Stat Papers (2020) 61:869–898
https://doi.org/10.1007/s00362-017-0967-8

REGULAR ARTICLE

Ridge-type pretest and shrinkage estimations in partially linear models

Bahadır Yüzbaşı[1] · S. Ejaz Ahmed[2] · Dursun Aydın[3]

Received: 17 January 2017 / Revised: 17 August 2017 / Published online: 21 November 2017

Abstract In this paper, we suggest pretest and shrinkage ridge regression estimators for a partially linear regression model, and compare their performance with some penalty estimators. We investigate the asymptotic properties of proposed estimators. We also consider a Monte Carlo simulation comparison, and a real data example is presented to illustrate the usefulness of the suggested methods.

Keywords Pretest estimation · Shrinkage estimation · Ridge regression · Smoothing spline · Partially linear model

1 Introduction

We are interested in estimating the following partially linear regression model (PLRM):

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + f(t_i) + \varepsilon_i, \quad i = 1, \dots, n, \tag{1}$$

where y_i's are observed values of response variable, $\mathbf{x}_i' = (x_{i1}, \dots, x_{ip})$ is the ith observed vector of explanatory variables including $p-$dimensional vector with $p \leq n$, t_i's are values of an extra univariate variable satisfying $t_1 \leq \cdots \leq t_n$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is an unknown $p-$dimensional vector of regression coefficients, $f(\cdot)$ is an unknown smooth function, and ε_i's are random disturbances assumed to be

✉ Bahadır Yüzbaşı
b.yzb@hotmail.com

1 Department of Econometrics, Inonu University, Malatya, Turkey

2 Department of Mathematics and Statistics, Brock University, St.Catharines, Canada

3 Department of Statistics, Mugla Sitki Kocman University, Mugla, Turkey

as $\mathcal{N}\left(0, \sigma^2\right)$. Also, the vector $\boldsymbol{\beta}$ is the parametric part of the model, and $f(\cdot)$ is the nonparametric part of the model. The model (1) also called a semi-parametric model, and in vector–matrix form is written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{f} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\mathbf{y} = (y_1, \ldots, y_n)'$, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$, $\boldsymbol{f} = (f(t_1), \ldots, f(t_n))'$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$ is random vector with $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}_n$.

The PLRM generalizes both parametric linear regression and nonparametric regression models which correspond to the cases $\boldsymbol{\beta} = \mathbf{0}$ and $\boldsymbol{f} = \mathbf{0}$, respectively. The key idea is to estimate the parameter vector $\boldsymbol{\beta}$, the function $\boldsymbol{f}$ and the mean vector $\mathbf{X}\boldsymbol{\beta} + \boldsymbol{f}$.

PLRMs have many applications. These models were originally studied by Engle et al. (1986) to determine the effect of weather on the electricity sales. In the following, several authors have investigated the PLRM, including Speckman (1988), Eubank et al. (1998), Schimek (2000), Liang (2006), Ahmed (2014), Aydın (2014) and Wu and Asar (2016), among others. The most popular approach for the PLRM is based on the fact that the cubic spline is a linear estimator for the nonparametric regression problem. Hence, the nonparametric procedure can be naturally extended to handle the PLRM.

In the using linear least squares regression, it is often encountered the problem of multicollinearity. In order to solve this issue, ridge regression has been proposed by Hoerl and Kennard (1970). It is well known that ridge estimator provides a slight improvement on the estimations of partial regression coefficients when the column vectors of the matrix in a linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ are highly correlated. In recent years a number of authors have proposed the use of the ridge type (biased) estimate approach to solve the problem of multicollinearity on estimating the parameters of the PLRMs, see Roozbeh and Arashi (2013), Arashi and Valizadeh (2015) and Yüzbaşı and Ahmed (2016). Contrary to these studies, we combine the idea of Speckman's smoothing spline with the Ridge type-estimation in a optimal way in order to controlling bias parameter because of several reasons. Here are two of them: (1-) The principle of adding a penalty term to a sum of squares or more generally to a log-likelihood applies to a wide variety of linear and non-linear problems. (2-) The researchers, especially Shiller (1984), Green et al. (1985) and Eubank (1986), think that this method simply seems to work well.

For PLRMs, Ahmed et al. (2007) considered a profile least squares approach based on using kernel estimates of $f(\cdot)$ to construct absolute penalty, shrinkage, and pretest estimators of $\boldsymbol{\beta}$ in the case where $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2'\right)'$. Similarly, for PLRMs, the suitability of estimating the nonparametric component based on the B-spline basis function is explored by Raheem et al. (2012).

In this paper, we introduce estimations techniques based on ridge regression when the matrix $\mathbf{X}'\mathbf{X}$ appears to be ill-conditioned in the PLRM using smoothing splines. Also, we consider that the coefficients $\boldsymbol{\beta}$ can be partitioned as $\left(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2\right)$ where $\boldsymbol{\beta}_1$ is the coefficient vector for main effects, and $\boldsymbol{\beta}_2$ is the vector for nuisance effects. We are essentially interested in the estimation of $\boldsymbol{\beta}_1$ when it is reasonable that $\boldsymbol{\beta}_2$ is close to zero. We suggest pretest ridge regression, shrinkage ridge regression and positive shrinkage ridge regression estimators for PLRMs. In the empirical applications, shrinkage estimators have been not paid attention to much due to the computational load till recently. However, with improvements in computing capability, this situation has changed. For example, as our real data example in Sect. 6, the annual salary of a baseball player may or may or not be effected by a number of situations (co-variates). A real baseball coach's opinion, experience, and knowledge often give precise information regarding certain parameter values in an annual salary of a baseball player model. Furthermore, some variable selections techniques give an idea about important co-variates. Hence, researchers may take into consideration this auxiliary information and choose either the full model or the candidate sub-model for following work. The Stein-rule and pretest estimation procedures has received considerable attention from researchers since these methods can be obtained by shrinking the full model estimates in the direction of the subspace leads to more efficient estimators when the shrinkage is adaptive and based on the estimated distance between the subspace and the full space, for more information.

The organization of this study is given as following: the full and sub-model estimators are given in Sect. 2. The pretest, shrinkage estimators and some penalized estimations, namely the least absolute shrinkage and selection operator (Lasso), the adaptive Lasso (aLasso) and the smoothly clipped absolute deviation (SCAD) are also presented in Sect. 3. The asymptotic investigations of listed estimators are given in Sect. 4. In order to demonstrate the relative performance with our suggested estimators, a Monte Carlo simulation study is conducted in Sect. 5. A real data example is presented to illustrate the usefulness of the suggested estimators in Sect. 6. Finally, the conclusions and remarks are given in Sect. 7.

2 Full model estimation

Generally, the back-fitting algorithm is considered for the estimation of the model (2) . In this paper, we consider Speckman approach based on penalized residual sum of squares method for estimation purpose. We estimate $\boldsymbol{\beta}$ and $\boldsymbol{f}$ by minimizing the following penalized sum of squares equation

$$\begin{aligned} SS(\boldsymbol{\beta}, f) &= \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i' \boldsymbol{\beta} - f(t_i)\right)^2 + \lambda \int_a^b \left(f''(t)\right)^2 dt \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{f})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{f}) + \lambda \boldsymbol{f}' \mathbf{K} \boldsymbol{f}, \end{aligned} \tag{3}$$

where $\mathbf{K}$ is positive definite penalty matrix with solution,

$$\widehat{\boldsymbol{f}} = \mathbf{S}_\lambda \mathbf{y},$$

where $\mathbf{S}_\lambda = (\mathbf{I}_n - \lambda \mathbf{K})^{-1}$ is a well-known positive-definite (symmetrical) smoother matrix which depends on fixed smoothing parameter $\lambda > 0$ and the knot points $t_1, \ldots, t_n$. The smoother matrix $\mathbf{S}_\lambda$ is obtained from univariate cubic spline smoothing (i.e. from penalized sum of squares equation (3) without parametric terms $\mathbf{X}\boldsymbol{\beta}$). Function $\widehat{f}_\lambda$, the estimator of function f, is obtained by cubic spline interpolation that rests on condition $\widehat{f}(t_i) = \left(\widehat{f}\right)_i$, $i = 1, \ldots, n$. The penalty $\mathbf{K}$ matrix in (3) is obtained by means of the knot points, and defined as following way:

$$\mathbf{K} = \mathbf{U}'\mathbf{R}^{-1}\mathbf{U},$$

where $h_j = t_{j+1} - t_j$, $j = 1, 2, \ldots, n-1$, $\mathbf{U}$ is tri-diagonal $(n-2) \times n$ matrix with $\mathbf{U}_{jj} = 1/h_j$, $\mathbf{U}_{j,j+1} = -\left(1/h_j + 1/h_{j+1}\right)$, $\mathbf{U}_{j,j+2} = 1/h_{j+1}$ and $\mathbf{R}$ is symmetric tri-diagonal matrix of order $(n-2)$ with $\mathbf{R}_{j-1,j} = \mathbf{R}_{j,j-1} = h_j/6$ and $\mathbf{R}_{jj} = \left(h_j + h_{j+1}\right)/3$.

The first term in the Eq. (3) denotes the residual sum of the squares and it penalizes the lack of fit. The second term in the same equation denotes the roughness penalty and it penalizes the curvature of the function. The amount of penalty is controlled by a smoothing parameter $\lambda > 0$. In general, large values of λ produce smoother estimators while smaller values produce more wiggly estimators. Thus, the λ plays a key role in controlling the trade-off between the goodness of fit represented by $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{f})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{f})$ and smoothness of the estimate measured by $\lambda \boldsymbol{f}'\mathbf{K}\boldsymbol{f}$.

In this paper we have discussed the partially linear model with a univariate non-parametric predictor t given in model (1). If $t > 1$, then a single smooth function in model (1) is replaced by two or more unspecified smooth functions. In this case, the fitted model is of the form

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \sum_{j=1}^{p} f_j\left(t_{ij}\right) + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{4}$$

The model (4) is also called as the partially linear additive model.

As stated previously, the main idea in PLRM is to estimate the vector $\boldsymbol{\beta}$ and $\boldsymbol{f}$ by minimizing the penalized residual sum of squares criterion (3). We carry this idea a step further for the partially linear additive model (4). In this context, the optimization problem is to minimize

$$(\widehat{\boldsymbol{\beta}}; \widehat{\boldsymbol{f}}) = \underset{\boldsymbol{\beta},\boldsymbol{f}}{\arg\min} \left\{ \sum\nolimits_{i=1}^{n} \left(\tilde{y}_i - \tilde{\mathbf{x}}_i'\boldsymbol{\beta} - \sum_{j=1}^{p} f_j\left(t_{ij}\right) \right)^2 + \sum_{j=1}^{p} \lambda_j \int_a^b \left(f_j''(t) \right)^2 dt \right\}, \tag{5}$$

over all twice differentiable functions f_j defined on $[a, b]$. In here f_j is a unspecified univariate function and λ_j's are separate smoothing parameters for each smooth functions f_j. As in the case with a single smooth function, if the λ_j's are all zero, we get a smooth system that interpolates the data. Also, when each λ_j goes to ∞, we obtain a standard least squares fit.

In the partially linear additive regression models, the functions λ_j can be estimated by a single smoothing spline manner. Using a straightforward extension of the arguments used in a univariate smoothing spline, the solution to Eq. (5) can be obtained by minimizing the matrix–vector form of Eq. (5), given by

$$(\widehat{\boldsymbol{\beta}};\, \widehat{\boldsymbol{f}}) = \underset{\boldsymbol{\beta},\boldsymbol{f}}{\arg\min} \left\{ \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \sum_{j=1}^{p} \boldsymbol{f_j} \right)' \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \sum_{j=1}^{p} \boldsymbol{f_j} \right) + \sum_{j=1}^{p} \lambda_j \boldsymbol{f_j}' \mathbf{K}_j \boldsymbol{f_j} \right\},$$

where $\mathbf{K}_j$'s are the penalty matrices for each predictor, similarly to the $\mathbf{K}$ for a univariate predictor given in Eq. (3), see Hastie and Tibshirani (1990) for additive models.

The resulting estimator is called as partial spline, see Wahba (1990). On the other hand, Eq. (3) is also known as the roughness penalty approach Green and Silverman (1994). This estimation concept is based on iterative solution of the normal equations Rice (1986) indicated that partial spline estimator is asymptotically biased for the optimal choice as the components $\mathbf{X}$ depend on t. Applying results due to Speckman (1988), this bias can be substantially reduced. In the following section, we present full model semi-parametric estimation based on ridge regression.

2.1 Full model and sub-model semi-parametric ridge strategies

For a pre-specified value of λ the corresponding estimators $\boldsymbol{\beta}$ and $\boldsymbol{f}$ for based on model (2) can be obtained by

$$\widehat{\boldsymbol{\beta}} = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{y}} \text{ and } \widehat{\boldsymbol{f}} = \mathbf{S}_\lambda \left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right),$$

where $\tilde{\mathbf{X}} = (\mathbf{I}_n - \mathbf{S}_\lambda)\,\mathbf{X}$ and $\tilde{\mathbf{y}} = (\mathbf{I}_n - \mathbf{S}_\lambda)\,\mathbf{y}$, respectively.

By multiplying both sides of model (2) with $(\mathbf{I}_n - \mathbf{S}_\lambda)$,

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}, \tag{6}$$

where $\tilde{\boldsymbol{f}} = (\mathbf{I}_n - \mathbf{S}_\lambda)\,\boldsymbol{f}$, $\tilde{\boldsymbol{\varepsilon}} = \tilde{\boldsymbol{f}} + \boldsymbol{\varepsilon}^*$ and $\boldsymbol{\varepsilon}^* = (\mathbf{I}_n - \mathbf{S}_\lambda)\,\boldsymbol{\varepsilon}$.

