Selection of Model Selection Criteria for Multivariate Ridge Regression

(Last Modified: January, 24)

Isamu NAGAI[1],

[1] *Department of Mathematics, Graduate School of Science, Hiroshima University*
1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8626, Japan
E-mail: d093481@hiroshima-u.ac.jp

Abstract

In the present study, we consider the selection of model selection criteria for multivariate ridge regression. There are several model selection criteria for selecting the ridge parameter in multivariate ridge regression, e.g., the C_p criterion and the modified C_p (MC_p) criterion. We propose the generalized C_p (GC_p) criterion, which includes C_p and MC_p criteria as special cases. The GC_p criterion is specified by a non-negative parameter λ, which is referred to as the penalty parameter. We attempt to select an optimal penalty parameter such that the predictive mean square error (PMSE) of the predictor of ridge regression after optimizing the ridge parameter is minimized. Through numerical experiments, we verify that the proposed optimization methods exhibit better performance than conventional optimization methods, i.e., optimizing only the ridge parameter by minimizing the C_p or MC_p criterion.

Key words: Asymptotic expansion; Generalized C_p criterion; Model selection criterion; Multivariate linear regression model; Ridge regression; Selection of the model selection criterion.

1. Introduction

In the present paper, we deal with a multivariate linear regression model with n observations of a p-dimensional vector of response variables and a k-dimensional vector of regressors (for more detailed information, see, for example, Srivastava, 2002, Chapter 9; Timm, 2002, Chapter 4). Let $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)'$, $\boldsymbol{X}$, and $\boldsymbol{\mathcal{E}} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n)'$ be the $n \times p$ matrix of response variables, the $n \times k$ matrix of non-stochastic centerized explanatory variables (i.e., $\boldsymbol{X}'\boldsymbol{1}_n = \boldsymbol{0}_k$) of $\mathrm{rank}(\boldsymbol{X}) = k$ $(< n)$, and the $n \times p$ matrix of error variables, respectively, where n is the sample size, $\boldsymbol{1}_n$ is an n-dimensional vector of ones, and $\boldsymbol{0}_k$ is a k-dimensional vector of zeros. Suppose that $n - k - p - 2 > 0$ and $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n \overset{\text{i.i.d.}}{\sim} N_p(\boldsymbol{0}_p, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$

is a $p \times p$ unknown covariance matrix. Then, the matrix form of the multivariate linear regression model is expressed as

$$\boldsymbol{Y} = \boldsymbol{1}_n \boldsymbol{\mu}' + \boldsymbol{X}\Xi + \boldsymbol{\mathcal{E}},$$

where $\boldsymbol{\mu}$ is a p-dimensional unknown location vector, and Ξ is a $k \times p$ unknown regression coefficient matrix. This model can also be expressed as

$$\boldsymbol{Y} \sim N_{n\times p}(\boldsymbol{1}_n \boldsymbol{\mu}' + \boldsymbol{X}\Xi, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n).$$

Note that $\boldsymbol{X}$ is centerized. The maximum likelihood or the least squares (LS) estimators of $\boldsymbol{\mu}$ and Ξ are given by $\hat{\boldsymbol{\mu}} = \boldsymbol{Y}'\boldsymbol{1}_n/n$ and $\hat{\Xi} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$, respectively. Since $\hat{\boldsymbol{\mu}}$ and $\hat{\Xi}$ are simple, and the unbiased estimators of $\boldsymbol{\mu}$ and Ξ, LS estimators are widely used in actual data analysis (see, e.g., Dien *et al.*, 2006; Sârbu *et al.*, 2008, Saxén and Sundell, 2006; Skagerberg, Macgregor, and Kiparissides, 1992; Yoshimoto, Yanagihara, and Ninomiya, 2005). However, the problem of an estimator of Ξ becoming unstable when multicollinearity occurs in $\boldsymbol{X}$ is well known. In order to avoid this problem, a ridge regression was proposed by Hoerl and Kennard (1970) when $p = 1$. Several studies extended this univariate ridge regression to the multivariate case, e.g., Brown and Zidek (1980), Haitovsky (1987), and Yanagihara and Satoh (2010). The ridge regression estimator of Ξ is given as

$$\hat{\Xi}_\theta = \boldsymbol{M}_\theta^{-1}\boldsymbol{X}'\boldsymbol{Y},$$

where $\boldsymbol{M}_\theta = \boldsymbol{X}'\boldsymbol{X} + \theta\boldsymbol{I}_k$, and θ is a nonnegative value, which is referred to as a ridge parameter. Since an estimate of $\hat{\Xi}_\theta$ depends strongly on the value of the ridge parameter θ, the optimization of θ is an important problem in ridge regression.

