Journal of Multivariate Analysis 116 (2013) 35–44

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Feasible ridge estimator in partially linear models

M. Roozbeh [a,*], M. Arashi [b]

[a] Department of Mathematics, Statistics and Computer sciences, School of Sciences, Semnan University, P.O. Box 35131-19111, Semnan, Iran

[b] Faculty of Mathematics, Shahrood University of Technology, P.O. Box 316-3619995161, Shahrood, Iran

ARTICLE INFO

Article history:
Received 20 November 2011
Available online 8 December 2012

AMS subject classifications:
primary 62G08
62J05
secondary 62J07
62G20

Keywords:
Linear restrictions
Kernel smoothing
Multicollinearity
Feasible ridge estimator
Partial linear model

ABSTRACT

In a partial linear model, some non-stochastic linear restrictions are imposed under a multicollinearity setting. Semiparametric ridge and non-ridge type estimators, in a restricted manifold are defined. For practical use, it is assumed that the covariance matrix of the error term is unknown and thus feasible estimators are replaced and their asymptotic distributional properties are derived. Also, necessary and sufficient conditions, for the superiority of the ridge type estimator over its counterpart, for selecting the ridge parameter k are obtained. Lastly, a Monte Carlo simulation study is conducted to estimate the parametric and non-parametric parts. In this regard, kernel smoothing and cross validation methods for estimating the non-parametric function are used.

1. Introduction

Consider the partial linear model given by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + f(\boldsymbol{t}) + \boldsymbol{\epsilon}, \tag{1.1}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ is a $n \times p$ matrix, $\boldsymbol{x}_i's$ are known p-vectors for $i = 1, \ldots, n$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a vector of unknown parameters, $f(\boldsymbol{t}) = (f(t_1), \ldots, f(t_n))'$ is a vector of unknown nonparametric functions and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$ is the error term. We assume that in general, $f(\cdot)$ is an unknown function, the t's have bounded support, say the unit interval, and have been reordered so that $t_1 \le t_2 \le \cdots \le t_n$. All we know about $f(\cdot)$ is that it's first derivative is bounded by a constant, say L. This model is first considered by Engle et al. [3] to study the effect of weather on electricity demand, in which they assumed that the mean relationship between temperature and electricity usage was unknown while other related factors such as income and price were parameterized linearly. In our study, $\boldsymbol{\epsilon}$ is a n-vector of disturbances with the characteristics $E(\boldsymbol{\epsilon}) = \boldsymbol{0}$ and $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \boldsymbol{V}$, where $\boldsymbol{V}$ is a symmetric, positive definite unknown matrix.

For the main purposes of this paper we will employ the ridge regression concept that was proposed in the 1970's to combat the multicollinearity, in the partial linear model. The existence of multicollinearity may lead to wide confidence intervals for individual parameters or linear combination of the parameters and may produce estimates with wrong signs, etc. Most of the literature judges the performance of ridge regression estimators on the basis of the concentration of estimates around the true value of the parameter (see e.g. [15,13,5,8,20,19,16,10,11,2]).

The rest of the paper is organized as follows: In Section 2, the estimators under study are given, while their asymptotic biases and distribution risks are derived in Section 3. Section 4 includes some comparison result of the proposed estimators.

* Corresponding author.
E-mail address: m.roozbeh.stat@gmail.com (M. Roozbeh).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.11.006

The relevant selection of shrinking parameter is discussed in Section 5 and we end this approach by some numerical computations in Section 6. Finally, some concluding important results are stated in Section 7.

2. The proposed estimators

In this paper we confine ourselves to the partial kernel smoothing estimator of $\boldsymbol{\beta}$, which attains the usual parametric convergence rate $n^{1/2}$ without under smoothing the nonparametric component $f(\cdot)$ [14]. Assume that $(\boldsymbol{x}_i', t_i, y_i; i = 1, \ldots, n)$ satisfy model (1.1). Since $E(\varepsilon_i) = 0$, we have $f(t_i) = E(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})$ for $i = 1, \ldots, n$. Hence, if we know $\boldsymbol{\beta}$, a natural nonparametric estimator of $f(\cdot)$ is

$$\hat{f}(t, \boldsymbol{\beta}) = \sum_{i=1}^{n} W_{ni}(t)(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}), \tag{2.1}$$

where the positive weight functions $W_{ni}(\cdot)$ satisfy three conditions below:

(i) $\max_{1\le i\le n} \sum_{j=1}^{n} W_{ni}(t_j) = O(1)$,
(ii) $\max_{1\le i,j\le n} W_{ni}(t_j) = O(n^{-2/3})$,
(iii) $\max_{1\le i\le n} \sum_{j=1}^{n} W_{ni}(t_j) I(|t_i - t_j| > c_n) = O(d_n)$,

where I is the indicator function, c_n satisfies $\limsup_{n\to\infty} nc_n^3 < \infty$, and d_n satisfies $\lim sup_{n\to\infty} nd_n^3 < \infty$.

The above assumptions guarantee the existence of $\hat{f}(t, \boldsymbol{\beta})$ at the optimal convergence rate $n^{-4/5}$, in partial linear models with probability one. See [9] for more details.

To estimate $\boldsymbol{\beta}$, we use the weighted least squares estimator given by

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, SS(\boldsymbol{\beta}) = \boldsymbol{C}^{-1}\widetilde{\boldsymbol{X}}'\boldsymbol{V}^{-1}\widetilde{\boldsymbol{y}}, \quad \boldsymbol{C} = \widetilde{\boldsymbol{X}}'\boldsymbol{V}^{-1}\widetilde{\boldsymbol{X}} \tag{2.2}$$

where $SS(\boldsymbol{\beta}) = (\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta})'\boldsymbol{V}^{-1}(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta})$, $\widetilde{\boldsymbol{y}} = (\widetilde{y}_1, \ldots, \widetilde{y}_n)'$, $\widetilde{\boldsymbol{X}} = (\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_n)'$, $\widetilde{y}_i = y_i - \sum_{j=1}^{n} W_{nj}(t_i)y_j$ and $\widetilde{\boldsymbol{x}}_i = \boldsymbol{x}_i - \sum_{j=1}^{n} W_{nj}(t_i)\boldsymbol{x}_j$ for $i = 1, \ldots, n$.

It is observed from (2.2) that the properties of the $\hat{\boldsymbol{\beta}}$ depend heavily on the characteristics of the information matrix $\boldsymbol{C}$. If the $\boldsymbol{C}$ matrix is ill-conditioned (near dependency among various columns of $\boldsymbol{C}$), then the $\hat{\boldsymbol{\beta}}$ produces unduly large sampling variances. Moreover, some of the regression coefficients may be statistically insignificant with wrong signs and meaningful statistical inferences become difficult for the researcher. As a remedy, Hoerl and Kennard [7] suggested to use the following estimator instead of $\hat{\boldsymbol{\beta}}$, say ridge weighted least squares estimator given by

$$\hat{\boldsymbol{\beta}}(k) = \boldsymbol{T}_k\hat{\boldsymbol{\beta}}, \quad \boldsymbol{T}_k = (k\boldsymbol{C}^{-1} + \boldsymbol{I}_p)^{-1}, \tag{2.3}$$

where $k \ge 0$ is the shrinking parameter.