Therefore, model (6) is transformed into an optimal problem to estimate semi-parametric estimator. We now consider model (6) with ridge penalty to estimate semi-parametric ridge estimator. We formulate this as follows:

$$\underset{\boldsymbol{\beta}}{\arg\min} \left(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}\right)' \left(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}\right) + k\boldsymbol{\beta}'\boldsymbol{\beta}, \tag{7}$$

where $k \geq 0$ is the tuning parameter. By solving (7), we get full model semi-parametric ridge regression estimator of $\boldsymbol{\beta}$ as follows:

$$\widehat{\boldsymbol{\beta}}^{\text{Ridge}} = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + k\mathbf{I}_p\right)^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{y}}.$$

Let $\widehat{\boldsymbol{\beta}}_1^{\text{FM}}$ be the semi-parametric unrestricted or full model ridge estimator of $\boldsymbol{\beta}_1$. From model (7), the semi-parametric full model ridge estimator $\widehat{\boldsymbol{\beta}}_1^{\text{FM}}$ of $\boldsymbol{\beta}_1$ is

$$\widehat{\boldsymbol{\beta}}_1^{\text{FM}} = \left(\tilde{\mathbf{X}}_1'\tilde{\mathbf{M}}_2^{\text{R}}\tilde{\mathbf{X}}_1 + k\mathbf{I}_{p_1}\right)^{-1} \tilde{\mathbf{X}}_1'\tilde{\mathbf{M}}_2^{\text{R}}\tilde{\mathbf{y}},$$

where $\tilde{\mathbf{M}}_2^{\mathrm{R}} = \mathbf{I}_n - \tilde{\mathbf{X}}_2 \left(\tilde{\mathbf{X}}_2'\tilde{\mathbf{X}}_2 + k\mathbf{I}_{p_2}\right)^{-1} \tilde{\mathbf{X}}_2'$ and $\tilde{\mathbf{X}}_i = (\mathbf{I}_n - \mathbf{S}_\lambda)\, \mathbf{X}_i$, $i = 1, 2$.

Now, consider $\boldsymbol{\beta}_2 = \mathbf{0}$, and add ridge penalty function on model (1) ,

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + f(t_i) + \varepsilon_i \;\text{ subject to } \boldsymbol{\beta}'\boldsymbol{\beta} \le \phi^2 \text{ and } \boldsymbol{\beta}_2 = \mathbf{0}\ .$$

Hence we have the following partially linear sub-model

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{f} + \boldsymbol{\varepsilon} \;\text{ subject to } \boldsymbol{\beta}_1'\boldsymbol{\beta}_1 \le \phi^2. \tag{8}$$

Let us denote $\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}$ the semi-parametric sub-model or restricted ridge estimator of $\boldsymbol{\beta}_1$ as defined subsequently. Generally speaking, $\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}$ performs better than $\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}$ when $\boldsymbol{\beta}_2$ close to $\mathbf{0}$. However, for $\boldsymbol{\beta}_2$ away from the origin $\mathbf{0}$, $\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}$ can be inefficient. From model (8), the semi-parametric sub-model ridge estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}$ of $\boldsymbol{\beta}_1$ has the form

$$\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}} = \left(\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_1 + k\mathbf{I}_{p_1}\right)^{-1} \tilde{\mathbf{X}}_1'\tilde{\mathbf{y}}.$$

3 Pretest, shrinkage and some penalty estimation strategies

The pretest estimator is a combination of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}$ and $\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}$ via an indicator function $\mathrm{I}\left(\mathcal{T}_n \le c_{n,\alpha}\right)$, where $\mathcal{T}_n$ is an appropriate test statistic to test $\mathrm{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ versus $\mathrm{H}_A : \boldsymbol{\beta}_2 \neq \mathbf{0}$. Moreover, $c_{n,\alpha}$ is an $\alpha-$level critical value using the distribution of $\mathcal{T}_n$.

We define test statistics for testing null hypothesis $\mathrm{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ as follows:

$$\mathcal{T}_n = \frac{n}{\widehat{\sigma}^2}\widehat{\boldsymbol{\beta}}_2' \left(\tilde{\mathbf{X}}_2'\tilde{\mathbf{M}}_1\tilde{\mathbf{X}}_2\right)\widehat{\boldsymbol{\beta}}_2,$$

where

$$\widehat{\sigma}^2 = \frac{1}{n}.\frac{\|(\mathbf{I}_n - \mathbf{H}_\lambda)\,\tilde{\mathbf{y}}\|^2}{\mathrm{tr}\,(\mathbf{I}_n - \mathbf{H}_\lambda)'\,(\mathbf{I}_n - \mathbf{H}_\lambda)},$$

and

$$\widehat{\boldsymbol{\beta}}_2 = \left(\tilde{\mathbf{X}}_2'\tilde{\mathbf{M}}_1\tilde{\mathbf{X}}_2\right)^{-1} \tilde{\mathbf{X}}_2'\tilde{\mathbf{M}}_1\tilde{\mathbf{y}},$$

with $\tilde{\mathbf{M}}_1 = \mathbf{I}_n - \tilde{\mathbf{X}}_1 \left(\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_1\right)^{-1} \tilde{\mathbf{X}}_1'$ and $\mathbf{H}_\lambda$ is called as smoother matrix for the model (1).

The $\mathbf{H}_\lambda$ matrix is obtained as follows:

$$\begin{aligned}\widehat{\mathbf{y}} &= \mathbf{X}\widehat{\boldsymbol{\beta}}^{\mathrm{FM}} + \widehat{\boldsymbol{f}}\\ &= \mathbf{X}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + k\mathbf{I}_p\right)^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}} + \boldsymbol{S}_\lambda\left(\mathbf{y} - \mathbf{X}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} + k\mathbf{I}_p\right)^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}\right)\end{aligned}$$

$$
\begin{aligned}
&= \mathbf{X}\left(\mathbf{I}-\mathbf{S}_\lambda\right)\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}+k\mathbf{I}_p\right)^{-1}\tilde{\mathbf{X}}'\mathbf{y} \\
&\quad+\mathbf{S}_\lambda\left(\mathbf{y}-\mathbf{X}\left(\mathbf{I}-\mathbf{S}_\lambda\right)\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}+k\mathbf{I}_p\right)^{-1}\tilde{\mathbf{X}}'\mathbf{y}\right) \\
&= \tilde{\mathbf{X}}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}+k\mathbf{I}_p\right)^{-1}\tilde{\mathbf{X}}'\mathbf{y}+\mathbf{S}_\lambda\left(\mathbf{y}-\tilde{\mathbf{X}}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}+k\mathbf{I}_p\right)^{-1}\tilde{\mathbf{X}}'\mathbf{y}\right) \\
&= \mathbf{H}\mathbf{y}+\mathbf{S}_\lambda\mathbf{y}-S_\lambda\mathbf{H}\mathbf{y} \\
&= \left(\mathbf{S}_\lambda+\left(\mathbf{I}_n-\mathbf{S}_\lambda\right)\mathbf{H}\right)\mathbf{y} \\
&= \mathbf{H}_\lambda\mathbf{y},
\end{aligned}
$$

where $\mathbf{H}=\tilde{\mathbf{X}}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}+k\mathbf{I}_p\right)^{-1}\tilde{\mathbf{X}}'$. Thus, the mentioned smoother matrix is

$$
\mathbf{H}_\lambda=\mathbf{S}_\lambda+\left(\mathbf{I}_n-\mathbf{S}_\lambda\right)\mathbf{H}.
$$

Under H_0, the test statistic $\mathcal{T}_n$ follows chi-square distribution with p_2 degrees of freedom for large n values. Then, we can choose an $\alpha-$level critical value $c_{n,\alpha}$.

The semi-parametric ridge pretest estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}}$ of $\boldsymbol{\beta}_1$ is defined by

$$
\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}}=\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)\mathrm{I}\left(\mathcal{T}_n\leq c_{n,\alpha}\right).
$$

The shrinkage estimator for a PLRM was introduced by Ahmed et al. (2007). This shrinkage estimator is a smooth function of the test statistic.

The semi-parametric ridge shrinkage or Stein-type estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}$ of $\boldsymbol{\beta}_1$ is defined by

$$
\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}=\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}+\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)\left(1-(p_2-2)\mathcal{T}_n^{-1}\right),\ p_2\geq 3.
$$

The positive part of the semi-parametric ridge shrinkage estimator $\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}$ of $\boldsymbol{\beta}_1$ defined by

$$
\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}=\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}+\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)\left(1-(p_2-2)\mathcal{T}_n^{-1}\right)^{+},
$$

where $z^{+}=max\,(0,z)$

3.1 Some penalty estimation strategies

Now, we suggest the semi-parametric penalty estimators by using the smoothing spline method. For a given penalty function $\pi\left(\cdot\right)$ and regularization parameter λ, the general form of the objective function of semi-parametric penalty estimators can be written as

$$
\sum\nolimits_{i=1}^{n}\left(\tilde{y}_i-\tilde{\mathbf{x}}_i'\boldsymbol{\beta}\right)^2+\lambda\pi\left(\cdot\right),
$$

where $\tilde{y}_i$ is the ith observation of $\tilde{\mathbf{y}}$, $\tilde{\mathbf{x}}_i'$ is the i th row of $\tilde{\mathbf{X}}$ and $\pi(\cdot) = \sum_{i=1}^{p} |\beta_i|^{\iota}$, $\iota > 0$.

If $\iota = 2$, then the ridge regression estimator can be written

$$\widehat{\boldsymbol{\beta}}^{\text{Ridge}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum\nolimits_{i=1}^{n} \left(\tilde{y}_i - \tilde{\mathbf{x}}_i'\boldsymbol{\beta}\right)^2 + \lambda \sum_{j=1}^{p} \left|\beta_j\right|^2 \right\}.$$

For $\iota = 1$, it is related to the Lasso, that is,

$$\widehat{\boldsymbol{\beta}}^{\text{Lasso}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum\nolimits_{i=1}^{n} \left(\tilde{y}_i - \tilde{\mathbf{x}}_i'\boldsymbol{\beta}\right)^2 + \lambda \sum_{j=1}^{p} \left|\beta_j\right| \right\}.$$

The aLasso estimator $\boldsymbol{\beta}^{\text{aLasso}}$ is defined as

$$\widehat{\boldsymbol{\beta}}^{\text{aLasso}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum\nolimits_{i=1}^{n} \left(\tilde{y}_i - \tilde{\mathbf{x}}_i'\boldsymbol{\beta}\right)^2 + \lambda \sum_{j=1}^{p} \widehat{\zeta}_j \left|\beta_j\right| \right\},$$

where the weight function is

$$\widehat{\zeta}_j = \frac{1}{|\widehat{\beta}_j^*|^{\iota}}; \quad \iota > 0,$$

and $\widehat{\beta}_j^*$ is a root-n consistent estimator of β.

The SCAD estimator $\widehat{\boldsymbol{\beta}}^{\text{SCAD}}$ is defined as

$$\widehat{\boldsymbol{\beta}}^{\text{SCAD}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum\nolimits_{i=1}^{n} \left(\tilde{y}_i - \tilde{\mathbf{x}}_i'\boldsymbol{\beta}\right)^2 + \sum_{j=1}^{p} J_{\alpha,\lambda}\left(\beta_j\right) \right\},$$

where

$$J_{\alpha,\lambda}(x) = \lambda \left\{ \mathrm{I}\left(|x| \le \lambda\right) + \frac{(\alpha\lambda - |x|)_+}{(\alpha - 1)\lambda} \mathrm{I}\left(|x| > \lambda\right) \right\}, \; x \ge 0.$$

In order to select the optimal regularization parameter λ, we used *glmnet* and *ncvreg* packages in R for Lasso and SCAD, respectively. Also, the aLasso is obtained by 10 fold cross-validation with weights from the 10 fold cross-validated Lasso.

4 Asymptotic analysis

In this section, we define expressions for asymptotic distributional biases (ADBs), asymptotic covariance matrices and asymptotic distributional risks (ADRs) of the pretest and shrinkage along with full model and sub-model estimators. For this purpose we consider a sequence $\{K_n\}$ is given by

$$K_n : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{2(n)} = \frac{\boldsymbol{w}}{\sqrt{n}}, \ \boldsymbol{w} = \left(w_1, \ldots, w_{p_2}\right)' \in \mathbb{R}^{p_2}.$$

Now, we define a quadratic loss function using a positive definite matrix (p.d.m) $\mathbf{W}$, by

$$L\left(\boldsymbol{\beta}_1^*\right) = n\left(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1\right)' \mathbf{W}\left(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1\right),$$

where $\boldsymbol{\beta}_1^*$ is anyone of suggested estimators. Now, under $\{K_n\}$, we can write the asymptotic distribution function of $\boldsymbol{\beta}_1^*$ as

$$F(\mathbf{x}) = \lim_{n\to\infty} P\left(\sqrt{n}\left(\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_1\right) \le \mathbf{x} | K_n\right),$$

where $F(\mathbf{x})$ is non degenerate. Then ADR of $\boldsymbol{\beta}_1^*$ is defined as follows:

$$\mathrm{ADR}\left(\boldsymbol{\beta}_1^*\right) = \mathrm{tr}\left(\mathbf{W} \int_{\mathbb{R}^{p_1}} \int \mathbf{x}\mathbf{x}' dF(\mathbf{x})\right) = \mathrm{tr}(\mathbf{W}\mathbf{V}),$$

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

Asymptotic distributional bias of an estimator $\boldsymbol{\beta}_1^*$ is defined as

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

We make the following two regularity conditions:

(i) $\frac{1}{n} \max_{1\le i\le n} \tilde{\mathbf{x}}_i'(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{x}}_i \to 0$ as $n \to \infty$, where $\tilde{\mathbf{x}}_i'$ is the ith row of $\tilde{\mathbf{X}}$,

(ii) $\frac{1}{n}\sum_{i=1}^n \tilde{\mathbf{X}}'\tilde{\mathbf{X}} \to \tilde{\boldsymbol{Q}}$, where $\tilde{\boldsymbol{Q}}$ is a finite positive-definite matrix.