An optimal θ is commonly determined by minimizing the predicted mean square error (PMSE) of the predictor of $\boldsymbol{Y}$, i.e., $\hat{\boldsymbol{Y}}_\theta = \boldsymbol{1}_n\hat{\boldsymbol{\mu}}' + \boldsymbol{X}\hat{\Xi}_\theta$. However, we cannot directly use the PMSE to optimize θ, because unknown parameters are included in the PMSE. Hence, we adopt an optimization method using a model selection criterion, i.e., an estimator of PMSE, instead of the unknown PMSE. As an estimator of PMSE, Yanagihara and Satoh (2010) proposed a C_p criterion. This criterion includes C_p criteria for selecting variables in a univariate linear model, which was proposed by Mallows (1973; 1995), for selecting variables in a multivariate linear model, which was proposed by Sparks, Coutsourides, and Troskie (1983) as a special case. Yanagihara and Satoh (2010) also proposed the modified C_p (MC_p) criterion such that the bias of the C_p criterion for choosing the ridge parameter for PMSE is completely corrected under a fixed θ. This criterion coincides with the bias-corrected C_p criterion proposed by Fujikoshi and Satoh (1997) when $\theta = 0$. The MC_p criterion has several desirable properties as the estimator of PMSE as described by, e.g., Fujikoshi, Yanagihara, and Wakaki (2005), and Yanagihara and Satoh (2010).

Unfortunately, optimizing θ by minimizing MC_p, i.e., an unbiased estimator of PMSE, does not always minimize the PMSE of $\hat{\boldsymbol{Y}}_\theta$. This indicates that there will be an optimal model selection criterion for selecting θ. Thus, we propose a generalized C_p (GC_p) criterion that includes the C_p and MC_p criteria as special cases (originally, the GC_p criterion was proposed by Atkinson, 1980, for selecting variables in the univariate linear model). The GC_p criterion is specified by a non-negative parameter λ, which is referred to as the penalty parameter. From the viewpoint of making the PMSE of the predictor of $\boldsymbol{Y}$ after optimizing θ small, we select the optimal penalty parameter λ, which is basically the selection of the model selection criterion. In the present paper, we optimize λ by the following three methods:

- (Double optimization): We optimize θ and λ simultaneously by minimizing GC_p and the penalty selection criteria, respectively.
- (Optimization of λ with an approximated value of an optimal θ): We optimize λ by minimizing the penalty selection criterion made from the approximated value of optimal θ.
- (Asymptotic optimization of λ): We calculate an asymptotic optimal λ from an asymptotic expansion of the PMSE. We then estimate the asymptotic optimal λ.

From the optimization of the model selection criterion, we will perform a reasonable optimization of θ.

The remainder of the present paper is organized as follows: In Section 2, we propose the GC_p criterion, which includes criteria proposed by Yanagihara and Satoh (2010) as special cases. In Section 3, we propose three optimization methods for λ. In Section 4, we compare the optimization methods by conducting numerical studies. Finally, technical details are provided in the Appendix.

2. Generalized C_p Criterion

In this section, we propose the GC_p criterion for optimizing the ridge parameter, which includes C_p and MC_p criteria proposed by Yanagihara and Satoh (2010). Moreover, we present several mathematical properties of the optimal θ by minimizing the GC_p criterion.