Now consider the exact linear non-stochastic constraint $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$, for a given $m \times p$ matrix $\boldsymbol{R}$ with rank $m < p$ and a given $m \times 1$ vector $\boldsymbol{r}$. The full row rank assumption is chosen for convenience and can be justified by the fact that every consistent linear equation can be transformed into an equivalent equation with a coefficient matrix of full row rank. Subject to the imposed linear restriction, the restricted weighted least squares estimator is given by

$$\hat{\boldsymbol{\beta}}_r = \hat{\boldsymbol{\beta}} - \boldsymbol{C}^{-1}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{R}')^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}). \tag{2.4}$$

So, the restricted ridge weighted least squares estimator can be written as

$$\hat{\boldsymbol{\beta}}_r(k) = \hat{\boldsymbol{\beta}}(k) - \boldsymbol{C}_k^{-1}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R})^{-1}[\boldsymbol{R}\hat{\boldsymbol{\beta}}(k) - \boldsymbol{r}], \quad \boldsymbol{C}_k = \boldsymbol{C} + k\boldsymbol{I}_p. \tag{2.5}$$

Since the covariance matrix $\boldsymbol{V}$ is unknown, as it usually is, the $\hat{\boldsymbol{\beta}}$ in (2.2) is non-operational as it depends upon the unknown covariance matrix of the errors and we must define the estimators $\hat{\boldsymbol{\beta}}_r(k)$ and $\hat{\boldsymbol{\beta}}_r$ based on two stage feasible weighted least squares estimator by replacing the unknown $\boldsymbol{V}$ with a consistent estimator of it as follows:

$$\hat{\boldsymbol{\beta}}^F = \boldsymbol{C}^{*-1}\widetilde{\boldsymbol{X}}'\boldsymbol{S}^{-1}\widetilde{\boldsymbol{y}}, \tag{2.6}$$

where $\boldsymbol{C}^* = \widetilde{\boldsymbol{X}}'\boldsymbol{S}^{-1}\widetilde{\boldsymbol{X}}$, $\boldsymbol{S} = \frac{1}{n-(p-m)}(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{b})(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\boldsymbol{b})'$ and $\boldsymbol{b}$ is ordinary least squares estimator, $(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})^{-1}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{y}}$. As it has been shown in [18], $\hat{\boldsymbol{\beta}}^F = \hat{\boldsymbol{\beta}} + O(n^{-1})$ and so, $\sqrt{n}(\hat{\boldsymbol{\beta}}^F - \boldsymbol{\beta})$ and $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ have the same asymptotic normal distribution. So, when the covariance matrix is unknown, the feasible estimators are defined as follows:

$$\hat{\boldsymbol{\beta}}_r^F = \hat{\boldsymbol{\beta}}^F - \boldsymbol{C}^{*-1}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{C}^{*-1}\boldsymbol{R}')^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}^F - \boldsymbol{r}), \tag{2.7}$$

$$\hat{\boldsymbol{\beta}}_r^F(k) = \hat{\boldsymbol{\beta}}^F(k) - \boldsymbol{C}_k^{*-1}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{C}_k^{*-1}\boldsymbol{R}')^{-1}[\boldsymbol{R}\hat{\boldsymbol{\beta}}^F(k) - \boldsymbol{r}], \tag{2.8}$$

where $\hat{\boldsymbol{\beta}}^F(k) = \boldsymbol{C}_k^{*-1}\widetilde{\boldsymbol{X}}\boldsymbol{S}^{-1}\widetilde{\boldsymbol{y}}$ and $\boldsymbol{C}_k^* = \boldsymbol{C}^* + k\boldsymbol{I}_p$.

Then it is easy to see that the proposed estimators are restricted with respect to $\boldsymbol{R\beta} = \boldsymbol{r}$. It is also clear that for $k = 0$, we get $\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}(0) = \hat{\boldsymbol{\beta}}_{\boldsymbol{r}}$, $\hat{\boldsymbol{\beta}}_r^F(\mathbf{0}) = \hat{\boldsymbol{\beta}}_r^F$.

We close this section with the following assumptions required to derive the main results. These assumptions are quite general and it can be shown that under these assumptions the proposed estimators are asymptotically normal estimators of $\boldsymbol{\beta}$ at the rate of $\sqrt{n}$. See [6] for more details.

Assumption 1. There exist bounded functions $h_s(\cdot)$ over $[0, 1]$, $s = 1, \dots, p$, such that

$$x_{is} = h_s(t_i) + u_{is}, \quad i = 1, \dots, n,\ s = 1, \dots, p, \tag{2.9}$$

where $\mathbf{u}_i = (u_{i1}, \dots, u_{ip})$ are real vectors satisfying

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} u_{ij}u_{il}}{n} = z_{lj}, \quad \text{for } l = 1, \dots, p,\ j = 1, \dots, p \tag{2.10}$$

and the $p \times p$ matrix $\boldsymbol{Z} = (z_{lj})$ is nonsingular.

Assumption 2. The functions $f(\cdot)$ and $h_s(\cdot)$ satisfy the Lipschitz condition of order 1 on $[0, 1]$ for $s = 1, \dots, p$.

Assumption 3. For any permutation $(j_1, \dots, j_n)$ of $(1, \dots, n)$ and $m = 1, \dots, p$, as $n \to \infty$,

$$\frac{1}{a_n} \max_{1\le k\le n} \left| \sum_{i=1}^{k} u_{j_i m} \right| < \infty, \tag{2.11}$$

where $a_n = n^{1/2} \log n$ and $\boldsymbol{Z}$ is a positive definite matrix.

3. Asymptotic characteristics

In this section, we derive expressions for asymptotic distributional biases (***adb***s) and asymptotic distributional risks (adrs) of the estimators considered in Section 2. The objective is to estimate the unknown parameter vector $\boldsymbol{\beta}$ by an estimator $\hat{\boldsymbol{\beta}}$ when the performance is evaluated by the squared error loss. To study the asymptotic quadratic risk of $\hat{\boldsymbol{\beta}}$, we define a quadratic loss function using a positive definite and symmetric matrix $\boldsymbol{Q}$, by

$$L(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}) = n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{Q}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \tag{3.1}$$

where $\hat{\boldsymbol{\beta}}$ can be any estimator of $\boldsymbol{\beta}$.

The entire vectorial parameter space is subjected to lie in the following null hypothesis

$$H_o : \boldsymbol{R\beta} = \boldsymbol{r}. \tag{3.2}$$

Further, following Saleh [12], consider the following regularity conditions hold

(i) $\max_{1\le i\le n} \widetilde{\boldsymbol{x}}_i'(\widetilde{\boldsymbol{X}}'\boldsymbol{V}^{-1}\widetilde{\boldsymbol{X}})^{-1}\widetilde{\boldsymbol{x}}_i \to 0$ as $n \to \infty$, where $\widetilde{\boldsymbol{x}}_i'$ is the i^th row of $\widetilde{\boldsymbol{X}}$.
(ii) $\lim_{n\to\infty} n^{-1}\boldsymbol{C} = \mathbf{B}_{p\times p}$, for symmetric and finite matrix $\mathbf{B}$.

Consequently, the asymptotic distributional bias (***adb***) of the estimator can be evaluated through

$$\boldsymbol{adb}(\hat{\boldsymbol{\beta}}) = \lim_{n\to\infty} E\Big(\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\Big). \tag{3.3}$$

Assume that the asymptotic distribution function of $\hat{\boldsymbol{\beta}}$ exists and is given by

$$F(x) = \lim_{n\to\infty} P\Big(\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \le x\Big), \tag{3.4}$$

where $F(x)$ is non-degenerate. Then the asymptotic distributional risk (adr) of $\hat{\boldsymbol{\beta}}$ is also defined as

$$\mathrm{adr}_{\boldsymbol{Q}}(\hat{\boldsymbol{\beta}}) = \mathrm{tr}\left(\boldsymbol{Q}\int_{\Re^p} (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}})(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}})'dF(\boldsymbol{x})\right) = \mathrm{tr}\Big(\boldsymbol{Q}^{1/2}\boldsymbol{V}_x\boldsymbol{Q}^{1/2}\Big), \tag{3.5}$$

where $\boldsymbol{V}_{\boldsymbol{x}}$ is the dispersion matrix for the distribution $F(x)$.