By virtue of Lemma 1, which is defined at appendix, assumed regularity conditions, and local alternatives, the ADBs of the estimators are:

Theorem 1

$$\begin{aligned}
\mathrm{ADB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right) &= -\boldsymbol{\eta}_{11.2}, \\
\mathrm{ADB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right) &= -\boldsymbol{\xi}, \\
\mathrm{ADB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}}\right) &= -\boldsymbol{\eta}_{11.2} - \boldsymbol{\delta}\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha}; \Delta\right), \\
\mathrm{ADB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right) &= -\boldsymbol{\eta}_{11.2} - (p_2-2)\boldsymbol{\delta}\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right), \\
\mathrm{ADB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}\right) &= -\boldsymbol{\eta}_{11.2} - \boldsymbol{\delta}\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha}; \Delta\right), \\
&\quad -(p_2-2)\boldsymbol{\delta}\mathrm{E}\left\{\chi^{-2}_{p_2+2}(\Delta)\,\mathrm{I}\left(\chi^2_{p_2+2}(\Delta) > p_2-2\right)\right\},
\end{aligned}$$

where $\tilde{\boldsymbol{Q}} = \begin{pmatrix} \tilde{\boldsymbol{Q}}_{11} & \tilde{\boldsymbol{Q}}_{12} \\ \tilde{\boldsymbol{Q}}_{21} & \tilde{\boldsymbol{Q}}_{22} \end{pmatrix}$, $\Delta = \left(\boldsymbol{w}^{\top} \tilde{\boldsymbol{Q}}_{22.1}^{-1} \boldsymbol{w}\right) \sigma^{-2}$, $\tilde{\boldsymbol{Q}}_{22.1} = \tilde{\boldsymbol{Q}}_{22} - \tilde{\boldsymbol{Q}}_{21} \tilde{\boldsymbol{Q}}_{11}^{-1} \tilde{\boldsymbol{Q}}_{12}$, $\boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \end{pmatrix} = -\lambda_0 \tilde{\boldsymbol{Q}}^{-1} \boldsymbol{\beta}$, $\boldsymbol{\eta}_{11.2} = \boldsymbol{\eta}_1 - \tilde{\boldsymbol{Q}}_{12} \tilde{\boldsymbol{Q}}_{22}^{-1} \left((\boldsymbol{\beta}_2 - \boldsymbol{w}) - \boldsymbol{\eta}_2\right)$, $\boldsymbol{\xi} = \boldsymbol{\eta}_{11.2} - \boldsymbol{\delta}$, $\boldsymbol{\delta} = \tilde{\boldsymbol{Q}}_{11}^{-1} \tilde{\boldsymbol{Q}}_{12} \boldsymbol{\omega}$ *and* $\mathrm{H}_v(x, \Delta)$ *be the cumulative distribution function of the non-central chi-squared distribution with non-centrality parameter* Δ *and* v *degree of freedom, and*

$$\mathrm{E}\left(\chi_v^{-2j}(\Delta)\right) = \int_0^{\infty} x^{-2j} d\mathrm{H}_v(x, \Delta).$$

Proof See Appendix. □

Since the bias expressions for all the estimators are not in scaler form, we also convert them to quadratic forms. Thus, we define the asymptotic quadratic distributional bias (AQDB) of an estimator $\boldsymbol{\beta}_1^*$ is

$$\mathrm{AQDB}\left(\boldsymbol{\beta}_1^*\right) = \left(\mathrm{ADB}\left(\boldsymbol{\beta}_1^*\right)\right)' \tilde{\boldsymbol{Q}}_{11.2} \left(\mathrm{ADB}\left(\boldsymbol{\beta}_1^*\right)\right), \tag{9}$$

where $\tilde{\boldsymbol{Q}}_{11.2} = \tilde{\boldsymbol{Q}}_{11} - \tilde{\boldsymbol{Q}}_{12} \tilde{\boldsymbol{Q}}_{22}^{-1} \tilde{\boldsymbol{Q}}_{21}$.

Considering Eq. (9), we present the AQDBs of the estimators as follows:

$$\begin{aligned}
\mathrm{AQDB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right) &= \boldsymbol{\eta}_{11.2}' \tilde{\boldsymbol{Q}}_{11.2} \boldsymbol{\eta}_{11.2}, \\
\mathrm{AQDB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right) &= \boldsymbol{\xi}' \tilde{\boldsymbol{Q}}_{11.2} \boldsymbol{\xi}, \\
\mathrm{AQDB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}}\right) &= \boldsymbol{\eta}_{11.2}' \tilde{\boldsymbol{Q}}_{11.2} \boldsymbol{\eta}_{11.2} + \boldsymbol{\eta}_{11.2}' \tilde{\boldsymbol{Q}}_{11.2} \boldsymbol{\delta} \mathrm{H}_{p_2+2}\left(\chi_{p_2,\alpha}^2; \Delta\right) \\
&\quad + \boldsymbol{\delta}' \tilde{\boldsymbol{Q}}_{11.2} \boldsymbol{\eta}_{11.2} \mathrm{H}_{p_2+2}\left(\chi_{p_2,\alpha}^2; \Delta\right) \\
&\quad + \boldsymbol{\delta}' \tilde{\boldsymbol{Q}}_{11.2} \boldsymbol{\delta} \mathrm{H}_{p_2+2}^2\left(\chi_{p_2,\alpha}^2; \Delta\right), \\
\mathrm{AQDB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right) &= \boldsymbol{\eta}_{11.2}' \tilde{\boldsymbol{Q}}_{11.2} \boldsymbol{\eta}_{11.2} + (p_2 - 2) \boldsymbol{\eta}_{11.2}' \tilde{\boldsymbol{Q}}_{11.2} \boldsymbol{\delta} \mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right) \\
&\quad + (p_2 - 2) \boldsymbol{\delta}' \tilde{\boldsymbol{Q}}_{11.2} \boldsymbol{\eta}_{11.2} \mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right) \\
&\quad + (p_2 - 2)^2 \boldsymbol{\delta}' \tilde{\boldsymbol{Q}}_{11.2} \boldsymbol{\delta} \left(\mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right)\right)^2, \\
\mathrm{AQDB}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}\right) &= \boldsymbol{\eta}_{11.2}' \tilde{\boldsymbol{Q}}_{11.2} \boldsymbol{\eta}_{11.2} + \left(\boldsymbol{\delta}' \tilde{\boldsymbol{Q}}_{11.2} \boldsymbol{\eta}_{11.2} + \boldsymbol{\eta}_{11.2}' \tilde{\boldsymbol{Q}}_{11.2} \boldsymbol{\delta}\right) \\
&\quad \times \left[\mathrm{H}_{p_2+2}\left((p_2 - 2); \Delta\right)\right. \\
&\quad \left. + (p_2 - 2) \mathrm{E}\left\{\chi_{p_2+2}^{-2}(\Delta) \mathrm{I}\left(\chi_{p_2+2}^{-2}(\Delta) > p_2 - 2\right)\right\}\right] \\
&\quad + \boldsymbol{\delta}' \tilde{\boldsymbol{Q}}_{11.2} \boldsymbol{\delta} \left[\mathrm{H}_{p_2+2}\left((p_2 - 2); \Delta\right)\right. \\
&\quad \left. + (p_2 - 2) \mathrm{E}\left\{\chi_{p_2+2}^{-2}(\Delta) \mathrm{I}\left(\chi_{p_2+2}^{-2}(\Delta) > p_2 - 2\right)\right\}\right]^2.
\end{aligned}$$

Assuming that $\tilde{\boldsymbol{Q}}_{12} \neq \mathbf{0}$, then

(i) The AQDB of $\widehat{\boldsymbol{\beta}}_1^{\text{FM}}$ is an constant with $\boldsymbol{\eta}'_{11.2}\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\eta}_{11.2}$.
(ii) The AQDB of $\widehat{\boldsymbol{\beta}}_1^{\text{SM}}$ is an unbounded function of $\boldsymbol{\xi}'\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\xi}$.
(iii) The AQDB of $\widehat{\boldsymbol{\beta}}_1^{\text{PT}}$ begins from $\boldsymbol{\eta}'_{11.2}\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\eta}_{11.2}$ at $\Delta = 0$. For $\Delta > 0$, it increases to a maximum and then decreases towards 0.
(iv) Similarly, the AQDB of $\widehat{\boldsymbol{\beta}}_1^{\text{S}}$ starts from $\boldsymbol{\eta}'_{11.2}\tilde{\boldsymbol{Q}}_{11.2}\boldsymbol{\eta}_{11.2}$ at $\Delta = 0$, and it increases to a point and then decreases towards zero for non-zero Δ values because of $\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right)$ being a non increasing log convex function of Δ. Lastly, for all Δ values, the behaviour of the AQDB of $\widehat{\boldsymbol{\beta}}_1^{\text{PS}}$ is almost the same $\widehat{\boldsymbol{\beta}}_1^{\text{S}}$, but the quadratic bias curve of $\widehat{\boldsymbol{\beta}}_1^{\text{PS}}$ remains on below the curve of $\widehat{\boldsymbol{\beta}}_1^{\text{S}}$.

Now, we present the asymptotic covariance matrices of the proposed estimators which are given by as follows:

Theorem 2

$$
\begin{aligned}
\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}\right) &= \sigma^2\tilde{\boldsymbol{Q}}_{11.2}^{-1} + \boldsymbol{\eta}_{11.2}\boldsymbol{\eta}'_{11.2},\\
\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\text{SM}}\right) &= \sigma^2\tilde{\boldsymbol{Q}}_{11}^{-1} + \boldsymbol{\xi}\boldsymbol{\xi}',\\
\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\text{PT}}\right) &= \sigma^2\tilde{\boldsymbol{Q}}_{11.2}^{-1} + \boldsymbol{\eta}_{11.2}\boldsymbol{\eta}'_{11.2} + 2\boldsymbol{\eta}'_{11.2}\boldsymbol{\delta}\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad + \sigma^2\left(\tilde{\boldsymbol{Q}}_{11.2}^{-1} - \tilde{\boldsymbol{Q}}_{11}^{-1}\right)\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\
&\quad + \boldsymbol{\delta}\boldsymbol{\delta}'\left[2\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) - \mathrm{H}_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)\right],\\
\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\text{S}}\right) &= \sigma^2\tilde{\boldsymbol{Q}}_{11.2}^{-1} + \boldsymbol{\eta}_{11.2}\boldsymbol{\eta}'_{11.2} + 2(p_2-2)\boldsymbol{\delta}\boldsymbol{\eta}'_{11.2}\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right)\\
&\quad - (p_2-2)\sigma^2\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\left\{2\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right)\right.\\
&\quad \left. - (p_2-2)\mathrm{E}\left(\chi^{-4}_{p_2+2}(\Delta)\right)\right\}\\
&\quad + (p_2-2)\boldsymbol{\delta}\boldsymbol{\delta}'\left\{2\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right)\right.\\
&\quad \left. - 2\mathrm{E}\left(\chi^{-2}_{p_2+4}(\Delta)\right) - (p_2-2)\mathrm{E}\left(\chi^{-4}_{p_2+4}(\Delta)\right)\right\},\\
\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\text{PS}}\right) &= \mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\text{S}}\right)\\
&\quad - 2\boldsymbol{\delta}\boldsymbol{\eta}'_{11.2}\mathrm{E}\left(\left\{1-(p_2-2)\chi^{-2}_{p_2+2}(\Delta)\right\}\mathrm{I}\left(\chi^2_{p_2+2}(\Delta) \le p_2-2\right)\right)\\
&\quad + (p_2-2)\sigma^2\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\\
&\quad \times\left[2\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\,\mathrm{I}\left(\chi^2_{p_2+2}(\Delta) \le p_2-2\right)\right)\right.\\
&\quad \left. - (p_2-2)\mathrm{E}\left(\chi^{-4}_{p_2+2}(\Delta)\,\mathrm{I}\left(\chi^2_{p_2+2}(\Delta) \le p_2-2\right)\right)\right]\\
&\quad - \sigma^2\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\mathrm{H}_{p_2+2}\left((p_2-2);\Delta\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\boldsymbol{\delta}\boldsymbol{\delta}'\left[2\mathrm{H}_{p_2+2}\left((p_2-2);\Delta\right)-\mathrm{H}_{p_2+4}\left((p_2-2);\Delta\right)\right]\\
&-(p_2-2)\boldsymbol{\delta}\boldsymbol{\delta}'\Big[2\mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\,\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&-2\mathrm{E}\left(\chi_{p_2+4}^{-2}(\Delta)\,\mathrm{I}\left(\chi_{p_2+4}^{2}(\Delta)\le p_2-2\right)\right)\\
&+(p_2-2)\mathrm{E}\left(\chi_{p_2+2}^{-4}(\Delta)\,\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\Big].
\end{aligned}
$$

Proof See Appendix. □

Finally, we obtain the ADRs of the estimators under $\{K_n\}$ given as:

