end{aligned}
$$

Thus, it yields

$$
\lim_{n \to \infty} nE\left[(\hat{\boldsymbol{\beta}}_1^{\mathrm{UR}} - \boldsymbol{\beta})^T \boldsymbol{M}(\hat{\boldsymbol{\beta}}_1^{\mathrm{UR}} - \hat{\boldsymbol{\beta}}_1^{\mathrm{RR}})g(T_n)\right] = -\boldsymbol{\eta}_{11.2}^T \boldsymbol{M}\boldsymbol{\delta}
$$

$$
\begin{aligned}
&E\left[g(\chi_{p_2+2}^2(\Delta))\right] \\
&+\sigma^2 tr[\boldsymbol{M}(\boldsymbol{C}_{11.2}^{-1} - \boldsymbol{C}_{11}^{-1})] \\
&E\left[g(\chi_{p_2+2}^2(\Delta))\right] \\
&+\boldsymbol{\delta}^T \boldsymbol{M}\boldsymbol{\delta} E\left[g(\chi_{p_2+4}^2(\Delta))\right] \\
&- \boldsymbol{\delta}^T \boldsymbol{M}\boldsymbol{\delta} E\left[g(\chi_{p_2+2}^2(\Delta))\right]
\end{aligned}
$$

In a similar manner, we get

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

Gathering all required expressions, we finally have

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

The proof is complete using Table A1.

References

1. Ahmed SE.(2014). Penalty, shrinkage and Pretest Strategies Variable Selection and estimation. *Cham: Springer International Publishing* 2014.
2. Ahmed SE, Bahadir Y. (2016). Big Data Analytics: Integrating Penalty Strategies. *Big Data Analytics: Integrating Penalty Strategies.* 11(2), 105–115.
3. Al-Momani M.(2023). Liu-type pretest and shrinkage estimation for the conditional autoregressive model. *PLOS ONE. 18* (4) *doi:10.1371/journal.pone.0283339* 2023.
4. Al-Momani M, Dawod AB.(2022). Model selection and post selection to improve the estimation of the arch model. *Journal of Risk and Financial Management. 15* (4), 174.
5. Al-Momani M, Hussein AA, Ahmed SE.(2016). Penalty and related estimation strategies in the spatial error model. *Statistica Neerlandica. 71* (1), 4–30.

6. Al-Momani M, Riaz M, Saleh MF. (2022). Pretest and shrinkage estimation of the regression parameter vector of the marginal model with multinomial responses. *Statistical Papers. doi:10.1007/s00362-022-01372-2* 2022.

7. Arashi M, Kibria BMG, Norouzirad M, Nadarajah S.(2014). Improved preliminary test and Stein-rule Liu estimators for the ill-conditioned elliptical linear regression model. *Journal of Multivariate Analysis.* (126), 53–74.

8. Arashi M, Norouzirad M, Ahmed SE., Bahadir Y.(2018). Rank-based Liu Regression. *Computational Statistics. 33* (3),1525–1561.

9. Arashi M, Norouzirad M, Roozbeh M, Khan NM.(2021). A high-dimensional counterpart for the ridge estimator in multicollinear situations. *Mathematics. 9* (23),3057.

10. Arashi M, Roozbeh M.(2016). Some improved estimation strategies in high-dimensional semiparametric regression models with application to riboflavin production data. *Statistical Papers. 60* (3), 667–686.

11. Arashi M, Roozbeh M, Hamzah NA, Gasparini M.(2021). Ridge regression and its applications in genetic studies. *PLOS ONE. 16* (4) *doi:10.1371/journal.pone.0245376* 2023.

12. Bahadir Y, Ahmed SE, Ahmed F. (2023). Post-shrinkage strategies in statistical and machine learning for high dimensional data. *S.l.: CRC PRESS* 2023.

13. Bahadir Y, Ahmed SE, and Gungor M. (2017). Robust estimation approach for spatial error model. *REVSTAT-Statistical Journal. 15* (2),251–276.

14. Bahadir Y, Arashi M, Ahmed SE.(2020). Shrinkage estimation strategies in Generalised Ridge Regression Models: Low/high-dimension regime. *International Statistical Review. 88* (1),229–251.

15. Baltagi BH, Lesage JP, Pace RK.(2017). Spatial econometrics: Qualitative and limited dependent variables. *Bingley, UK: Emerald Group Publishing Limited.* 2017.

16. Bivand R. (2022). R packages for Analyzing Spatial Data: A comparative case study with Areal Data. *Geographical Analysis. 54* (3),488–518.

17. Boonstra PS, Mukherjee B, Taylor JM.(2015). A small-sample choice of the tuning parameter in ridge regression. *Statistica Sinica. doi:10.5705/ss.2013.284* 2015.