Now for calculating the ***adb***s and adrs of the proposed estimators in the last section and deriving a necessary and sufficient condition for the superiority of the $\hat{\boldsymbol{\beta}}_r^F(k)$ over the $\hat{\boldsymbol{\beta}}_r^F$, first we obtain a new formula for $\hat{\boldsymbol{\beta}}_r(k)$ that simplifies the calculation of ***adb*** and adr as follows:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_r(k) &= \hat{\boldsymbol{\beta}}(k) - \boldsymbol{C}_k^{-1}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}')^{-1}[\boldsymbol{R}\hat{\boldsymbol{\beta}}(k) - \boldsymbol{r}]\\ &= \boldsymbol{C}_k^{-1}\boldsymbol{C}_k\hat{\boldsymbol{\beta}}(k) - \boldsymbol{C}_k^{-1}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}')^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{C}_k\hat{\boldsymbol{\beta}}(k) + \boldsymbol{C}_k^{-1}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}')^{-1}\boldsymbol{R}\boldsymbol{\beta}_0 - \boldsymbol{C}_k^{-1}\boldsymbol{C}_k\boldsymbol{\beta}_0 + \boldsymbol{\beta}_0\\ &= \boldsymbol{M}_k\widetilde{\boldsymbol{X}}'\boldsymbol{V}^{-1}\widetilde{\boldsymbol{y}} - \boldsymbol{M}_k\boldsymbol{C}_k\boldsymbol{\beta}_0 + \boldsymbol{\beta}_0,\end{aligned} \tag{3.6}$$

where, $\boldsymbol{\beta}_0 = \boldsymbol{R}'(\boldsymbol{R}\boldsymbol{R}')^{-1}\boldsymbol{r}$ and $\boldsymbol{M}_k = \boldsymbol{C}_k^{-1} - \boldsymbol{C}_k^{-1}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{C}_k^{-1}\boldsymbol{R}')^{-1}\boldsymbol{R}\boldsymbol{C}_k^{-1}$.

Now, we can calculate the bias and covariance matrix of $\hat{\boldsymbol{\beta}}_r(k)$ by using Eq. (3.6) as follows:

$$\begin{aligned}E[\hat{\boldsymbol{\beta}}_r(k) - \boldsymbol{\beta}] &= \boldsymbol{M}_k\boldsymbol{C}\boldsymbol{\beta} - \boldsymbol{M}_k\boldsymbol{C}_k\boldsymbol{\beta}_0 + \boldsymbol{\beta}_0 - \boldsymbol{\beta}\\ &= \boldsymbol{M}_k\boldsymbol{C}_k(\boldsymbol{\beta} - \boldsymbol{\beta}_0) - k\boldsymbol{M}_k\boldsymbol{\beta} + \boldsymbol{\beta}_0 - \boldsymbol{\beta}\\ &= \boldsymbol{\beta} - \boldsymbol{\beta}_0 - k\boldsymbol{M}_k\boldsymbol{\beta} + \boldsymbol{\beta}_0 - \boldsymbol{\beta}\\ &= -k\boldsymbol{M}_k\boldsymbol{\beta},\end{aligned} \tag{3.7}$$

$$\mathrm{Cov}[\hat{\boldsymbol{\beta}}_r(k)] = \boldsymbol{M}_k\boldsymbol{C}\boldsymbol{M}_k. \tag{3.8}$$

Thus we get

$$\begin{aligned}\boldsymbol{adb}[\hat{\boldsymbol{\beta}}_r(k)] &= \lim_{n\to\infty} E\Big(\sqrt{n}(\hat{\boldsymbol{\beta}}_r(k) - \boldsymbol{\beta})\Big)\\ &= -k\sqrt{n}\boldsymbol{M}_k\boldsymbol{\beta} + o(1),\end{aligned}$$

$$\mathrm{adr}_{\boldsymbol{Q}}[\hat{\boldsymbol{\beta}}_r(k)] = n\,\mathrm{tr}\Big(\boldsymbol{Q}^{1/2}\boldsymbol{M}_k\boldsymbol{C}\boldsymbol{M}_k\boldsymbol{Q}^{1/2}\Big) + o(1). \tag{3.9}$$

Using assumption (ii) from page 6, further straightforward algebra shows that

$$\begin{aligned}\boldsymbol{adb}[\hat{\boldsymbol{\beta}}_r(k)] &= \lim_{n\to\infty} -k\sqrt{n}\boldsymbol{M}_k\boldsymbol{\beta}\\ &= \lim_{n\to\infty} -k(\sqrt{n})^{-1}\big[\mathbf{B}^{-1} - \mathbf{B}^{-1}\boldsymbol{R}'(\boldsymbol{R}\mathbf{B}^{-1}\boldsymbol{R}')^{-1}\boldsymbol{R}\mathbf{B}^{-1}\big]\boldsymbol{\beta}\\ &= \mathbf{0},\end{aligned}$$

$$\begin{aligned}\mathrm{adr}_{\boldsymbol{Q}}[\hat{\boldsymbol{\beta}}_r(k)] &= \lim_{n\to\infty}\mathrm{tr}\Big(\boldsymbol{n}\boldsymbol{Q}^{\mathbf{1/2}}\boldsymbol{M}_k\boldsymbol{C}\boldsymbol{M}_k\boldsymbol{Q}^{\mathbf{1/2}}\Big)\\ &= \lim_{n\to\infty}\mathrm{tr}\Big(\boldsymbol{Q}^{\mathbf{1/2}}(n\boldsymbol{M}_k)(n^{-1}\boldsymbol{C})(n\boldsymbol{M}_k)\boldsymbol{Q}^{\mathbf{1/2}}\Big)\\ &= \mathrm{tr}\Big(\boldsymbol{Q}^{1/2}\big[\mathbf{B}^{-1} - \mathbf{B}^{-1}\boldsymbol{R}'(\boldsymbol{R}\mathbf{B}^{-1}\boldsymbol{R}')^{-1}\boldsymbol{R}\mathbf{B}^{-1}\big]\mathbf{B}\big[\mathbf{B}^{-1} - \mathbf{B}^{-1}\boldsymbol{R}'(\boldsymbol{R}\mathbf{B}^{-1}\boldsymbol{R}')^{-1}\boldsymbol{R}\mathbf{B}^{-1}\big]\boldsymbol{Q}^{1/2}\Big)\\ &= \mathrm{tr}\Big(\boldsymbol{Q}^{1/2}\big[\mathbf{B}^{-1} - \mathbf{B}^{-1}\boldsymbol{R}'(\boldsymbol{R}\mathbf{B}^{-1}\boldsymbol{R}')^{-1}\boldsymbol{R}\mathbf{B}^{-1}\big]\boldsymbol{Q}^{1/2}\Big).\end{aligned}$$

So, $\hat{\boldsymbol{\beta}}_r(k)$ is an asymptotically unbiased estimator of $\boldsymbol{\beta}$ with stabilized asymptotic dispersion matrix $\boldsymbol{Q}^{1/2}\big[\mathbf{B}^{-1} - \mathbf{B}^{-1}\boldsymbol{R}'(\boldsymbol{R}\mathbf{B}^{-1}\boldsymbol{R}')^{-1}\boldsymbol{R}\mathbf{B}^{-1}\big]\boldsymbol{Q}^{1/2}$.

By the fact that $\boldsymbol{C} = \boldsymbol{C}_k - k\boldsymbol{I}_p$, it can be obtained $\boldsymbol{M}_k\boldsymbol{C}\boldsymbol{M}_k = \boldsymbol{M}_k - k\boldsymbol{M}_k^2$. Thus the amse function of $\hat{\boldsymbol{\beta}}_r(k)$ is

$$\begin{aligned}amse_{\boldsymbol{Q}}[\hat{\boldsymbol{\beta}}_r(k), \boldsymbol{\beta}] &= \boldsymbol{adb}'[\hat{\boldsymbol{\beta}}_r(k)]\boldsymbol{adb}[\hat{\boldsymbol{\beta}}_r(k)] + \mathrm{adr}_{\boldsymbol{Q}}[\hat{\boldsymbol{\beta}}_r(k)]\\ &= nk^2\boldsymbol{\beta}'\boldsymbol{M}_k^2\boldsymbol{\beta} + n\,\mathrm{tr}\Big(\boldsymbol{Q}^{1/2}(\boldsymbol{M}_k - k\boldsymbol{M}_k^2)\boldsymbol{Q}^{1/2}\Big) + o(1).\end{aligned} \tag{3.10}$$

As it has been shown in [18], $\hat{\boldsymbol{\beta}}_r^F$ has asymptotic normal distribution with mean $\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{C}^{*^{-1}} + o(n^{-1})$, So, the amse functions of the feasible form of proposed estimators are

$$amse_{\boldsymbol{Q}}[\hat{\boldsymbol{\beta}}_r^F(k), \boldsymbol{\beta}] = nk^2\boldsymbol{\beta}'\boldsymbol{M}_k^{*^2}\boldsymbol{\beta} + n\,\mathrm{tr}\Big(\boldsymbol{Q}^{1/2}(\boldsymbol{M}_k^* - k\boldsymbol{M}_k^{*^2})\boldsymbol{Q}^{1/2}\Big) + o(n^{-1}), \tag{3.11}$$

$$amse_{\boldsymbol{Q}}(\hat{\boldsymbol{\beta}}_r^F, \boldsymbol{\beta}) = n\,\mathrm{tr}\Big(\boldsymbol{Q}^{1/2}\boldsymbol{M}_0^*\boldsymbol{Q}^{1/2}\Big) + o(n^{-1}), \tag{3.12}$$

where $\boldsymbol{M}_k^* = \boldsymbol{C}_k^{*^{-1}} - \boldsymbol{C}_k^{*^{-1}}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{C}_k^{*^{-1}}\boldsymbol{R}')^{-1}\boldsymbol{R}\boldsymbol{C}_k^{*^{-1}}$.