Theorem 3

$$
\begin{aligned}
\mathrm{ADR}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right)&=\sigma^2\mathrm{tr}\left(\boldsymbol{W}\tilde{\boldsymbol{Q}}_{11.2}^{-1}\right)+\boldsymbol{\eta}_{11.2}'\boldsymbol{W}\boldsymbol{\eta}_{11.2},\\
\mathrm{ADR}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)&=\sigma^2\,\mathrm{tr}\left(\boldsymbol{W}\tilde{\boldsymbol{Q}}_{11}^{-1}\right)+\boldsymbol{\xi}'\boldsymbol{W}\boldsymbol{\xi},\\
\mathrm{ADR}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}}\right)&=\mathrm{ADR}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right)-2\boldsymbol{\eta}_{11.2}'\boldsymbol{W}\boldsymbol{\delta}\mathrm{H}_{p_2+2}\left(\chi_{p_2,\alpha}^2;\Delta\right)\\
&\quad-\sigma^2\mathrm{tr}\left(\boldsymbol{W}\tilde{\boldsymbol{Q}}_{11.2}^{-1}-\boldsymbol{W}\tilde{\boldsymbol{Q}}_{11}^{-1}\right)\mathrm{H}_{p_2+2}\left(\chi_{p_2,\alpha}^2;\Delta\right)\\
&\quad+\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}\left\{2\mathrm{H}_{p_2+2}\left(\chi_{p_2,\alpha}^2;\Delta\right)-\mathrm{H}_{p_2+4}\left(\chi_{p_2,\alpha}^2;\Delta\right)\right\},\\
\mathrm{ADR}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)&=\mathrm{ADR}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right)+2(p_2-2)\boldsymbol{\eta}_{11.2}'\boldsymbol{W}\boldsymbol{\delta}\mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right)\\
&\quad-(p_2-2)\sigma^2\mathrm{tr}\left(\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\boldsymbol{W}\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\right)\Big\{2\mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right)\\
&\quad-(p_2-2)\mathrm{E}\left(\chi_{p_2+2}^{-4}(\Delta)\right)\Big\}\\
&\quad+(p_2-2)\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}\Big\{2\mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right)\\
&\quad-2\mathrm{E}\left(\chi_{p_2+4}^{-2}(\Delta)\right)-(p_2-2)\mathrm{E}\left(\chi_{p_2+4}^{-4}(\Delta)\right)\Big\},\\
\mathrm{ADR}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}\right)&=\mathrm{ADR}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)\\
&\quad-2\boldsymbol{\eta}_{11.2}'\boldsymbol{W}\boldsymbol{\delta}\mathrm{E}\left(\left\{1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta)\right\}\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad+(p_2-2)\sigma^2\mathrm{tr}\left(\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\boldsymbol{W}\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\right)\\
&\quad\times\Big[2\mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\,\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad-(p_2-2)\mathrm{E}\left(\chi_{p_2+2}^{-4}(\Delta)\,\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\Big]\\
&\quad-\sigma^2\mathrm{tr}\left(\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\boldsymbol{W}\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\right)\mathrm{H}_{p_2+2}\left((p_2-2);\Delta\right)\\
&\quad+\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}\left[2\mathrm{H}_{p_2+2}\left((p_2-2);\Delta\right)-\mathrm{H}_{p_2+4}\left((p_2-2);\Delta\right)\right]\\
&\quad-(p_2-2)\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}\Big[2\mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\,\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad-2\mathrm{E}\left(\chi_{p_2+4}^{-2}(\Delta)\,\mathrm{I}\left(\chi_{p_2+4}^{2}(\Delta)\le p_2-2\right)\right)
\end{aligned}
$$

$$+(p_2-2)\mathrm{E}\left(\chi^{-4}_{p_2+2}(\Delta)\,\mathrm{I}\left(\chi^{2}_{p_2+2}(\Delta)\le p_2-2\right)\right)\Big].$$

Proof See Appendix. □

If $\tilde{\boldsymbol{Q}}_{12}=\mathbf{0}$, then $\boldsymbol{\delta}=\mathbf{0}$, $\boldsymbol{\xi}=\boldsymbol{\eta}_{11.2}$ and $\tilde{\boldsymbol{Q}}_{11.2}=\tilde{\boldsymbol{Q}}_{11}$, all the ADRs reduce to common value $\sigma^2\mathrm{tr}\left(\mathbf{W}\tilde{\boldsymbol{Q}}_{11}^{-1}\right)+\boldsymbol{\eta}'_{11.2}\mathbf{W}\boldsymbol{\eta}_{11.2}$ for all $\boldsymbol{\omega}$. On the other hand, assuming $\tilde{\boldsymbol{Q}}_{12}\neq 0$, then

(i) As Δ moves away from 0, the ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)$ becomes unbounded. Furthermore, the ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}}\right)$ perform better than ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right)$ for all values of $\Delta\ge 0$, that is ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}}\right)\le$ ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right)$.

(ii) For all $\mathbf{W}$ and $\boldsymbol{\omega}$, ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)\le$ ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right)$, if

$$\frac{\mathrm{tr}\left(\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\mathbf{W}\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\right)}{ch_{max}\left(\tilde{\boldsymbol{Q}}_{21}\tilde{\boldsymbol{Q}}_{11}^{-1}\mathbf{W}\tilde{\boldsymbol{Q}}_{11}^{-1}\tilde{\boldsymbol{Q}}_{12}\tilde{\boldsymbol{Q}}_{22.1}^{-1}\right)}\ge\frac{p_2+2}{2},$$

where $ch_{max}(\cdot)$ is the maximum characteristic root.

(iii) To compare $\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}$ and $\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}$, we observe that the ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}\right)$ overshadows ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)$ for all the values of $\boldsymbol{\omega}$. Moreover, with result (ii), we have ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}\right)\le$ ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)\le$ ADR $\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right)$ all $\mathbf{W}$ and $\boldsymbol{\omega}$.

5 Simulation studies

In this section, we consider a Monte Carlo simulation to evaluate the relative quadratic risk performance of the listed estimators. All calculations were carried out in R Development Core Team (2010). We simulate the response from the following model:

$$y_i=x_{1i}\beta_1+x_{2i}\beta_2+\cdots+x_{pi}\beta_p+f(t_i)+\varepsilon_i\ ,\ i=1,\ldots,n, \tag{10}$$

where $\mathbf{x}_i\sim\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_x)$ and ε_i are i.i.d. $\mathcal{N}(0,1)$. We also define $\boldsymbol{\Sigma}_x$ that is positive definite covariance matrix. The off-diagonal elements of the covariance matrix $\boldsymbol{\Sigma}_x$ are considered to be equal to ρ with $\rho=0.25, 0.5, 0.75$. The condition number (CN) is used to test the multicollinearity, which is defined as the ratio of the largest eigenvalue to the smallest eigenvalue of matrix $\mathbf{X}'\mathbf{X}$. If CN is larger than 30, then it implies the existence of multicollinearity in the data set, Belsley (1991). We also get $\alpha=0.05$ and $t_i=(i-0.5)/n$. Furthermore, we consider the hypothesis $\mathrm{H}_0:\beta_j=0$, for $j=p_1+1,p_1+2,\ldots,p$, with $p=p_1+p_2$. Hence, we partition the regression coefficients as $\boldsymbol{\beta}=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)=(\boldsymbol{\beta}_1,\mathbf{0})$ with $\boldsymbol{\beta}_1=(1,1,1,1,1)$. In (10), we consider two different the nonparametric functions $f_1(t_i)=\sqrt{t_i(1-t_i)}\sin\left(\frac{2.1\pi}{t_i+0.05}\right)$ and $f_2(t_i)=0.5\sin(4\pi t_i)$ to generate response y_i.

Table 1 RMSEs for $n = 50$, $p_1 = 5$, $p_2 = 10$ and $\rho = 0.25$

Δ^*	CN	$\widehat{\beta}_1$	$\widehat{\beta}_1^{\text{SM}}$	$\widehat{\beta}_1^{\text{PT}}$	$\widehat{\beta}_1^{\text{S}}$	$\widehat{\beta}_1^{\text{PS}}$
0.00		0.924	1.442	1.350	1.216	1.340
0.25		0.869	1.369	1.212	1.165	1.296
0.50		0.920	1.223	0.989	1.263	1.271
0.75		0.903	1.040	0.928	1.176	1.176
1.00	33.807	0.957	0.887	1.000	1.154	1.154
1.25		0.875	0.618	1.000	1.097	1.097
1.50		0.866	0.506	1.000	1.060	1.060
2.00		0.851	0.329	1.000	1.032	1.032
4.00		0.935	0.128	1.000	1.007	1.007

In literature, there are a number of studies about bandwidth selection for a PLRM. Some recent studies are: Li et al. (2011) provide a theoretical justification for the earlier empirical observations of an optimal zone of bandwidths. Further, Li and Palta (2009) introduced a bandwidth selection for semi-parametric varying-coefficient. In our study, we use generalized cross-validation (GCV) to select the optimal λ value for given k. By Wahba (1990), the GCV score function can be procured by

$$\text{GCV}(\lambda, k) = \frac{n \left\|\mathbf{y} - \widehat{\mathbf{y}}\right\|^2}{\{\text{tr}(\mathbf{I}_n - \mathbf{H}_\lambda)\}^2}.$$

For further information about selection of the optimal ridge parameter and the optimal bandwidth, we refer to Amini and Roozbeh (2015) and Roozbeh (2015).

Each realization was repeated 5000 times. We define $\Delta^* = \left\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\right\|$, where $\boldsymbol{\beta}_0 = \left(\boldsymbol{\beta}_1, \mathbf{0}\right)$, and $\|\cdot\|$ is the Euclidean norm. In order to investigate of the behaviour of the estimators for $\Delta^* > 0$, further datasets were generated from those distributions under local alternative hypothesis.

The performance of an estimator was evaluated by using mean squared error (MSE). In order to easy comparison, we also calculate the relative mean squared efficiency (RMSE) of the $\boldsymbol{\beta}_1^{\blacktriangledown}$ to the $\widehat{\boldsymbol{\beta}}_1^{\text{FM}}$ is given by

$$\text{RMSE}\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}} : \boldsymbol{\beta}_1^{\blacktriangledown}\right) = \frac{\text{MSE}\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}\right)}{\text{MSE}\left(\boldsymbol{\beta}_1^{\blacktriangledown}\right)},$$

where $\boldsymbol{\beta}_1^{\blacktriangledown}$ is one of the suggested estimators. If the RMSE of an estimator is larger than one, it is superior to the full model estimator. Results are reported briefly in Table 1, and plotted to easier comparison in Figs. 1 and 2.

We summary the results as follows:

(i) When $\Delta^* = 0$, SM outperforms all the other estimators. On the other hand, after the small interval near $\Delta^* = 0$, the RMSE of $\widehat{\boldsymbol{\beta}}_1^{\text{SM}}$ decreases and goes to zero.

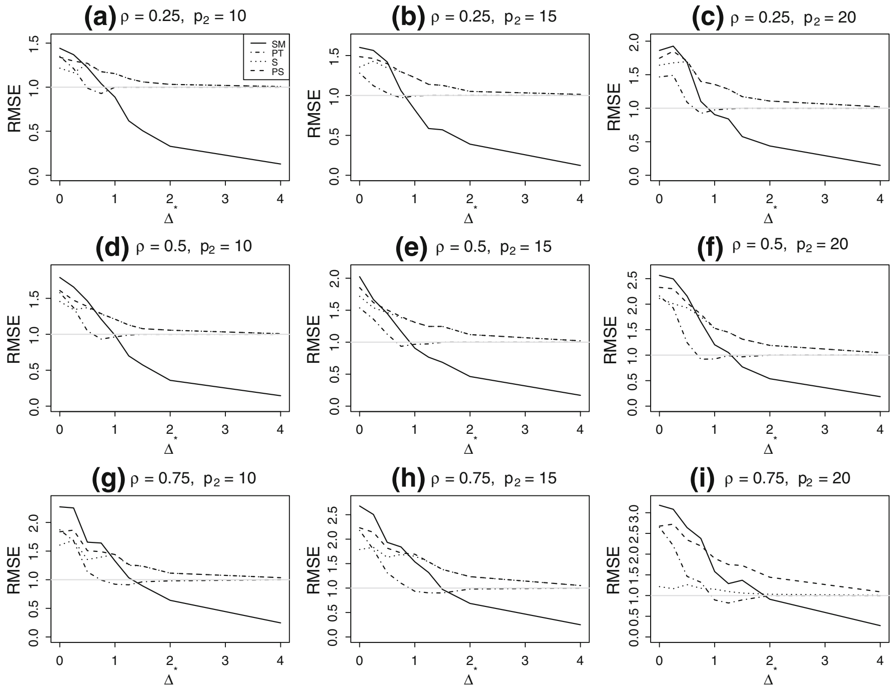

Fig. 1 Relative efficiency of the estimators as a function of Δ^* for $n = 50$

(ii) The ordinary least squares (OLS) estimator $\widehat{\boldsymbol{\beta}}_1$ performs much worse than ridge-type suggested estimators when ρ is large.

(iii) For large p_2, the CN increases, whereas the RMSE of $\widehat{\boldsymbol{\beta}}_1^{\text{FM}}$ decreases, the RMSE of $\widehat{\boldsymbol{\beta}}_1^{\text{SM}}$ increases.

(iv) The PT outperforms shrinkage ridge regression estimators at $\Delta^* = 0$ when p_1 and p_2 close to each other. But, for large p_2 values, $\widehat{\boldsymbol{\beta}}_1^{\text{PS}}$ has biggest RMSE. As Δ^* increases, the RMSE of $\widehat{\boldsymbol{\beta}}_1^{\text{PT}}$ decreases, and it remains on below 1, and then it increases and approaches one.

(v) It is seen that the RMSE of $\widehat{\boldsymbol{\beta}}_1^{\text{S}}$ is smaller than the RMSE of $\widehat{\boldsymbol{\beta}}_1^{\text{PS}}$ for all Δ^* values.