The PMSE of $\hat{\boldsymbol{Y}}_\theta$ is defined as

$$\mathrm{PMSE}[\hat{\boldsymbol{Y}}_\theta] = E_{\boldsymbol{Y}}[E_{\boldsymbol{U}}[\mathrm{tr}\{(\boldsymbol{U} - \hat{\boldsymbol{Y}}_\theta)'(\boldsymbol{U} - \hat{\boldsymbol{Y}}_\theta)\boldsymbol{\Sigma}^{-1}\}]],$$

where $\boldsymbol{U}$ is a random variable matrix that is independent of $\boldsymbol{Y}$ and has the same distribution as $\boldsymbol{Y}$.

The C_p criterion proposed by Yanagihara and Satoh (2010) is a rough estimator of the PMSE of $\hat{\boldsymbol{Y}}_\theta$, which is defined by

$$C_p(\theta) = \mathrm{tr}(\boldsymbol{W}_\theta \boldsymbol{S}^{-1}) + 2p\mathrm{tr}(\boldsymbol{M}_\theta^{-1}\boldsymbol{M}_0),$$

where $\boldsymbol{W}_\theta$ is a residual sum of squares matrix defined by $\boldsymbol{W}_\theta = (\boldsymbol{Y} - \hat{\boldsymbol{Y}}_\theta)'(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_\theta)$, $\boldsymbol{S}$ is an unbiased estimator of $\boldsymbol{\Sigma}$ defined by $\boldsymbol{S} = \boldsymbol{W}_0/(n-k-1)$. From the definition of the C_p criterion, the first term of C_p measures the closeness of the ridge regression to the data, and the second term evaluates the penalty for the complexity of the ridge regression. However, the C_p criterion has the bias to the PMSE. The MC_p proposed by Yanagihara and Satoh (2010) is an exact unbiased estimator of the PMSE. By neglecting terms that are independent of θ, MC_p is defined as

$$MC_p(\theta) = c_{\mathrm{M}}\mathrm{tr}(\boldsymbol{W}_\theta \boldsymbol{S}^{-1}) + 2p\mathrm{tr}(\boldsymbol{M}_\theta^{-1}\boldsymbol{M}_0),$$

where $c_{\mathrm{M}} = 1 - (p+1)/(n-k-1)$. By comparing the two criteria, we can see that the difference between C_p and MC_p is a coefficient before $\mathrm{tr}(\boldsymbol{W}_\theta \boldsymbol{S}^{-1})$.

Thus, we can generalize the model selection criterion for optimizing the ridge parameter as

$$GC_p(\theta, \lambda) = \lambda\mathrm{tr}(\boldsymbol{W}_\theta \boldsymbol{S}^{-1}) + 2p\mathrm{tr}(\boldsymbol{M}_\theta^{-1}\boldsymbol{M}_0), \tag{2.1}$$

where λ is a non-negative parameter. Note that $GC_p(\theta, 1) = C_p(\theta)$ and $GC_p(\theta,\ c_{\mathrm{M}}) = MC_p(\theta)$. In this criterion, the penalty for the complexity of the model, which is in the second term of (2.1), becomes large when λ becomes small. This means that λ controls the penalty for the complexity of the model in the criterion (2.1). Hence, we can regard λ as a penalty parameter. In the present paper, we consider the optimization of λ to obtain the optimal θ, which further reduces the PMSE.

When λ is fixed, the optimized ridge parameter $\hat{\theta}(\lambda)$ is obtained by

$$\hat{\theta}(\lambda) = \arg\min_{\theta \in [0,\infty]} GC_p(\theta, \lambda). \tag{2.2}$$

Since $\hat{\theta}(\lambda)$ is a minimizer of $GC_p(\theta, \lambda)$, the following equation holds:

$$\left.\frac{\partial GC_p(\theta, \lambda)}{\partial \theta}\right|_{\theta = \hat{\theta}(\lambda)} = 0. \tag{2.3}$$

Note that $\hat{\theta}(\lambda)$ changes with λ.