Lemma 3.1. *The matrix $\boldsymbol{M}_k^*$ can be written as*

$$\boldsymbol{M}_k^* = \Big(\boldsymbol{P}\boldsymbol{C}_k^*\boldsymbol{P}\Big)^+ = \boldsymbol{H}\begin{pmatrix}(\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{H}',$$

where $\boldsymbol{P} = \boldsymbol{I}_p - \boldsymbol{R}'(\boldsymbol{R}\boldsymbol{R}')^{-1}\boldsymbol{R}$, $\boldsymbol{\Lambda}$ *is a diagonal matrix containing the* $p - m$ *nonzero eigenvalues of* $\boldsymbol{P}\boldsymbol{C}^*\boldsymbol{P}$ *on its diagonal, and* $\boldsymbol{H}$ *is an orthogonal matrix. Note that by a + superscript we denote the unique Moore–Penrose inverse of a matrix. Moreover,* $\frac{\partial \boldsymbol{M}_k^*}{\partial k} = -\boldsymbol{M}_k^{*^2}$.

Proof. The first identity $\boldsymbol{M}_k^* = \left(\boldsymbol{P}\boldsymbol{C}_k^*\boldsymbol{P}\right)^+$ is a special case of a more general result by Baksalary et al. [1, Theorem 3]. For the second identity, consider the matrix

$$\boldsymbol{P}\boldsymbol{C}_k^*\boldsymbol{P} = \boldsymbol{P}\boldsymbol{C}^*\boldsymbol{P} + k\boldsymbol{P}.$$

Since $\boldsymbol{P}\boldsymbol{C}^*\boldsymbol{P}$ and $\boldsymbol{P}$ are two commuting symmetric matrices with the same range and with rank $p - m$, they can be diagonalized by the same orthogonal matrix in such a way that

$$\boldsymbol{P}\boldsymbol{C}^*\boldsymbol{P} = \boldsymbol{H}\begin{pmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{H}', \quad \boldsymbol{P} = \boldsymbol{H}\begin{pmatrix} \boldsymbol{I}_{p-m} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{H}',$$

where the diagonal elements of $\boldsymbol{\Lambda}$ are the $p-m$ positive eigenvalues of $\boldsymbol{P}\boldsymbol{C}^*\boldsymbol{P}$. From this, the second asserted identity follows immediately.

For the second assertion, from the formula of matrix differential we have

$$\frac{\partial \boldsymbol{C}_k^{*^{-1}}}{\partial k} = -\boldsymbol{C}_k^{*^{-1}} \frac{\partial \boldsymbol{C}_k^{*^{-1}}}{\partial k} \boldsymbol{C}_k^{*^{-1}} = -\boldsymbol{C}_k^{*^{-2}}.$$

So we have

$$\begin{aligned}
\frac{\partial \boldsymbol{M}_k^*}{\partial k} &= -\boldsymbol{C}_k^{*^{-2}} - \Big(-\boldsymbol{C}_k^{*^{-2}}\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{C}_k^{*^{-1}}\boldsymbol{R}']^{-1}\boldsymbol{R}\boldsymbol{C}_k^{*^{-1}} + \boldsymbol{C}_k^{*^{-1}}\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{C}_k^{*^{-1}}\boldsymbol{R}']^{-1}\boldsymbol{R}\boldsymbol{C}_k^{*^{-2}}\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{C}_k^{*^{-1}}\boldsymbol{R}']^{-1}\boldsymbol{R}\boldsymbol{C}_k^{*^{-1}} \\
&\quad - \boldsymbol{C}_k^{*^{-1}}\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{C}_k^{*^{-1}}\boldsymbol{R}']^{-1}\boldsymbol{R}\boldsymbol{C}_k^{*^{-2}}\Big) \\
&= \left(\boldsymbol{C}_k^{*^{-1}} - \boldsymbol{C}_k^{*^{-1}}\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{C}_k^{*^{-1}}\boldsymbol{R}']^{-1}\boldsymbol{R}\boldsymbol{C}_k^{*^{-1}}\right)^2 \\
&= -\boldsymbol{M}_k^{*^2}. \quad \square
\end{aligned}$$

4. Superiority conditions

In this section, we provide necessary and sufficient conditions for amse-superiority of the $\hat{\boldsymbol{\beta}}_r^F(k)$ over $\hat{\boldsymbol{\beta}}_r^F$.

From (3.10) and (3.11), the difference $\delta = amse_{\boldsymbol{Q}}(\hat{\boldsymbol{\beta}}_r^F, \boldsymbol{\beta}) - amse_{\boldsymbol{Q}}[\hat{\boldsymbol{\beta}}_r^F(k), \boldsymbol{\beta}]$ is given by

$$\delta = n\mathrm{tr}\Big(\boldsymbol{Q}^{1/2}\boldsymbol{M}_0^*\boldsymbol{Q}^{1/2} + \boldsymbol{Q}^{1/2}(k\boldsymbol{M}_k^{*^2} - \boldsymbol{M}_k^*)\boldsymbol{Q}^{1/2}\Big) - nk^2\boldsymbol{\beta}'\boldsymbol{M}_k^{*^2}\boldsymbol{\beta}. \tag{4.1}$$

Lemma 4.1. *The* $\boldsymbol{ADR_Q}[\hat{\boldsymbol{\beta}}_r^F(k)]$ *is consistently smaller than* $\boldsymbol{ADR_Q}(\hat{\boldsymbol{\beta}}_r^F)$*, that is to say the following inequality always holds for arbitrary* $k > 0$

$$\tilde{\boldsymbol{\Delta}} = \boldsymbol{ADR_Q}(\hat{\boldsymbol{\beta}}_r^F) - \boldsymbol{ADR_Q}[\hat{\boldsymbol{\beta}}_r^F(k)] > 0,$$

where $\boldsymbol{ADR_Q}[\hat{\boldsymbol{\beta}}_r^F(k)] = n\boldsymbol{Q}^{1/2}(\boldsymbol{M}_k^* - k\boldsymbol{M}_k^{*^2})\boldsymbol{Q}^{1/2} + o(n^{-1})$ *is the matrix form of* $\mathrm{adr}_{\boldsymbol{Q}}[\hat{\boldsymbol{\beta}}_r^F(k)]$*. Moreover,* $\tilde{\boldsymbol{\Delta}}$ *is monotonously increased with respect to k.*

Proof. According to Lemma 3.1, we have

$$\begin{aligned}
\tilde{\boldsymbol{\Delta}} &= n\boldsymbol{Q}^{1/2}\boldsymbol{M}_0^*\boldsymbol{Q}^{1/2} - \boldsymbol{Q}^{1/2}\boldsymbol{M}_k^*\boldsymbol{Q}^{1/2} + k\boldsymbol{Q}^{1/2}\boldsymbol{M}_k^{*^2}\boldsymbol{Q}^{1/2} \\
&= n\boldsymbol{Q}^{1/2}\boldsymbol{H}\begin{pmatrix} \boldsymbol{\Lambda}^{-1} - (\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-1} + k(\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{H}'\boldsymbol{Q}^{1/2}.
\end{aligned}$$

Since $k > 0$ and $\boldsymbol{\Lambda} > \boldsymbol{0}$, we get $\boldsymbol{\Lambda}^{-1} - (\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-1} > \boldsymbol{0}$. Then we have

$$\boldsymbol{\Lambda}^{-1} - (\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-1} + k(\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-2} > \boldsymbol{0},$$

from which we can conclude the first assertion.