(vi) Overall, our results are consistent with the studies of Ahmed et al. (2007); Raheem et al. (2012).

In Table 2, we show the results the comparison the suggested estimators with penalty estimators. From the simulation results, the SM outperforms all other estimators. We observe that ridge pretest and ridge shrinkage estimators perform better than penalty estimators when both ρ and p_2 are large. Especially, when ρ is large, performance of penalty estimators decrease, whereas the performance of ridge pretest and ridge shrinkage estimators increase. Therefore, the OLS performs much worse than ridge-type suggested estimators, since covariates are designed to be correlated.

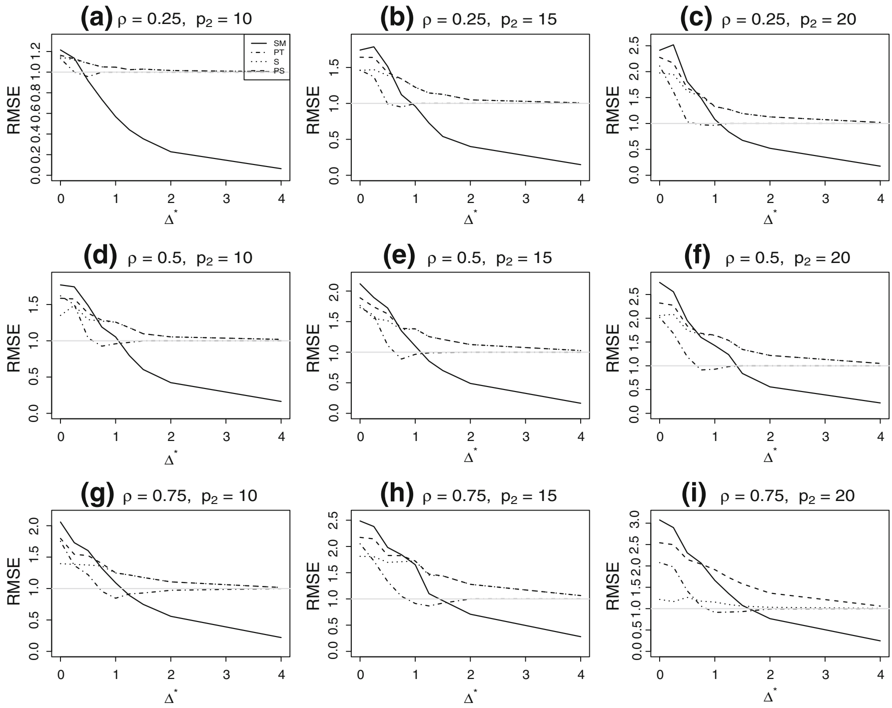

Fig. 2 Relative efficiency of the estimators as a function of Δ^* for $n = 100$

In Fig. 3, we plotted estimations of the nonparametric functions f_1 and f_2. The curves estimated by smoothing spline denoted a similar behaviours to real functions especially for larger sample size.

6 Application

We implement proposed strategies to the Baseball data which is analyzed by Friendly (2002). The data contains 322 rows and 22 variables. We also omit missing values and four covariates which are not scaler. Hence, we have 263 sample and 17 covariates, and Table 3 lists the detailed descriptions of both the dependent variable and covariates.

We calculated the CN value is 5830 which implies the existence of multicollinearity in the data set.

The chosen variables via BIC and AIC are shown in Table 4, and AIC selects a model with more variables than BIC. So, in Table 5, the full- and sub- models are given. As it can be seen in Table 5, we omit the intercept term in this analysis since this term was very close to zero in calculations.

In Table 5, as stated previously, f denotes a smooth function. To select the covariate which can be modelled non-parametrically, we used White Neural Network test (see *tseries* package in R) for nonlinearity of each of the covariates. According to the results

Table 2 CNs and RMSEs for $\Delta^* = 0$

(n, p_1)	ρ	p_2	CN	$\widehat{\beta}_1$	$\widehat{\beta}_1^{SM}$	$\widehat{\beta}_1^{PT}$	$\widehat{\beta}_1^{S}$	$\widehat{\beta}_1^{PS}$	$\widehat{\beta}_1^{Lasso}$	$\widehat{\beta}_1^{aLasso}$	$\widehat{\beta}_1^{SCAD}$
(100,5)	0.25	10	16.906	0.979	1.293	1.255	1.219	1.249	1.001	1.101	0.957
		15	21.330	0.982	1.435	1.376	1.283	1.366	1.012	1.171	1.001
		20	32.509	1.092	1.534	1.460	1.412	1.517	1.119	1.348	1.215
	0.5	10	24.492	0.917	1.091	1.107	1.123	1.157	1.098	1.076	0.849
		15	51.180	1.028	1.687	1.570	1.531	1.602	1.086	1.152	0.881
		20	118.785	1.071	1.880	1.651	1.626	1.760	1.121	1.230	0.998
	0.75	10	103.700	0.870	1.876	1.738	1.444	1.637	0.942	0.835	0.657
		15	129.756	0.974	2.096	1.755	1.712	1.839	0.999	0.875	0.704
		20	226.138	0.999	2.591	2.097	1.880	2.232	1.095	0.970	0.734
(100,10)	0.25	10	19.494	0.916	1.150	1.104	1.114	1.122	0.992	0.995	0.926
		15	42.708	0.931	1.357	1.258	1.250	1.291	0.996	1.042	0.979
		20	46.569	0.941	1.387	1.336	1.342	1.361	1.044	1.122	1.060
	0.5	10	56.776	0.879	1.453	1.388	1.333	1.336	0.989	0.914	0.818
		15	106.365	0.901	1.553	1.405	1.434	1.447	1.032	0.995	0.810
		20	129.272	0.870	1.648	1.496	1.521	1.541	1.053	1.051	0.801
	0.75	10	241.912	0.755	1.955	1.833	1.704	1.677	0.843	0.524	0.756
		15	331.234	0.783	2.244	1.980	1.871	1.919	0.867	0.595	0.727
		20	455.707	0.804	2.462	2.035	2.069	2.101	0.901	0.647	0.714

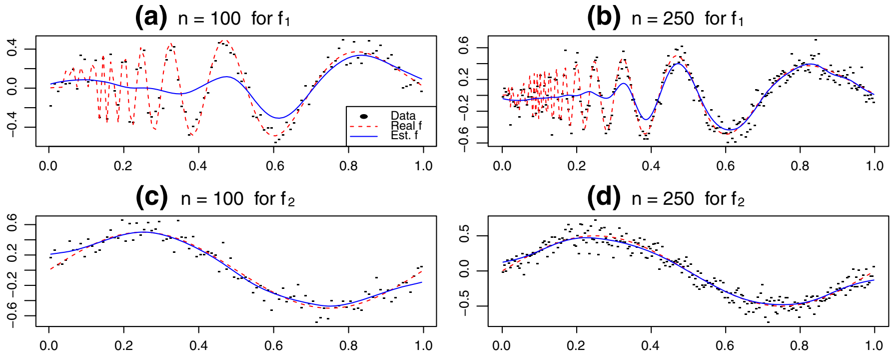

Fig. 3 Estimation of non-parametric functions for $p_1 = 5$ and $p_2 = 10$

Table 3 List of Variable

Variable	Description
Dependent Variable	
*ln*Sal	The logarithm of annual salary (in thousands) on opening day 1987
Covariates	
Atbat	Number of times at bat in 1986
Hits	Number of hits in 1986
Homer	Number of home runs in 1986
Runs	Number of runs in 1986
RBI	Batted in during 1986
Walks	Number of walks in 1986
Years	Number of years in the major leagues
Atbatc	Number of times at bat in his career
Hitsc	Number of hits in career
Homerc	Number of home runs in career
Runsc	Number of runs in career
RBIc	Number of Runs Batted In in career
Walksc	Number of walks in career
Putouts	Number of putouts in 1986
Assists	Number of assists in 1986
Errors	Number of errors in 1986

Table 4 Candidate sub-models

Methods	Chosen variables
AIC	Atbat, Runs, Walks, Years, Atbatc, Hitsc, Homerc, Assists, Errors
BIC	Atbat, Years, Atbatc, Hitsc, Homerc, Assists, Errors

Table 5 Fitting models

Models	Formula
Full model	$ln\text{Sal} = \beta_1\text{Atbat} + \beta_2\text{Hits} + \beta_3\text{Homer} + \beta_4\text{Runs} + \beta_5\text{RBI} + \beta_6\text{Walks} + \beta_7\text{Atbatc} + \beta_8\text{Hitsc} + \beta_9\text{Homerc} + \beta_{10}\text{Runsc} + \beta_{11}\text{RBIc} + \beta_{12}\text{Walksc} + \beta_{13}\text{Putouts} + \beta_{14}\text{Assists} + \beta_{15}\text{Errors} + f(\text{Years})$
Sub-model(AIC)	$ln\text{Sal} = \beta_1\text{Atbat} + \beta_2\text{Runs} + \beta_3\text{Walks} + \beta_4\text{Atbatc} + \beta_5\text{Hitsc} + \beta_6\text{Homerc} + \beta_7\text{Assists} + \beta_8\text{Errors} + f(\text{Years})$
Sub-model(BIC)	$ln\text{Sal} = \beta_1\text{Atbat} + \beta_2\text{Atbatc} + \beta_3\text{Hitsc} + \beta_4\text{Homerc} + \beta_5\text{Assists} + \beta_6\text{Errors} + f(\text{Years})$

Table 6 Relative prediction errors

Estimators	AIC				BIC				Lasso	aLasso	SCAD
	SM	PT	S	PS	SM	PT	S	PS			
RPE	1.501	1.405	1.411	1.443	1.475	1.377	1.450	1.452	1.407	1.388	1.356

of this test, we have found that the Years has a significant nonlinear relationship with the response lnSal.

To evaluate the performance of each method, we obtain prediction errors by using 10-fold cross validation following 999 resampled bootstrap samples. Further, we also calculate the Relative Prediction Error (RPE) of each method with respect to the full model estimator. If the RPE of any estimator is larger than one, then this indicates the superiority of that method over the full model estimator. The results are shown in Table 6. According to these results, not surprisingly the SM has maximum RPE since this estimator is computed based on the true model. Further, shrinkage methods outperform penalty estimators although pretest method may less efficient. Finally, we may suggest to use BIC method to construct suggested techniques.

In Table 7, we present the coefficients of parametric part of model. Moreover, the curve estimated by smoothing spline which the smoothing parameter is selected by GCV is shown in Fig. 4.

7 Conclusion

In this paper, we suggest pretest, shrinkage and penalty estimation for PLRMs. The parametric components is estimated by using ridge regression approach, the non-parametric component is estimated by using Speckman approach based on penalized residual sum of squares method. The advantages of listed estimators are studied both theoretically and numerically. Our results show that the sub-model estimator outperforms shrinkage and penalty estimators when the null hypothesis is true, i.e., $\Delta^* = 0$. Moreover, the pretest and shrinkage estimators performs better than the full model estimator. On the other hand, as the restriction moves away from $\Delta^* = 0$, i.e., the assumption of null hypothesis is violated, the efficiency of sub-model estimator gradually decreases even worse than full model estimator. Also, the pretest estimation may

Table 7 Estimation of parametric coefficients

Variable	FM	SM(AIC)	PT(AIC)	S(AIC)	PS(AIC)	SM(BIC)	PT(BIC)	S(BIC)	PS(BIC)	Lasso	aLasso	SCAD
Atbat	0.0205	−0.0511	−0.0249	−0.0165	−0.0132	0.1333	0.0572	0.0621	0.0620	0	0	0
Hits	0.0240	0	0.0107	0.0120	0.0131	0	0.0176	0.0157	0.0157	0.0286	0	0.0449
Homer	0.0169	0	0.0076	0.0085	0.0093	0	0.0123	0.0109	0.0109	0.0325	0	0.0454
Runs	0.0238	0.1407	0.0981	0.0891	0.0813	0	0.0174	0.0155	0.0155	0.0432	0	0
RBI	0.0206	0	0.0091	0.0103	0.0112	0	0.0150	0.0133	0.0133	0	0	0
Walks	0.0211	0.0940	0.0666	0.0587	0.0560	0	0.0157	0.0138	0.0139	0.0571	0.0539	0.0656
Atbatc	0.0778	−0.7611	−0.4415	−0.3687	−0.3232	−0.6979	−0.1796	−0.2169	−0.2157	0	0	0
Hitsc	0.0781	13.747	0.8795	0.7641	0.6990	13.284	0.4905	0.5509	0.5490	0.5292	0.5896	0
Homerc	0.0289	0.1662	0.1140	0.1034	0.0952	0.1999	0.0860	0.0949	0.0946	0	0	0
Runsc	0.0663	0	0.0296	0.0334	0.0364	0	0.0487	0.0431	0.0432	0.0146	0	0
RBIc	0.0594	0	0.0267	0.0299	0.0326	0	0.0434	0.0385	0.0385	0.1451	0.0806	0
Walksc	0.0381	0	0.0168	0.0195	0.0209	0	0.0283	0.0249	0.0250	0	0	0
Putouts	0.0196	0	0.0110	0.0115	0.0119	0	0.0157	0.0133	0.0133	0.0829	0.0839	0.0892
Assists	0.0006	0.0005	0.0039	0.0010	0.0012	−0.0179	−0.0054	−0.0063	−0.0063	0	0	0
Errors	0.0023	0.0027	−0.0018	0.0002	0.0005	−0.0082	−0.0031	−0.0020	−0.0021	0	0	0

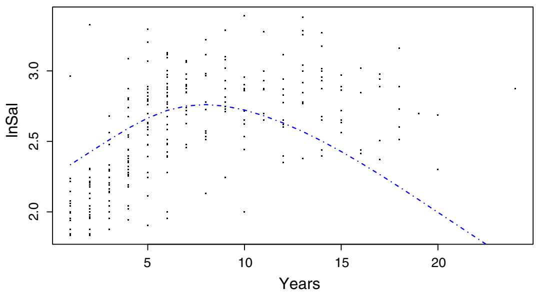

Fig. 4 Graph of the estimation of nonparametric function

not perform well for little violations of null hypothesis while the performance of it act like full model when the violations of null hypothesis is large. Finally, shrinkage estimation outperforms the full model estimator in every case. We also compare listed estimators with penalty estimators through Monte Carlo simulation. Our asymptotic theory is well supported by numerical analysis. In summary, construction estimators outshine penalty estimators when p_2 is large, and these estimators much more consistent than penalty estimators in the presence of multicollinearity.