18. Cressie N. (1993). Statistics for Spatial Data. *Nashville, TN: John Wiley & Sons* 1993.

19. Cressie N, Wikle CK.(2011). Statistics for Spatio-Temporal Data. *Chichester, England: Wiley-Blackwell* 2011.

20. Dai X, Li E, Tian M.(2019). Quantile regression for varying coefficient spatial error models. *Communications in Statistics - Theory and Methods. 50* (10),2382–2397.

21. Ehsanes S A K M. (2006). Theory of preliminary test and Stein-type estimation with applications. *Hoboken (N.J.): Wiley-Interscience.* 2006.

22. Ehsanes S A K M, Arashi M, Golam K B M (2019). Tank-based methods for shrinkage and selection: With application to machine learning. *Hoboken (N.J.):John Wiley & amp; Sons, Inc.* 2019.

23. Ehsanes S A K M, Arashi M, Saleh R A, Norouzirad M. (2022). Theory of ridge regression estimation with applications. *Hoboken (N.J.):John Wiley & amp; Sons, Inc.* 2019.

24. Gilley OW, Pace RK.(1996). On the Harrison and Rubinfeld data. *Journal of Environmental Economics and Management. 31* (3),403–405.

25. Haining R. (2003). Spatial Data Analysis: Theory and Practice. *Cambridge: Cambridge University Press;* 2003.

26. Harrison D, Rubinfeld DL.(1978). Hedonic housing prices and the demand for Clean Air. *Journal of Environmental Economics and Management. 5* (1),81–102.

27. Higazi SF, Abdel-Hady DH, Al-Oulfi SA.(2013). Application of spatial regression models to income poverty ratios in Middle Delta contiguous counties in Egypt. *Pakistan Journal of Statistics and Operation Research. . 9* (1),93.

28. Hoerl AE, Kennard RW. (1970). A new Liu-type estimator in linear regression model. *Technometrics. 12* (1),55–67.

29. Judge, George G., and M. E. Bock. (1970). The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics. *North-Holland Pub. Co..* 1978.

30. Kejian L.(1993). A new class of blased estimate in linear regression. *Communications in Statistics - Theory and Methods. 22* (2),393–402.

31. Li Y, Yang H.(2010). A new Liu-type estimator in linear regression model. *Statistical Papers. 53* (2),427-437.

32. Lisawadi S, Ahmed SE, Reangsephet O.(2020). Post estimation and prediction strategies in negative binomial regression model. *International Journal of Modelling and Simulation. 41* (6),463–477.

33. Liu R, Yu C, Liu C, Jiang J, Xu J.(2018). Impacts of haze on housing prices: An empirical analysis based on data from Chengdu (China). *International Journal of Environmental Research and Public Health. 15* (6),1161.

34. Mardia KV, Marshall RJ. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika. 71* (1),135–146.

35. Nkurunziza S, Al-Momani M, Lin EY. (2016). Shrinkage and lasso strategies in high-dimensional heteroscedastic models. *Communications in Statistics - Theory and Methods. 45* (15),4454–4470.

36. Wenjiang Fu and Keith Knight. (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics. 28* (5), 1356–1378.

37. Pace R K, Gilley, O W. (1997). Using the Spatial Configuration of the Data to Improve Estimation. *Journal of Real Estate Finance and Economics.* (14),333–340.

38. Piscitelli A. (2019). Spatial Regression of Juvenile Delinquency: Revisiting Shaw and McKay. *International Journal of Criminal Justice Sciences. 14* (2),132–147.

39. Roozbeh M, Babaie–Kafaki S, Aminifard Z.(2021). Two penalized mixed–integer nonlinear programming approaches to tackle multicollinearity and outliers effects in linear regression models. *Journal of Industrial and Management Optimization. 17* (6),3475.

40. Roozbeh M, Babaie-Kafaki S, Manavi M.(2022). A heuristic algorithm to combat outliers and multicollinearity in regression model analysis. *Iranian Journal of Numerical Analysis and Optimization. 12* (1),173–186.

41. Solow AR. (1985). Bootstrapping correlated data. *Journal of the International Association for Mathematical Geology. 17* (7),769–775.

42. Waller LA, Gotway CA. (2004). Applied Spatial Statistics for Public Health Data. *Hoboken (N.J.):Wiley-Interscience.* 2004.

43. Yildirim V, Mert K Y. (2020). Robust estimation approach for spatial error model. *Journal of Statistical Computation and Simulation. 90* (9),1618–1638.

44. Yuzbasi, B., Arashi, M., and Ahmed, S. E. (2020). Shrinkage estimation strategies in generalized ridge regression models under low/high-dimension regime. *International Statistical Review. 88* (1), 229-251.