For proving the second assertion, we just need to prove that

$$\boldsymbol{Q}^{1/2}\boldsymbol{H}\begin{pmatrix} \boldsymbol{\Lambda}^{-1} - (\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-1} + k(\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{H}'\boldsymbol{Q}^{1/2},$$

is increased monotonously with respect to k, i.e., the function

$$\boldsymbol{g}(k) = \boldsymbol{\Lambda}^{-1} - (\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-1} + k(\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-2},$$

is increased monotonously with respect to k. After simplifying, we obtain

$$\frac{\partial \boldsymbol{g}(k)}{\partial k} = (\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-2}[\boldsymbol{I}_{p-m} - (k^{-1}\boldsymbol{\Lambda}^{-1} + \boldsymbol{I}_{p-m})^{-1}] > \mathbf{0}.$$

Since $k > 0$ and $\boldsymbol{\Lambda} > \mathbf{0}$, we know that $(\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-2} > 0$ and $\boldsymbol{I}_{p-m} > (k^{-1}\boldsymbol{\Lambda}^{-1} + \boldsymbol{I}_{p-m})^{-1}$, So the lemma is proved. $\square$

The results of Lemma 4.1 are correct for $\tilde{\delta} = \mathrm{adr}_{\boldsymbol{Q}}(\hat{\boldsymbol{\beta}}_r^F) - \mathrm{adr}_{\boldsymbol{Q}}[\hat{\boldsymbol{\beta}}_r^F(k)]$.

Lemma 4.2 ([4]). *Let $\boldsymbol{A}$ be a symmetric positive definite $n \times n$ matrix, $\boldsymbol{a}$ an $n \times 1$ vector and α a positive number. Then $\alpha\boldsymbol{A} - \boldsymbol{a}\boldsymbol{a}'$ is nonnegative definite if and only if $\boldsymbol{a}'\boldsymbol{A}^{+}\boldsymbol{a} \le \alpha$ is satisfied.*

Lemma 4.3 ([17]). *If $\hat{\boldsymbol{\beta}}_2$ is superior to $\hat{\boldsymbol{\beta}}_1$ with respect to amse, then it is superior to $\hat{\boldsymbol{\beta}}_1$ with respect to MSE and vice versa.*

Theorem 4.1. *Let us be given the estimator $\hat{\boldsymbol{\beta}}_r^F(k)$ under the linear regression model. If $k > 0$, then the amse difference δ is nonnegative if and only if*

$$\boldsymbol{\beta}'\boldsymbol{G}^{+}\boldsymbol{\beta} \le 1, \tag{4.2}$$

where $\boldsymbol{G} = 2(\sqrt{n}k)^{-1}\boldsymbol{P} + (\sqrt{n})^{-1}(\boldsymbol{P}\boldsymbol{C}^{*}\boldsymbol{P})^{+}$.

Proof. We prove the necessary and sufficient condition for the $AMSE_{\boldsymbol{Q}}$ difference $\boldsymbol{\Delta}$. Then, the Eq. (4.2) follows by using Lemma 4.3 and letting $\boldsymbol{Q} = \boldsymbol{I}_p$. We can write

$$\begin{aligned}
&\boldsymbol{\Delta} = \boldsymbol{AMSE}_{\boldsymbol{Q}}(\hat{\boldsymbol{\beta}}_r^F, \boldsymbol{\beta}) - \boldsymbol{AMSE}_{\boldsymbol{Q}}[\hat{\boldsymbol{\beta}}_r^F(k), \boldsymbol{\beta}] \ge \mathbf{0}\\
&\leftrightarrow \boldsymbol{ADR}_{\boldsymbol{Q}}(\hat{\boldsymbol{\beta}}_r^F) - \boldsymbol{ADR}_{\boldsymbol{Q}}[\hat{\boldsymbol{\beta}}_r^F(k)] - nk^2\boldsymbol{M}_k^{*}\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{M}_k^{*} \ge \mathbf{0}\\
&\leftrightarrow (\sqrt{n}k)^{-2}\tilde{\boldsymbol{\Delta}} - (\boldsymbol{M}_k^{*}\boldsymbol{\beta})(\boldsymbol{M}_k^{*}\boldsymbol{\beta})' \ge \mathbf{0}.
\end{aligned}$$

From Lemma 4.1, we know that $\tilde{\boldsymbol{\Delta}}$ is symmetric nonnegative definite and $\boldsymbol{M}_k^{*}\boldsymbol{\beta}$ is a $p \times 1$ vector. According to Lemma 4.2, the last equation is equivalent to $\boldsymbol{\beta}'\boldsymbol{M}_k^{*}\tilde{\boldsymbol{\Delta}}^{+}\boldsymbol{M}_k^{*}\boldsymbol{\beta} \le (\sqrt{n}k)^{-2}$. From Lemmas 3.1 and 4.1, the last expression is equivalent to

$$\boldsymbol{\beta}'(\boldsymbol{Q}^{1/2})^{+}\boldsymbol{H}\begin{pmatrix}\boldsymbol{\Gamma} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{H}'(\boldsymbol{Q}^{1/2})^{+}\boldsymbol{\beta} \le (\sqrt{n}k)^{-2},$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-1}\Big[\boldsymbol{\Lambda}^{-1} - (\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-1} + k(\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-2}\Big]^{-1}(\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-1}$.

It can be simplified as follows:

$$\begin{aligned}
&\boldsymbol{\beta}'(\boldsymbol{Q}^{1/2})^{+}\boldsymbol{H}\begin{pmatrix}(\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-1}\left[\dfrac{(\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^2 - \boldsymbol{\Lambda}^2}{(\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^2\boldsymbol{\Lambda}}\right]^{-1}(\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{H}'(\boldsymbol{Q}^{1/2})^{+}\boldsymbol{\beta} \le (\sqrt{n}k)^{-2}\\
&\leftrightarrow nk^2\boldsymbol{\beta}'(\boldsymbol{Q}^{1/2})^{+}\boldsymbol{H}\begin{pmatrix}\dfrac{\boldsymbol{\Lambda}}{(\boldsymbol{\Lambda} + k\boldsymbol{I}_{p-m})^2 - \boldsymbol{\Lambda}^2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{H}'(\boldsymbol{Q}^{1/2})^{+}\boldsymbol{\beta} \le 1\\
&\leftrightarrow nk^2\boldsymbol{\beta}'(\boldsymbol{Q}^{1/2})^{+}\boldsymbol{H}\begin{pmatrix}\dfrac{\boldsymbol{\Lambda}}{2k\boldsymbol{\Lambda} + k^2\boldsymbol{I}_{p-m}} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{H}'(\boldsymbol{Q}^{1/2})^{+}\boldsymbol{\beta} \le 1\\
&\leftrightarrow n\boldsymbol{\beta}'(\boldsymbol{Q}^{1/2})^{+}\boldsymbol{H}\begin{pmatrix}\dfrac{\boldsymbol{\Lambda}}{2k^{-1}\boldsymbol{\Lambda} + \boldsymbol{I}_{p-m}} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{H}'(\boldsymbol{Q}^{1/2})^{+}\boldsymbol{\beta} \le 1\\
&\leftrightarrow n\boldsymbol{\beta}'(\boldsymbol{Q}^{1/2})^{+}\boldsymbol{H}\begin{pmatrix}(2k^{-1}\boldsymbol{I}_{p-m} + \boldsymbol{\Lambda}^{-1})^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{H}'(\boldsymbol{Q}^{1/2})^{+}\boldsymbol{\beta} \le 1\\
&\leftrightarrow n\boldsymbol{\beta}'\left[\boldsymbol{Q}^{1/2}\boldsymbol{H}\begin{pmatrix}2k^{-1}\boldsymbol{I}_{p-m} + \boldsymbol{\Lambda}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{H}'\boldsymbol{Q}^{1/2}\right]^{+}\boldsymbol{\beta} \le 1\\
&\leftrightarrow n\boldsymbol{\beta}'\left[\boldsymbol{Q}^{1/2}\boldsymbol{H}\begin{pmatrix}2k^{-1}\boldsymbol{I}_{p-m} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{H}'\boldsymbol{Q}^{1/2} + \boldsymbol{Q}^{1/2}\boldsymbol{H}\begin{pmatrix}\boldsymbol{\Lambda}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\boldsymbol{H}'\boldsymbol{Q}^{1/2}\right]^{+}\boldsymbol{\beta} \le 1\\
&\leftrightarrow \boldsymbol{\beta}'\big[2(\sqrt{n}k)^{-1}\boldsymbol{Q}^{1/2}\boldsymbol{P}\boldsymbol{Q}^{1/2} + (\sqrt{n})^{-1}\boldsymbol{Q}^{1/2}(\boldsymbol{P}\boldsymbol{C}^{*}\boldsymbol{P})^{+}\boldsymbol{Q}^{1/2}\big]^{+}\boldsymbol{\beta} \le 1. \quad \square
\end{aligned}$$