Acknowledgements The authors thank the editor and two reviewers for their detailed reading of the manuscript and their valuable comments and suggestions that led to a considerable improvement of the paper. Research of Professor Bahadır Yüzbaşı is supported by The Scientific and Research Council of Turkey under grant Tubitak-Bideb-2214/A during this study at Brock University in Canada. Research of Professor S. Ejaz Ahmed is supported by the Natural Sciences and the Engineering Research Council of Canada (NSERC).

Appendix

We present the following two lemmas below, which will enable us to derive the results of Theorems 1 and 3 in this paper

Lemma 1 *If* $k/\sqrt{n} \to \lambda_0 \geq 0$ *and* $\tilde{\boldsymbol{Q}}$ *is non-singular, then*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{\text{FM}} - \boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}\left(-\lambda_0 \tilde{\boldsymbol{Q}}^{-1}\boldsymbol{\beta}, \sigma^2 \tilde{\boldsymbol{Q}}^{-1}\right),$$

where "$\xrightarrow{d}$" denotes convergence in distribution.

Proof Let define $V_n(\mathbf{u})$ as follows:

$$\sum_{i=1}^{n}\left[\left(\tilde{\varepsilon}_i - \mathbf{u}'\tilde{\mathbf{x}}_i/\sqrt{n}\right)^2 - \tilde{\varepsilon}_i^2\right] + k\sum_{j=1}^{p}\left[\left|\beta_j + u_j/\sqrt{n}\right|^2 - \left|\beta_j\right|^2\right],$$

where $\mathbf{u} = (u_1, \ldots, u_p)'$. Following Knight and Fu (2000), it can be shown that

$$\sum_{i=1}^{n}\left[\left(\tilde{\varepsilon}_i - \mathbf{u}'\tilde{\mathbf{x}}_i/\sqrt{n}\right)^2 - \tilde{\varepsilon}_i^2\right] \xrightarrow{d} -2\mathbf{u}'\mathbf{D} + \mathbf{u}'\tilde{\boldsymbol{Q}}\mathbf{u},$$

where $\mathbf{D} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_p)$, with finite-dimensional convergence holding trivially. Hence,

$$k\sum_{j=1}^{p}\left[\left|\beta_j + u_j/\sqrt{n}\right|^2 - \left|\beta_j\right|^2\right] \xrightarrow{d} \lambda_0 \sum_{j=1}^{p} u_j \text{sgn}(\beta_j)|\beta_j|.$$

Hence, $V_n(\mathbf{u}) \xrightarrow{d} V(\mathbf{u})$. Because V_n is convex and V has a unique minimum, by following Geyer (1996), it yields

$$\arg\min(V_n) = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{\text{FM}} - \boldsymbol{\beta}\right) \xrightarrow{d} \arg\min(V).$$

Hence,

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{\text{FM}} - \boldsymbol{\beta}\right) \xrightarrow{d} \tilde{\boldsymbol{Q}}^{-1}(\mathbf{D} - \lambda_0\boldsymbol{\beta}) \sim \mathcal{N}\left(-\lambda_0\tilde{\boldsymbol{Q}}^{-1}\boldsymbol{\beta}, \sigma^2\tilde{\boldsymbol{Q}}^{-1}\right).$$

□

Lemma 2 *Let* $\mathbf{X}$ *be* $q-$*dimensional normal vector distributed as* $\mathcal{N}\left(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_q\right)$, *then, for a measurable function of* φ, *we have*

$$\begin{aligned}\mathrm{E}\left[\mathbf{X}\varphi\left(\mathbf{X}'\mathbf{X}\right)\right] &= \boldsymbol{\mu}_x \mathrm{E}\left[\varphi\chi^2_{q+2}(\Delta)\right]\\ \mathrm{E}\left[\mathbf{X}\mathbf{X}'\varphi\left(\mathbf{X}'\mathbf{X}\right)\right] &= \boldsymbol{\Sigma}_q \mathrm{E}\left[\varphi\chi^2_{q+2}(\Delta)\right] + \boldsymbol{\mu}_x\boldsymbol{\mu}_x' \mathrm{E}\left[\varphi\chi^2_{q+4}(\Delta)\right]\end{aligned}$$

where $\chi^2_v(\Delta)$ *is a non-central chi-square distribution with* v *degrees of freedom and non-centrality parameter* Δ.

Proof It can be found in Judge and Bock (1978) □

We further consider the following proposition for proving theorems.

Proposition 1 *Under local alternative* $\{K_n\}$ *as* $n \to \infty$, *we have*

$$\begin{pmatrix}\vartheta_1\\ \vartheta_3\end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix}-\boldsymbol{\eta}_{11.2}\\ \boldsymbol{\delta}\end{pmatrix}, \begin{pmatrix}\sigma^2\tilde{\boldsymbol{Q}}^{-1}_{11.2} & \boldsymbol{\Phi}_*\\ \boldsymbol{\Phi}_* & \boldsymbol{\Phi}_*\end{pmatrix}\right],$$
$$\begin{pmatrix}\vartheta_3\\ \vartheta_2\end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix}\boldsymbol{\delta}\\ -\boldsymbol{\xi}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Phi}_* & \mathbf{0}\\ \mathbf{0} & \sigma^2\tilde{\boldsymbol{Q}}^{-1}_{11}\end{pmatrix}\right],$$

where $\vartheta_1 = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}} - \boldsymbol{\beta}_1\right)$, $\vartheta_2 = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\text{SM}} - \boldsymbol{\beta}_1\right)$ and $\vartheta_3 = \vartheta_1 - \vartheta_2$.

Proof Under the light of Lemmas 1 and 2, it can easily be obtained

$$\vartheta_1 \xrightarrow{d} \mathcal{N}\left(-\boldsymbol{\eta}_{11.2}, \sigma^2\tilde{\boldsymbol{Q}}^{-1}_{11.2}\right).$$

Define $\mathbf{y}^* = \tilde{\mathbf{y}} - \tilde{\mathbf{X}}_2 \widehat{\boldsymbol{\beta}}_2^{\mathrm{FM}}$, and

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}} &= \underset{\boldsymbol{\beta}_1}{\arg\min} \left\{ \left\| \mathbf{y}^* - \tilde{\mathbf{X}}_1 \boldsymbol{\beta}_1 \right\| + k \left\| \boldsymbol{\beta}_1 \right\|^2 \right\} \\
&= \left(\tilde{\mathbf{X}}_1' \tilde{\mathbf{X}}_1 + k \mathbf{I}_{p_1} \right)^{-1} \tilde{\mathbf{X}}_1' \mathbf{y}^* \\
&= \left(\tilde{\mathbf{X}}_1' \tilde{\mathbf{X}}_1 + k \mathbf{I}_{p_1} \right)^{-1} \tilde{\mathbf{X}}_1' \tilde{\mathbf{y}} - \left(\tilde{\mathbf{X}}_1' \tilde{\mathbf{X}}_1 + k \mathbf{I}_{p_1} \right)^{-1} \tilde{\mathbf{X}}_1' \tilde{\mathbf{X}}_2 \widehat{\boldsymbol{\beta}}_2^{\mathrm{FM}} \\
&= \widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}} - \left(\tilde{\mathbf{X}}_1' \tilde{\mathbf{X}}_1 + k \mathbf{I}_{p_1} \right)^{-1} \tilde{\mathbf{X}}_1' \tilde{\mathbf{X}}_2 \widehat{\boldsymbol{\beta}}_2^{\mathrm{FM}}.
\end{aligned} \tag{11}$$

By using Eq. (11),

$$\begin{aligned}
\mathrm{E}\left\{ \lim_{n\to\infty} \sqrt{n} \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}} - \boldsymbol{\beta}_1 \right) \right\} &= \mathrm{E}\left\{ \lim_{n\to\infty} \sqrt{n} \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}} + \tilde{\boldsymbol{Q}}_{11}^{-1} \tilde{\boldsymbol{Q}}_{12} \widehat{\boldsymbol{\beta}}_2^{\mathrm{FM}} - \boldsymbol{\beta}_1 \right) \right\} \\
&= \mathrm{E}\left\{ \lim_{n\to\infty} \sqrt{n} \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}} - \boldsymbol{\beta}_1 \right) \right\} \\
&\quad + \mathrm{E}\left\{ \lim_{n\to\infty} \sqrt{n} \left(\tilde{\boldsymbol{Q}}_{11}^{-1} \tilde{\boldsymbol{Q}}_{12} \widehat{\boldsymbol{\beta}}_2^{\mathrm{FM}} \right) \right\}
\end{aligned}$$

by Lemma 2,

$$\begin{aligned}
&= -\boldsymbol{\eta}_{11.2} + \tilde{\boldsymbol{Q}}_{11}^{-1} \tilde{\boldsymbol{Q}}_{12} \boldsymbol{\omega} \\
&= -\left(\boldsymbol{\eta}_{11.2} - \boldsymbol{\delta} \right) \\
&= -\boldsymbol{\xi}.
\end{aligned}$$

Hence, $\vartheta_2 \xrightarrow{d} \mathcal{N}\left(-\boldsymbol{\xi}, \sigma^2 \tilde{\boldsymbol{Q}}_{11}^{-1} \right)$.
Using the Eq. (11), we can obtain $\boldsymbol{\Phi}_*$ as follows:

$$\begin{aligned}
\boldsymbol{\Phi}_* &= \mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}} \right) \\
&= \mathrm{E}\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}} \right) \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}} - \widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}} \right)' \right] \\
&= \mathrm{E}\left[\left(\tilde{\boldsymbol{Q}}_{11}^{-1} \tilde{\boldsymbol{Q}}_{12} \widehat{\boldsymbol{\beta}}_2^{\mathrm{FM}} \right) \left(\tilde{\boldsymbol{Q}}_{11}^{-1} \tilde{\boldsymbol{Q}}_{12} \widehat{\boldsymbol{\beta}}_2^{\mathrm{FM}} \right)' \right] \\
&= \tilde{\boldsymbol{Q}}_{11}^{-1} \tilde{\boldsymbol{Q}}_{12} \mathrm{E}\left[\widehat{\boldsymbol{\beta}}_2^{\mathrm{FM}} \left(\widehat{\boldsymbol{\beta}}_2^{\mathrm{FM}} \right)' \right] \tilde{\boldsymbol{Q}}_{21} \tilde{\boldsymbol{Q}}_{11}^{-1} \\
&= \sigma^2 \tilde{\boldsymbol{Q}}_{11}^{-1} \tilde{\boldsymbol{Q}}_{12} \tilde{\boldsymbol{Q}}_{22.1}^{-1} \tilde{\boldsymbol{Q}}_{21} \tilde{\boldsymbol{Q}}_{11}^{-1}.
\end{aligned}$$

We also know that

$$\boldsymbol{\Phi}_* = \sigma^2 \tilde{\boldsymbol{Q}}_{11}^{-1} \tilde{\boldsymbol{Q}}_{12} \tilde{\boldsymbol{Q}}_{22.1}^{-1} \tilde{\boldsymbol{Q}}_{21} \tilde{\boldsymbol{Q}}_{11}^{-1} = \sigma^2 \left(\tilde{\boldsymbol{Q}}_{11.2}^{-1} - \tilde{\boldsymbol{Q}}_{11}^{-1} \right).$$

Hence, it is obtained $\vartheta_3 \xrightarrow{d} \mathcal{N}\left(\boldsymbol{\delta}, \boldsymbol{\Phi}_* \right)$. □