Assumptions (i)–(iii) and 1–3 are used throughout the manuscript.

5. Range of shrinkage parameter

It is difficult to give a satisfying answer about how to select k. This is because the best k always depends on the unknown $\boldsymbol{\beta}$ in the practical applications which make the problem to be complicated.

In the process of deciding k, on one side, we must control the condition number of $\boldsymbol{C}_k^*$ to a lesser level if we want to avoid the instability of estimated coefficients brought by the morbidity of $\boldsymbol{C}^*$. Hence, we must do our best to let the ridge parameter k be big. Furthermore, we know the bigger the k is, the smaller the covariance of the estimated coefficients is. It implies that the estimator is more stable. On the other side, we know, in view of the biased estimator, when the k is smaller, the estimator will be better (the $\hat{\boldsymbol{\beta}}_r^F(k)$ will be apart badly from the actual $\boldsymbol{\beta}$ as k increasing). In other words, we must comply some principles to select k.

As stated in Theorem 4.1, we do not need to find out the best k in the practice. That is to say, we just need to find a k which can make $\hat{\boldsymbol{\beta}}_r^F(k)$ be superior to the $\hat{\boldsymbol{\beta}}_r^F$ in the sense of amse.

Although the criterion mentioned above is simple, our problem to select k is not yet completely solved. Therefore, we give a range to select k in Theorem 5.1.

Theorem 5.1. *Let us be given the estimator $\hat{\boldsymbol{\beta}}_r^F(k)$ under the linear regression model with true restrictions $\boldsymbol{R\beta} = \boldsymbol{r}$ and $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. If*

$$0 < k \leq \frac{2}{\boldsymbol{\beta}'\boldsymbol{Q}^{1/2}\boldsymbol{P}\boldsymbol{Q}^{1/2}\boldsymbol{\beta}}, \tag{5.1}$$

then the amse difference δ is nonnegative.

Proof. Since nonnegativity of δ follows from the nonnegative definite of $\boldsymbol{\Delta}$, then it is sufficient that we prove $\boldsymbol{\Delta}$ is nonnegative definite if Eq. (4.2) holds. From Theorem 4.1, the difference $\boldsymbol{\Delta}$ is nonnegative definite for $k > 0$ if and only if $\boldsymbol{\beta}'\left[\boldsymbol{Q}^{1/2}\boldsymbol{G}\boldsymbol{Q}^{1/2}\right]^+\boldsymbol{\beta} \leq 1$. The matrix $\boldsymbol{G}$ is symmetric nonnegative definite and has the same range as $\boldsymbol{P}$. Therefore, we can equivalently write $\boldsymbol{\beta}'\boldsymbol{P}\left[\boldsymbol{Q}^{1/2}\boldsymbol{G}\boldsymbol{Q}^{1/2}\right]^+\boldsymbol{P\beta} \leq 1$, where $\boldsymbol{P\beta}$ belongs to the range of $\boldsymbol{G}$. From Lemma 4.2, this inequality is satisfied if and only if $\boldsymbol{Q}^{1/2}\boldsymbol{G}\boldsymbol{Q}^{1/2} - \boldsymbol{P\beta\beta}'\boldsymbol{P}$ is nonnegative definite. A sufficient condition for nonnegative definiteness of this matrix is the nonnegative definiteness of the matrix $2k^{-1}\boldsymbol{Q}^{1/2}\boldsymbol{P}\boldsymbol{Q}^{1/2} - \boldsymbol{P\beta\beta}'\boldsymbol{P}$, which in turn is equivalent to $k\boldsymbol{\beta}'\boldsymbol{Q}^{1/2}\boldsymbol{P}\boldsymbol{Q}^{1/2}\boldsymbol{\beta} \leq 2$, i.e. $k \leq \frac{2}{\boldsymbol{\beta}'\boldsymbol{Q}^{1/2}\boldsymbol{P}\boldsymbol{Q}^{1/2}\boldsymbol{\beta}}$. Note that $\boldsymbol{\beta}'\boldsymbol{Q}^{1/2}\boldsymbol{P}\boldsymbol{Q}^{1/2}\boldsymbol{\beta}$ is nonzero if and only if $\boldsymbol{P}\boldsymbol{Q}^{1/2}\boldsymbol{\beta} \neq \boldsymbol{0}$, which is satisfied for every vector $\boldsymbol{\beta}$ with $\boldsymbol{R\beta} = \boldsymbol{r}$ and $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. □

Remark 5.1. If we have no restriction at all, the result of Theorem 5.1 can also be used for the ridge estimation in which $\boldsymbol{R}$ could be regard as $\boldsymbol{0}$. Then, the necessary and sufficient condition (4.2), and sufficient condition (5.1) are, respectively, simplified as

$$\boldsymbol{\beta}'\left(2(\sqrt{n}k)^{-1}\boldsymbol{I}_p + \sqrt{n}\boldsymbol{C}^{*^{-1}}\right)^+\boldsymbol{\beta} \leq 1, \quad 0 < k \leq \frac{2}{\boldsymbol{\beta}'\boldsymbol{Q}\boldsymbol{\beta}}.$$

6. Method demonstration

In this section, we examine the amse function performance of the proposed estimators numerically. Our sampling experiment consists of different values of k, i.e., $k = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Adopting the model (1.1) we simulate the response for $n = 50$ with 10^5 iteration from the following model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{f}(\boldsymbol{t}) + \boldsymbol{\epsilon}, \tag{6.1}$$

where $\boldsymbol{\beta} = (-2, 1, 3, -2, -5, 4)'$, $\boldsymbol{x_i} \sim N_5(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ for $i = 1, \ldots, n$ with

$$\boldsymbol{\mu}_x = \begin{pmatrix} 2.5 \\ 2 \\ 3 \\ 1 \\ -1 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_x = \begin{pmatrix} 1.9 & 1.8 & 1.8 & 1 & 1 \\ 1.8 & 1.8 & 1.8 & 1 & 1 \\ 1.8 & 1.8 & 4.25 & 1 & 1 \\ 1 & 1 & 1 & 2.49 & 1 \\ 1 & 1 & 1 & 1 & 2.25 \end{pmatrix},$$

$$f(t_i) = \sqrt{t_i(1-t_i)}\sin\left(\frac{2.1\pi}{t_i + 0.05}\right),$$

that is called the Doppler function for $t_i = (i - 0.05)/n$ and $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{V})$ where the elements of $\boldsymbol{V}$ are $v_{ij} = (0.05)^{|i-j|}$. The main reason for selecting such a structure for the nonlinear part is to check the efficiency of nonparametric estimations for the wavy function.

Table 1
Evaluation of parameters at different k values in model (6.1).

Coefficients	k				
	0	1	2	3	4
$\hat{\beta}_0$	−2.013830	−2.013880	−2.013930	−2.013979	−2.014029
$\hat{\beta}_1$	1.007080	1.007106	1.007131	1.007156	1.007182
$\hat{\beta}_2$	3.008268	3.008297	3.008327	3.008356	3.008386
$\hat{\beta}_3$	−2.003012	−2.003023	−2.003034	−2.003045	−2.003056
$\hat{\beta}_4$	−5.000220	−5.000221	−5.000221	−5.000222	−5.000223
$\hat{\beta}_5$	3.974054	3.973961	3.973868	3.973775	3.973682
$\boldsymbol{adb}[\hat{\boldsymbol{\beta}}_r(k)]'\boldsymbol{adb}[\hat{\boldsymbol{\beta}}_r(k)]$	0	6.72675e−07	2.69018e−06	6.05170e−06	2.07564e−05
adr$[\hat{\boldsymbol{\beta}}_r(k), \boldsymbol{\beta}]$	0.019685	0.019680	0.019680	0.019675	0.019670
$amse[\hat{\boldsymbol{\beta}}_r(k), \boldsymbol{\beta}]$	0.019685	0.019680	0.019680	0.019680	0.019680
δ	0	3.20297e−06	5.05995e−06	5.57180e−06	4.37924e−06
$mse[\hat{f}(\boldsymbol{t}), f(\boldsymbol{t})]$	0.0569625	0.0569737	0.0569849	0.0569961	0.0570074
	k				
	5	6	7	8	9
$\hat{\beta}_0$	−2.014078	−2.014128	−2.014177	−2.014227	−2.014276
$\hat{\beta}_1$	1.007207	1.007232	1.007258	1.007283	1.007308
$\hat{\beta}_2$	3.008416	3.008445	3.008475	3.008504	3.008534
$\hat{\beta}_3$	−2.003066	−2.003077	−2.003088	−2.003099	−2.003109
$\hat{\beta}_4$	−5.000224	−5.000225	−5.000225	−5.000226	−5.000227
$\hat{\beta}_5$	3.973589	3.973496	3.973404	3.973311	3.973218
$\boldsymbol{adb}[\hat{\boldsymbol{\beta}}_r^F(k)]'\boldsymbol{adb}[\hat{\boldsymbol{\beta}}_r^F(k)]$	1.68037e−05	2.41925e−05	3.29223e−05	4.29921e−05	5.44010e−05
adr$[\hat{\boldsymbol{\beta}}_r^F(k), \boldsymbol{\beta}]$	0.019665	0.019660	0.019660	0.019655	0.019650
$amse[\hat{\boldsymbol{\beta}}_r^F(k), \boldsymbol{\beta}]$	0.019685	0.019685	0.019690	0.019700	0.019705
δ	2.56310e−06	−9.55835e−07	−5.81675e−06	−1.20189e−05	−1.95614e−05
$mse[\hat{f}(\boldsymbol{t}), f(\boldsymbol{t})]$	0.0570187	0.0570300	0.0570414	0.0570528	0.0570642

For the weight function $W_{ni}(t_j)$, we use

$$W_{ni}(t_j) = \frac{1}{nh_n} K\left(\frac{t_i - tj}{h_n}\right) = \frac{1}{nh_n} . \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(t_i - t_j)^2}{2h_n^2}\right\}, \quad h_n = 0.01,$$

which is Priestley and Chao's weight with the Gaussian kernel. We also apply the cross-validation (C.V.) method to select the optimal bandwidth h_n, which minimizes the following C.V. function

$$\text{C.V. } (h_n) = \frac{1}{n} \sum_{i=1}^{n} \left(\widetilde{\boldsymbol{y}}^{(-i)} - \widetilde{\boldsymbol{X}}^{(-i)} \hat{\boldsymbol{\beta}}_r^{(-i)}(k)\right)^2,$$

where $\hat{\boldsymbol{\beta}}_r^{(-i)}(k)$ is obtained by replacing $\widetilde{\boldsymbol{X}}$ and $\widetilde{\boldsymbol{y}}$ with $\widetilde{\boldsymbol{X}}^{(-i)} = \left(x_{jk}^{(-i)}\right)$, $1 \le k \le n$, $1 \le j \le p$, $\widetilde{\boldsymbol{y}}^{(-i)} = \left(y_1^{(-i)}, \ldots, y_n^{(-i)}\right)$, $x_{sk}^{(-i)} = x_{sk} - \sum_{j \ne i}^{n} W_{nj}(t_i) x_{sj}$, $y_k^{(-i)} = y_k - \sum_{j \ne i}^{n} W_{nj}(t_i) y_j$ in (2.5). Here $\widetilde{\boldsymbol{y}}^{(-i)}$ is the predicted value of $\boldsymbol{y} = (y_1, \ldots, y_n)$ at $\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{pi})$ with y_i and x_i left out of the estimation of the $\boldsymbol{\beta}$.

For the linear restriction, suppose the pre-specified matrix $\boldsymbol{R}$ is given by

$$\boldsymbol{R} = \begin{pmatrix} 1 & 5 & -3 & -1 & -1 & 0 \\ -2 & -1 & 0 & -2 & 3 & 1 \\ 1 & 2 & 1 & 3 & -2 & 0 \\ 4 & -1 & 2 & 2 & 0 & -2 \\ 5 & 3 & 4 & -5 & 1 & 0 \end{pmatrix}.$$

All computations were conducted using the statistical package R. The ratio of the largest eigenvalue to the smallest eigenvalue of matrix $\boldsymbol{C}^*$ is $\lambda_5/\lambda_1 = 354.91$ which implies the existence of multicollinearity in the data set.

In Table 1, we compute the restricted ridge estimators of linear parameter. We numerically estimate the $\boldsymbol{adb}'\boldsymbol{adb}$, adr, amse, δ and $mse[\hat{f}(\boldsymbol{t}), f(\boldsymbol{t})] = \frac{1}{n}\sum_{i=1}^{n}[\hat{f}(t_i) - f(t_i)]^2$ for different values of k and 10^5 samples with size of 50.

In Fig. 1, in the left part, the amse of $\hat{\boldsymbol{\beta}}_r^F(k)$ (dot line) and $\hat{\boldsymbol{\beta}}_r^F$ (solid line) and in the right part, the δ versus ridge parameter k are plotted for the model (6.1). For estimating the nonlinear part, we simulate response from model (6.1) for $n = 1000$ again. In Fig. 2, the nonparametric part of the model (6.1) is plotted in the top left plot. This function is difficult to estimate and provides a good test case for the nonparametric regression method. The function is spatially inhomogeneous which means that it's smoothness (second derivatives) vary over t. The top right plot shows $n = 1000$ data points after removing

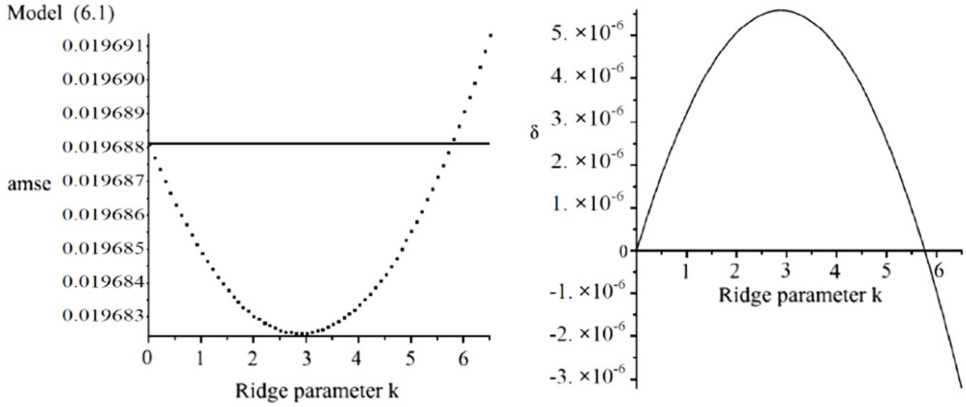

Fig. 1. The diagram of amse and δ versus k.