Proof (Theorem 1) ADB $\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}\right)$ and ADB $\left(\widehat{\boldsymbol{\beta}}_1^{\text{SM}}\right)$ are directly obtained from Proposition 1. Also, the ADBs of PT, S and PS are obtained as follows:

$$\begin{aligned}
\text{ADB}\left(\widehat{\boldsymbol{\beta}}_1^{\text{PT}}\right) &= \text{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\text{PT}}-\boldsymbol{\beta}_1\right)\right\}\\
&= \text{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}-\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}-\widehat{\boldsymbol{\beta}}_1^{\text{SM}}\right)\text{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)-\boldsymbol{\beta}_1\right)\right\}\\
&= \text{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}-\boldsymbol{\beta}_1\right)\right\}\\
&\quad -\text{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}-\widehat{\boldsymbol{\beta}}_1^{\text{SM}}\right)\text{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)\right)\right\}\\
&= -\boldsymbol{\eta}_{11.2}-\boldsymbol{\delta}\text{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right).\\
\text{ADB}\left(\widehat{\boldsymbol{\beta}}_1^{\text{S}}\right) &= \text{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\text{S}}-\boldsymbol{\beta}_1\right)\right\}\\
&= \text{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}-\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}-\widehat{\boldsymbol{\beta}}_1^{\text{SM}}\right)(p_2-2)\,\mathcal{T}_n^{-1}-\boldsymbol{\beta}_1\right)\right\}\\
&= \text{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}-\boldsymbol{\beta}_1\right)\right\}\\
&\quad -\text{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}-\widehat{\boldsymbol{\beta}}_1^{\text{SM}}\right)(p_2-2)\,\mathcal{T}_n^{-1}\right)\right\}\\
&= -\boldsymbol{\eta}_{11.2}-(p_2-2)\,\boldsymbol{\delta}\text{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right).\\
\text{ADB}\left(\widehat{\boldsymbol{\beta}}_1^{\text{PS}}\right) &= \text{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\text{PS}}-\boldsymbol{\beta}_1\right)\right\}\\
&= \text{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\text{SM}}+\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}-\widehat{\boldsymbol{\beta}}_1^{\text{SM}}\right)\right.\right.\\
&\quad \left.\left.\times\left(1-(p_2-2)\,T_n^{-1}\right)\text{I}\left(\mathcal{T}_n>p_2-2\right)-\boldsymbol{\beta}_1\right)\right\}\\
&= \text{E}\left\{\lim_{n\to\infty}\sqrt{n}\left[\widehat{\boldsymbol{\beta}}_1^{\text{SM}}+\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}-\widehat{\boldsymbol{\beta}}_1^{\text{SM}}\right)\left(1-\text{I}\left(\mathcal{T}_n\le p_2-2\right)\right)\right.\right.\\
&\quad \left.\left.-\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}-\widehat{\boldsymbol{\beta}}_1^{\text{SM}}\right)(p_2-2)\,\mathcal{T}_n^{-1}\text{I}\left(\mathcal{T}_n>p_2-2\right)-\boldsymbol{\beta}_1\right]\right\}\\
&= \text{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}-\boldsymbol{\beta}_1\right)\right\}\\
&\quad -\text{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}-\widehat{\boldsymbol{\beta}}_1^{\text{SM}}\right)\text{I}\left(\mathcal{T}_n\le p_2-2\right)\right\}\\
&\quad -\text{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\text{FM}}-\widehat{\boldsymbol{\beta}}_1^{\text{SM}}\right)(p_2-2)\,\mathcal{T}_n^{-1}\text{I}\left(\mathcal{T}_n>p_2-2\right)\right\}\\
&= -\boldsymbol{\eta}_{11.2}-\boldsymbol{\delta}\text{H}_{p_2+2}\left(p_2-2;(\Delta)\right)\\
&\quad -\boldsymbol{\delta}(p_2-2)\,\text{E}\left\{\chi^{-2}_{p_2+2}(\Delta)\,\text{I}\left(\chi^2_{p_2+2}(\Delta)>p_2-2\right)\right\}.
\end{aligned}$$

□

The asymptotic covariance of an estimator $\boldsymbol{\beta}_1^*$ is defined as follows:

$$\text{Cov}\left(\boldsymbol{\beta}_1^*\right)=\text{E}\left\{\lim_{n\to\infty}n\left(\boldsymbol{\beta}_1^*-\boldsymbol{\beta}_1\right)\left(\boldsymbol{\beta}_1^*-\boldsymbol{\beta}_1\right)'\right\}.$$

Proof (Theorem 2) Firstly, the asymptotic covariance of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}$ is given by

$$\begin{aligned}
\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}\right) &= \mathrm{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\boldsymbol{\beta}_1\right)\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\boldsymbol{\beta}_1\right)'\right\}\\
&= \mathrm{E}\left(\vartheta_1\vartheta_1'\right)\\
&= \mathrm{Cov}\left(\vartheta_1\vartheta_1'\right)+\mathrm{E}\left(\vartheta_1\right)\mathrm{E}\left(\vartheta_1'\right)\\
&= \sigma^2\tilde{\boldsymbol{Q}}_{11.2}^{-1}+\boldsymbol{\eta}_{11.2}\boldsymbol{\eta}_{11.2}'.
\end{aligned}$$

The asymptotic covariance of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}$ is given by

$$\begin{aligned}
\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right) &= \mathrm{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}-\boldsymbol{\beta}_1\right)\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}-\boldsymbol{\beta}_1\right)'\right\}\\
&= \mathrm{E}\left(\vartheta_2\vartheta_2'\right)\\
&= \mathrm{Cov}\left(\vartheta_2\vartheta_2'\right)+\mathrm{E}\left(\vartheta_2\right)\mathrm{E}\left(\vartheta_2'\right)\\
&= \sigma^2\tilde{\boldsymbol{Q}}_{11}^{-1}+\boldsymbol{\xi}\boldsymbol{\xi}',
\end{aligned}$$

The asymptotic covariance of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}}$ is given by

$$\begin{aligned}
\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}}\right) &= \mathrm{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}}-\boldsymbol{\beta}_1\right)\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}}-\boldsymbol{\beta}_1\right)'\right\}\\
&= \mathrm{E}\left\{\lim_{n\to\infty} n\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\boldsymbol{\beta}_1\right)-\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)\mathrm{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)\right]\right.\\
&\qquad\left.\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\boldsymbol{\beta}_1\right)-\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)\mathrm{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)\right]'\right\}\\
&= \mathrm{E}\left\{\left[\vartheta_1-\vartheta_3\mathrm{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)\right]\left[\vartheta_1-\vartheta_3\mathrm{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)\right]'\right\}\\
&= \mathrm{E}\left\{\vartheta_1\vartheta_1'-2\vartheta_3\vartheta_1'\mathrm{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)+\vartheta_3\vartheta_3'\mathrm{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)\right\}.
\end{aligned}$$

Now, by using Lemma 2 and the formula for a conditional mean of a bivariate normal, we have

$$\begin{aligned}
\mathrm{E}\left\{\vartheta_3\vartheta_1'\mathrm{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)\right\} &= \mathrm{E}\left\{\mathrm{E}\left(\vartheta_3\vartheta_1'\mathrm{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)|\vartheta_3\right)\right\}\\
&= \mathrm{E}\left\{\vartheta_3\mathrm{E}\left(\vartheta_1'\mathrm{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)|\vartheta_3\right)\right\}\\
&= \mathrm{E}\left\{\vartheta_3\left[-\eta_{11.2}+\left(\vartheta_3-\boldsymbol{\delta}\right)\right]'\mathrm{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)\right\}\\
&= -\mathrm{E}\left\{\vartheta_3\boldsymbol{\eta}_{11.2}'\mathrm{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)\right\}+\\
&\quad\ \mathrm{E}\left\{\vartheta_3\left(\vartheta_3-\boldsymbol{\delta}\right)'\mathrm{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)\right\}\\
&= -\boldsymbol{\eta}_{11.2}'\mathrm{E}\left\{\vartheta_3\mathrm{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)\right\}\\
&\quad +\mathrm{E}\left\{\vartheta_3\vartheta_3'\mathrm{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)\right\}\\
&\quad -\mathrm{E}\left\{\vartheta_3\boldsymbol{\delta}'\mathrm{I}\left(\mathcal{T}_n\le c_{n,\alpha}\right)\right\}\\
&= -\boldsymbol{\eta}_{11.2}'\boldsymbol{\delta}\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)+\left\{\mathrm{Cov}(\vartheta_3\vartheta_3')\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\right.
\end{aligned}$$

$$+\mathrm{E}\left(\vartheta_3\right)\mathrm{E}\left(\vartheta_3'\right)\mathrm{H}_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)-\boldsymbol{\delta}\boldsymbol{\delta}'\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\Big\}$$
$$=-\boldsymbol{\eta}_{11.2}'\boldsymbol{\delta}\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)+\boldsymbol{\Phi}_*\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)$$
$$+\boldsymbol{\delta}\boldsymbol{\delta}'\mathrm{H}_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)-\boldsymbol{\delta}\boldsymbol{\delta}'\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right),$$

then,

$$\begin{aligned}\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PT}}\right)&=\boldsymbol{\eta}_{11.2}\boldsymbol{\eta}_{11.2}'+2\boldsymbol{\eta}_{11.2}'\boldsymbol{\delta}\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\&\quad\sigma^2\tilde{\boldsymbol{Q}}_{11.2}^{-1}-\boldsymbol{\Phi}_*\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};(\Delta)\right)-\boldsymbol{\delta}\boldsymbol{\delta}'\mathrm{H}_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)\\&\quad+2\boldsymbol{\delta}\boldsymbol{\delta}'\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\&=\sigma^2\tilde{\boldsymbol{Q}}_{11.2}^{-1}+\boldsymbol{\eta}_{11.2}\boldsymbol{\eta}_{11.2}'+2\boldsymbol{\eta}_{11.2}'\boldsymbol{\delta}\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\&\quad+\sigma^2\left(\tilde{\boldsymbol{Q}}_{11.2}^{-1}-\tilde{\boldsymbol{Q}}_{11}^{-1}\right)\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)\\&\quad+\boldsymbol{\delta}\boldsymbol{\delta}'\left[2\mathrm{H}_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)-\mathrm{H}_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)\right].\end{aligned}$$

The asymptotic covariance of $\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}$ is given by

$$\begin{aligned}\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)&=\mathrm{E}\left\{\lim_{n\to\infty}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}-\boldsymbol{\beta}_1\right)\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}-\boldsymbol{\beta}_1\right)'\right\}\\&=\mathrm{E}\left\{\lim_{n\to\infty}n\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\boldsymbol{\beta}_1\right)-\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)(p_2-2)\mathcal{T}_n^{-1}\right]\right.\\&\quad\left.\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\boldsymbol{\beta}_1\right)-\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)(p_2-2)\mathcal{T}_n^{-1}\right]'\right\}\\&=\mathrm{E}\left\{\vartheta_1\vartheta_1'-2(p_2-2)\vartheta_3\vartheta_1'\mathcal{T}_n^{-1}+(p_2-2)^2\vartheta_3\vartheta_3'\mathcal{T}_n^{-2}\right\}.\end{aligned}$$

Note that, by using Lemma 2 and the formula for a conditional mean of a bivariate normal, we have

$$\begin{aligned}\mathrm{E}\left\{\vartheta_3\vartheta_1'\mathcal{T}_n^{-1}\right\}&=\mathrm{E}\left\{\mathrm{E}\left(\vartheta_3\vartheta_1'\mathcal{T}_n^{-1}|\vartheta_3\right)\right\}\\&=\mathrm{E}\left\{\vartheta_3\mathrm{E}\left(\vartheta_1'\mathcal{T}_n^{-1}|\vartheta_3\right)\right\}\\&=\mathrm{E}\left\{\vartheta_3\left[-\boldsymbol{\eta}_{11.2}+(\vartheta_3-\boldsymbol{\delta})\right]'\mathcal{T}_n^{-1}\right\}\\&=-\mathrm{E}\left\{\vartheta_3\boldsymbol{\eta}_{11.2}'\mathcal{T}_n^{-1}\right\}+\mathrm{E}\left\{\vartheta_3(\vartheta_3-\boldsymbol{\delta})'\mathcal{T}_n^{-1}\right\}\\&=-\boldsymbol{\eta}_{11.2}'\mathrm{E}\left\{\vartheta_3T_n^{-1}\right\}+\mathrm{E}\left\{\vartheta_3\vartheta_3'\mathcal{T}_n^{-1}\right\}\\&\quad-\mathrm{E}\left\{\vartheta_3\boldsymbol{\delta}'\mathcal{T}_n^{-1}\right\}\\&=-\boldsymbol{\eta}_{11.2}'\boldsymbol{\delta}\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right)+\left\{\mathrm{Cov}(\vartheta_3\vartheta_3')\mathrm{E}\left(\chi^{-2}_{p_2+2}(\Delta)\right)\right.\end{aligned}$$

$$
\begin{aligned}
&+\mathrm{E}\left(\vartheta_3\right)\mathrm{E}\left(\vartheta_3'\right)\mathrm{E}\left(\chi_{p_2+4}^{-2}(\Delta)\right)-\boldsymbol{\delta}\boldsymbol{\delta}'\mathrm{H}_{p_2+2}\left(\chi_{p_2,\alpha}^{2};\Delta\right)\Big\}\\
&=-\boldsymbol{\eta}_{11.2}'\boldsymbol{\delta}\mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right)+\boldsymbol{\Phi}_*\mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right)\\
&\quad+\boldsymbol{\delta}\boldsymbol{\delta}'\mathrm{E}\left(\chi_{p_2+4}^{-2}(\Delta)\right)-\boldsymbol{\delta}\boldsymbol{\delta}'\mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right),
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)&=\sigma^2\tilde{\boldsymbol{Q}}_{11.2}^{-1}+\boldsymbol{\eta}_{11.2}\boldsymbol{\eta}_{11.2}'+2\left(p_2-2\right)\boldsymbol{\eta}_{11.2}'\boldsymbol{\delta}\mathrm{E}\left(\chi_{p_2+2,\alpha}^{-2}(\Delta)\right)\\
&\quad-\left(p_2-2\right)\boldsymbol{\Phi}_*\left\{2\mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right)-\left(p_2-2\right)\mathrm{E}\left(\chi_{p_2+2}^{-4}(\Delta)\right)\right\}\\
&\quad+\left(p_2-2\right)\boldsymbol{\delta}\boldsymbol{\delta}'\left\{-2\mathrm{E}\left(\chi_{p_2+4}^{-2}(\Delta)\right)+2\mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\right)\right.\\
&\quad\left.+\left(p_2-2\right)\mathrm{E}\left(\chi_{p_2+4}^{-4}(\Delta)\right)\right\}.
\end{aligned}
$$