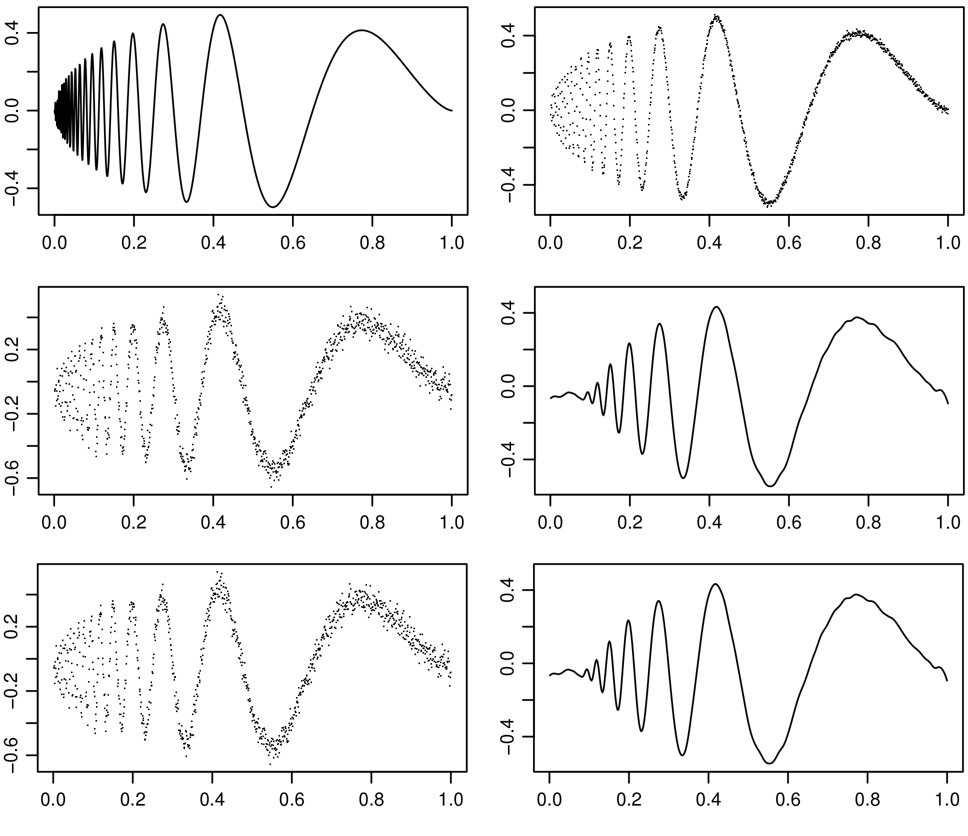

Fig. 2. Estimation of the nonparametric part by local linear regression for $n = 1000$.

the linear part, i.e., $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$. In the continuation, the middle left and the right plots show the residuals which obtained after estimation of the linear part of the model by $\hat{\boldsymbol{\beta}}_r^F$, that is, $\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_r^F$ and the fitted function, respectively. The bottom left and right plots are the middle part when $\hat{\boldsymbol{\beta}}_r^F$ is replaced with $\hat{\boldsymbol{\beta}}_r^F(k)$ when k equals to the median range of (5.1), which is approximately equal to 2.8806. The minimum of C.V. occurred at $h_n = 0.0512$ for the model (6.1) for $n = 1000$.

7. Summary and conclusions

In this paper, we proposed two new estimators in a partial linear model when the errors were dependent and some additional linear constraints held on the whole parameter space $\boldsymbol{\beta}$. It was also assumed that the covariance matrix of the errors are unknown for operational use. In the presence of multicollinearity in a partial linear model we introduced the restricted feasible ridge regression estimator $\hat{\boldsymbol{\beta}}_r^F(k)$ versus the non-ridge version $\hat{\boldsymbol{\beta}}_r$ under dependency among column vectors of the design matrix. The asymptotic properties of proposed estimators were derived and the superiority conditions of the feasible ridge estimator over non-ridge form based on ridge parameter k was proved by theorems. Applying Kernel smoothing and cross-validation methods, we estimated the nonlinear functions of the proposed model.

Some points arising from this study are listed below.

- A near dependency among the column of $\boldsymbol{C}^*$ from $\lambda_5/\lambda_1 = 354.91$, that is, the feasible restricted ridge weighted least squares estimator should be taken as a leading estimator in our studies.
- As it can be seen from Fig. 1, the δ increases (amse decreases) at first and then decreases (increases), that provides a reason for the assertion in Theorem 5.1. Furthermore, the maximum of δ (minimum of the amse) is obtained when k equals the median range of (5.1) i.e., $\frac{1}{\boldsymbol{\beta}'\boldsymbol{Q}^{1/2}\boldsymbol{P}\boldsymbol{Q}^{1/2}\boldsymbol{\beta}}$ which is approximately equal to 2.8806 in model (6.1).
- From Fig. 2, it can be concluded that our semiparametric approach is truly efficient in the sense of better nonparametric function estimation.

Acknowledgment

The authors would like to sincerely thank the Associate Editor for constructive and important comments that substantially improved the presentation of the paper.

References

[1] J.K. Baksalary, S. Puntanen, G.P.H. Styan, A property of the dispersion matrix of the best linear unbiased estimator in the general Gauss–Markov model, Sankhya A 52 (1990) 279–296.
[2] E.A. Duran, W.K. Härdle, M. Osipenko, Difference based ridge and Liu type estimators in semiparametric regression models, J. Multivariate Anal. 105 (2012) 164–175.
[3] R.F. Engle, C.W.J. Granger, J. Rice, A. Weiss, Semiparametric estimates of the relation between weather and electricity sales, J. Amer. Statist. Assoc. 80 (1986) 310–319.
[4] R. Farebrother, Further results on the mean square error of ridge regression, J. Roy. Statist. Soc. Ser. B 38 (1976) 248–250.
[5] J. Grob, Restricted ridge estimation, Statist. Probab. Lett. 65 (2003) 57–64.
[6] W. Härdle, H. Liang, J. Gao, Partially Linear Models, Physika Verlag, Heidelberg, 2000.
[7] A.E. Hoerl, R.W. Kennard, Ridge regression biased estimation for non-orthogonal problems, Technometrics 12 (1970) 69–82.
[8] B.M.G. Kibria, A.K.Md.E. Saleh, Preliminary test ridge regression estimators with students t errors and conflicting test-statistics, Metrika 59 (2004) 105–124.
[9] M. Müller, Semiparametric Extensions to Generalized Linear Models. Habilitationsschrift, 2000.
[10] M. Roozbeh, M. Arashi, H.A. Niroumand, Semiparametric ridge regression approach in partially linear models, Comm. Statist. Simulation Comput. 39 (2010) 449–460.
[11] M. Roozbeh, M. Arashi, H.A. Niroumand, Ridge regression methodology in partial linear models with correlated errors, J. Stat. Comput. Simul. 81 (2011) 517–528.
[12] A.K.Md.E. Saleh, Theory of Preliminary Test and Stein-type Estimation with Applications, John Wiley, New York, 2006.
[13] N. Sarkar, A new estimator combining the ridge regression and the restricted least squares methods of estimation, Comm. Statist. Theory Methods 21 (1992) 1987–2000.
[14] P. Speckman, Kernel smoothing in partial linear models, J. Roy. Statist. Soc. Ser. B 50 (1988) 413–437.
[15] P.A.V.B. Swamy, J.S. Mehta, P.N. Rappoport, Two methods of evaluating Hoerl and Kennard's ridge regression, Commun. Stat. 12 (1978) 1133–1155.
[16] G. Tabakan, F. Akdeniz, Difference-based ridge estimator of parameters in partial linear model, Statist. Papers 51 (2010) 357–368.
[17] C.M. Theobald, Generalizations of mean square error applied to ridge regression, J. Roy. Statist. Soc. Ser. B 36 (1974) 103–106.
[18] A. Zellner, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, J. Amer. Statist. Assoc. 57 (1962) 348–368.
[19] C. Zhang, H. Yang, The conditional ridge-type estimation in singular linear model with linear equality restriction, Statistics 41 (6) (2007) 485–494.
[20] Z. Zhong, H. Yang, Ridge estimation to the restricted linear model, Comm. Statist. Theory Methods 36 (2007) 2099–2115.