Finally, the asymptotic covariance matrix of positive shrinkage ridge regression estimator is derived as follows:

$$
\begin{aligned}
\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}\right)&=\mathrm{E}\left\{\lim_{n\to\infty}n\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}-\boldsymbol{\beta}_1\right)\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}-\boldsymbol{\beta}_1\right)'\right\}\\
&=\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)-2\mathrm{E}\left\{\lim_{n\to\infty}\sqrt{n}\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}-\boldsymbol{\beta}_1\right)'\right.\right.\\
&\quad\left.\left.\times\left\{1-\left(p_2-2\right)T_n^{-1}\right\}\mathrm{I}\left(\mathcal{T}_n\le p_2-2\right)\right]\right\}\\
&\quad+\mathrm{E}\left\{\lim_{n\to\infty}\sqrt{n}\left[\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{FM}}-\widehat{\boldsymbol{\beta}}_1^{\mathrm{SM}}\right)'\right.\right.\\
&\quad\left.\left.\times\left\{1-\left(p_2-2\right)T_n^{-1}\right\}^2\mathrm{I}\left(\mathcal{T}_n\le p_2-2\right)\right]\right\}\\
&=\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)-2\mathrm{E}\left\{\vartheta_3\vartheta_1'\left\{1-\left(p_2-2\right)\mathcal{T}_n^{-1}\right\}\mathrm{I}\left(\mathcal{T}_n\le p_2-2\right)\right\}\\
&\quad+2\mathrm{E}\left\{\vartheta_3\vartheta_3'\left(p_2-2\right)\mathcal{T}_n^{-1}\mathrm{I}\left(\mathcal{T}_n\le p_2-2\right)\right\}\\
&\quad-2\mathrm{E}\left\{\vartheta_3\vartheta_3'\left(p_2-2\right)^2\mathcal{T}_n^{-2}\mathrm{I}\left(\mathcal{T}_n\le p_2-2\right)\right\}\\
&\quad+\mathrm{E}\left\{\vartheta_3\vartheta_3'\mathrm{I}\left(\mathcal{T}_n\le p_2-2\right)\right\}\\
&\quad-2\mathrm{E}\left\{\vartheta_3\vartheta_3'\left(p_2-2\right)\mathcal{T}_n^{-1}\mathrm{I}\left(\mathcal{T}_n\le p_2-2\right)\right\}\\
&\quad+\mathrm{E}\left\{\vartheta_3\vartheta_3'\left(p_2-2\right)^2\mathcal{T}_n^{-2}\mathrm{I}\left(\mathcal{T}_n\le p_2-2\right)\right\}\\
&=\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)-2\mathrm{E}\left\{\vartheta_3\vartheta_1'\left\{1-\left(p_2-2\right)\mathcal{T}_n^{-1}\right\}\mathrm{I}\left(\mathcal{T}_n\le p_2-2\right)\right\}\\
&\quad-\mathrm{E}\left\{\vartheta_3\vartheta_3'\left(p_2-2\right)^2\mathcal{T}_n^{-2}\mathrm{I}\left(\mathcal{T}_n\le p_2-2\right)\right\}\\
&\quad+\mathrm{E}\left\{\vartheta_3\vartheta_3'\mathrm{I}\left(\mathcal{T}_n\le p_2-2\right)\right\}.
\end{aligned}
$$

Based on Lemma 2 and the formula for a conditional mean of a bivariate normal, we have

$$\begin{aligned}
&\mathrm{E}\left\{\vartheta_3\vartheta_1'\left\{1-(p_2-2)\,\mathcal{T}_n^{-1}\right\}\mathrm{I}\,(\mathcal{T}_n\le p_2-2)\right\}\\
&=\mathrm{E}\left\{\mathrm{E}\left(\vartheta_3\vartheta_1'\left\{1-(p_2-2)\,\mathcal{T}_n^{-1}\right\}\mathrm{I}\,(\mathcal{T}_n\le p_2-2)\,|\vartheta_3\right)\right\}\\
&=\mathrm{E}\left\{\vartheta_3\mathrm{E}\left(\vartheta_1'\left\{1-(p_2-2)\,\mathcal{T}_n^{-1}\right\}\mathrm{I}\,(\mathcal{T}_n\le p_2-2)\,|\vartheta_3\right)\right\}\\
&=\mathrm{E}\left\{\vartheta_3\left[-\boldsymbol{\eta}_{11.2}+(\vartheta_3-\boldsymbol{\delta})\right]'\left\{1-(p_2-2)\,\mathcal{T}_n^{-1}\right\}\mathrm{I}\,(\mathcal{T}_n\le p_2-2)\right\}\\
&=-\boldsymbol{\eta}_{11.2}\mathrm{E}\left(\vartheta_3\left\{1-(p_2-2)\,\mathcal{T}_n^{-1}\right\}\mathrm{I}\,(\mathcal{T}_n\le p_2-2)\right)\\
&\quad+\mathrm{E}\left(\vartheta_3\vartheta_3'\left\{1-(p_2-2)\,\mathcal{T}_n^{-1}\right\}\mathrm{I}\,(\mathcal{T}_n\le p_2-2)\right)\\
&\quad-\mathrm{E}\left(\vartheta_3\boldsymbol{\delta}'\left\{1-(p_2-2)\,\mathcal{T}_n^{-1}\right\}\mathrm{I}\,(\mathcal{T}_n\le p_2-2)\right)\\
&=-\boldsymbol{\delta}\boldsymbol{\eta}_{11.2}'\boldsymbol{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta)\right\}\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad+\boldsymbol{\Phi}_*\mathrm{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta)\right\}\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad+\boldsymbol{\delta}\boldsymbol{\delta}'\mathrm{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+4}^{-2}(\Delta)\right\}\mathrm{I}\left(\chi_{p_2+4}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad-\boldsymbol{\delta}\boldsymbol{\delta}'\mathrm{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta)\right\}\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right),
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{PS}}\right)&=\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)+2\boldsymbol{\delta}\boldsymbol{\eta}_{11.2}'\mathrm{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta)\right\}\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad-2\boldsymbol{\Phi}_*\mathrm{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta)\right\}\mathrm{I}\left(\chi_{p_2+2}^{-2}(\Delta)\le p_2-2\right)\right)\\
&\quad-2\boldsymbol{\delta}\boldsymbol{\delta}'\mathrm{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+4}^{-2}(\Delta)\right\}\mathrm{I}\left(\chi_{p_2+4}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad+2\boldsymbol{\delta}\boldsymbol{\delta}'\mathrm{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta)\right\}\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad-(p_2-2)^2\,\boldsymbol{\Phi}_*\mathrm{E}\left(\chi_{p_2+2,\alpha}^{-4}(\Delta)\,\mathrm{I}\left(\chi_{p_2+2,\alpha}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad-(p_2-2)^2\,\boldsymbol{\delta}\boldsymbol{\delta}'\mathrm{E}\left(\chi_{p_2+4}^{-4}(\Delta)\,\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad+\boldsymbol{\Phi}_*\mathrm{H}_{p_2+2}(p_2-2;\Delta)+\boldsymbol{\delta}\boldsymbol{\delta}'\mathrm{H}_{p_2+4}(p_2-2;\Delta)\\
&=\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{S}}\right)+2\boldsymbol{\delta}\boldsymbol{\eta}_{11.2}'\mathrm{E}\left(\left\{1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta)\right\}\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\\
&\quad+(p_2-2)\,\sigma^2\,\tilde{\boldsymbol{Q}}_{11}^{-1}\,\tilde{\boldsymbol{Q}}_{12}\,\tilde{\boldsymbol{Q}}_{22.1}^{-1}\,\tilde{\boldsymbol{Q}}_{21}\,\tilde{\boldsymbol{Q}}_{11}^{-1}\\
&\quad\times\left[2\mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\,\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\right.\\
&\quad\left.-(p_2-2)\,\mathrm{E}\left(\chi_{p_2+2}^{-4}(\Delta)\,\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\right]\\
&\quad-\sigma^2\,\tilde{\boldsymbol{Q}}_{11}^{-1}\,\tilde{\boldsymbol{Q}}_{12}\,\tilde{\boldsymbol{Q}}_{22.1}^{-1}\,\tilde{\boldsymbol{Q}}_{21}\,\tilde{\boldsymbol{Q}}_{11}^{-1}\mathrm{H}_{p_2+2}(p_2-2;\Delta)\\
&\quad+\boldsymbol{\delta}\boldsymbol{\delta}'\left[2\mathrm{H}_{p_2+2}(p_2-2;\Delta)-\mathrm{H}_{p_2+4}(p_2-2;\Delta)\right]\\
&\quad-(p_2-2)\,\boldsymbol{\delta}\boldsymbol{\delta}'\left[2\mathrm{E}\left(\chi_{p_2+2}^{-2}(\Delta)\,\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\le p_2-2\right)\right)\right.
\end{aligned}$$

$$-2\mathrm{E}\left(\chi_{p_2+4}^{-2}(\Delta)\,\mathrm{I}\left(\chi_{p_2+4}^{2}(\Delta)\leq p_2-2\right)\right)$$
$$+\,(p_2-2)\,\mathrm{E}\left(\chi_{p_2+2}^{-4}(\Delta)\,\mathrm{I}\left(\chi_{p_2+2}^{2}(\Delta)\leq p_2-2\right)\right)\Bigg].$$

□

Proof (Theorem 3) The asymptotic risks of the estimators can be derived by following the definition of ADR

$$\begin{aligned}\mathrm{ADR}\left(\boldsymbol{\beta}_1^*\right)&=n\mathrm{E}\left[\left(\boldsymbol{\beta}_1^*-\boldsymbol{\beta}_1\right)'\mathbf{W}\left(\boldsymbol{\beta}_1^*-\boldsymbol{\beta}_1\right)\right]\\&=n\mathrm{tr}\left[\mathbf{W}\mathrm{E}\left(\boldsymbol{\beta}_1^*-\boldsymbol{\beta}_1\right)\left(\boldsymbol{\beta}_1^*-\boldsymbol{\beta}_1\right)'\right]\\&=\mathrm{tr}\left(\mathbf{W}\mathrm{Cov}\left(\boldsymbol{\beta}_1^*\right)\right).\end{aligned}$$

□

References

Ahmed SE (2014) Penalty, shrinkage and pretest strategies: variable selection and estimation. Springer, New York

Ahmed SE, Doksum KA, Hossain S, You J (2007) Shrinkage, pretest and absolute penalty estimators in partially linear models. Aust N Z J Stat 49(4):435–454

Amini M, Roozbeh M (2015) Optimal partial ridge estimation in restricted semi-parametric regression models. J Multivar Anal 136:26–40

Arashi M, Valizadeh T (2015) Performance of Kibria's methods in partial linear ridge regression model. Stat Papers 56(1):231–246

Aydın D (2014) Estimation of partially linear model with smoothing spline based on different selection methods: A comparative study. Pak J Stat 30:35–56

Belsley DA (1991) Conditioning diagnostics. Wiley, New York

Engle RF, Granger CWJ, Rice CA, Weiss A (1986) Semi-parametric estimates of the relation between weather and electricity sales. J Am Stat Assoc 81:310–320

Eubank RL (1986) A note on smoothness priors and nonlinear regression. J Am Stat Assoc 81(394):514–517

Eubank RL, Kambour EL, Kim TC, Kipple K, Reese SC, Schimek M (1998) Estimation in partially linear models. Comput Stat Data Anal 29:27–34

Friendly M (2002) Corrgrams: exploratory displays for correlation matrices. Am Stat 56:316–324

Green PJ, Silverman BW (1994) Nonparametric regression and generalized linear model. Chapman & Hall, Boca Raton

Green PJ, Jennison C, Seheult A (1985) Analysis of field experiments by least square smoothing. J R Stat Soc B 47:299–315

Geyer CJ (1996) On the asymptotics of convex stochastic optimization. Unpublished manuscript

Hastie TJ, Tibshirani RJ (1990) Generalized additive models, vol 43. CRC Press, Boca Raton

Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for non-orthogonal problems. Technometrics 12:69–82

Judge GG, Bock ME (1978) The statistical implications of pre-test and stein-rule estimators in econometrics. North Holland, Amsterdam

Knight K, Fu W (2000) Asymptotics for Lasso-type estimators. Ann Stat 28(5):1356–1378

Liang H (2006) Estimation partially linear models and numerical comparison. Comput Stat Data Anal 50:675–687

Li J, Palta M (2009) Bandwidth selection through cross-validation for semi-parametric varying-coefficient partially linear models. J Stat Comput Simul 79:1277–1286

Li J, Zhang W, Wu Z (2011) Optimal zone for bandwidth selection in semi-parametric models. J Nonparametr Stat 23(3):701–717

R Development Core Team (2010) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0
Raheem SE, Ahmed SE, Doksum KA (2012) Absolute penalty and shrinkage estimation in partially linear models. Comput Stat Data Anal 56(4):874–891
Rice J (1986) Convergence rates for partially spline models. Stat Prob Lett 4:203–208
Roozbeh M (2015) Shrinkage ridge estimators in semi-parametric regression models. J Multivariate Anal 136:56–74
Roozbeh M, Arashi M (2013) Feasible ridge estimator in partially linear models. J Multivariate Anal 116:35–44
Schimek GM (2000) Estimation and inference in partially linear models with smoothing splines. J Stat Plann Inference 91(2):525–540
Shiller RJ (1984) Smoothness priors and nonlinear regression. J Am Stat Assoc 79(387):609–615
Speckman P (1988) Kernel smoothing in partially linear model. J R Stat Soc B 50:413–436
Yüzbaşı B, Ahmed SE (2016) Shrinkage and penalized estimation in semi-parametric models with multicollinear data. J Stat Comput Simul. https://doi.org/10.1080/00949655.2016.1171868
Wahba G (1990) Spline model for observational data. SIAM, Philadelphia, PA
Wu J, Asar Y (2016) A weighted stochastic restricted ridge estimator in partially linear model. Commun Stat Theory Methods. https://doi.org/10.1080/03610926.2016.1206936