Here, we obtain the following mathematical properties of $\hat{\theta}(\lambda)$ (The proof is provided in Appendix A.1.):

Theorem 2.1. *Let*

$$(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k)' = \boldsymbol{Q}'\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{S}^{-1/2}, \tag{2.4}$$

$$r_{\lambda,j} = \frac{\lambda\|\boldsymbol{z}_j\|^2 - pd_j}{pd_j^2}, \ (j = 1, \ldots, k), \tag{2.5}$$

where $\boldsymbol{z}_i$ *is a* p*-dimensional vector,* $\boldsymbol{Q}$ *is a* $k \times k$ *orthogonal matrix which diagonalizes* $\boldsymbol{X}'\boldsymbol{X}$*, i.e.,* $\boldsymbol{Q}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{Q} = \boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_k)$ *and* d_i $(i = 1, \ldots, k)$ *are eigenvalues of* $\boldsymbol{X}'\boldsymbol{X}$*, and* $r^+_{\lambda,1} \le \cdots \le r^+_{\lambda,m}$ $(m \le k)$ *are positive values of* $r_{\lambda,1}, \ldots, r_{\lambda,k}$*. Then,* $\hat{\theta}(\lambda)$ *has the following properties:*

1. $\hat{\theta}(\lambda)$ *is a monotonic decreasing function with respect to* λ*.*
2. $\hat{\theta}(\lambda)$ *is not* 0 *when* $\lambda \in [0, \infty)$*.*
3. $\hat{\theta}(\lambda) > (r^+_{\lambda,1})^{-1}$ *when* $r^+_{\lambda,1}$ *exists.* $\hat{\theta}(\lambda) = \infty$ *when* $r^+_{\lambda,1}$ *does not exist, i.e.,* $\max_{j=1,\ldots,m} r_{\lambda,j} \le 0$*.*
4. $\hat{\theta}(\infty) = 0$*,* $\hat{\theta}(0) = \infty$*.*
5. $\hat{\theta}(\lambda) = \infty$ *for any* $\lambda < \min_{j=1,\ldots,k} pd_j^2/\|\boldsymbol{z}_j\|^2$*.*

We suppose that $d_i = O(n)$. However, we must use an iterative computational algorithm to optimize θ because we cannot obtain $\hat{\theta}(\lambda)$ in closed form. In order to reduce the number of computational tasks, we consider approximating $\hat{\theta}(\lambda)$ using an asymptotic expansion. Equation (2.3) implies asymptotic expansion of the GC_p criterion. From this expansion, we obtain the asymptotic expansion of $\hat{\theta}(\lambda)$ as the follows:

Theorem 2.2. *Here,* $\hat{\theta}(\lambda)$ *can be expanded as*

$$\hat{\theta}(\lambda) = \tilde{\theta}_{(L)}(\lambda) + O_p(n^{-L}),$$

where

$$\begin{aligned}\tilde{\theta}_{(L)}(\lambda) \ &= \frac{pb_1}{\lambda a_1} \\ &+ \frac{1}{2\lambda a_1}\sum_{\ell=0}^{L-1}\frac{1}{n^\ell}(-1)^{\ell+1}(\ell+1)\tilde{\theta}^{\ell}_{(L-1)}(\lambda)\{\lambda(\ell+2)a_{\ell+1}\tilde{\theta}_{(L-1)}(\lambda) - 2pb_{\ell+1}\}.\end{aligned} \tag{2.6}$$

Here, $\tilde{\theta}_{(0)}(\lambda) = 0$*,* $\boldsymbol{V} = \boldsymbol{X}'\boldsymbol{Y}\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{X}$*,* $a_j = n^j\mathrm{tr}(\boldsymbol{V}\boldsymbol{M}_0^{-(j+2)})$*, and* $b_j = n^j\mathrm{tr}(\boldsymbol{M}_0^{-j})$*, and* $\tilde{\theta}^{\ell}_{(L)}(\lambda)$ *refers to* $\{\tilde{\theta}_{(L)}(\lambda)\}^\ell$*.*

The proofs of this theorem are given in Appendix A.2. Note that $\tilde{\theta}_{(L)}(\lambda)$ can be used as an approximated value of $\hat{\theta}(\lambda)$. There is a one-to-one correspondence between $\tilde{\theta}_{(1)}(\lambda) = pb_1/(\lambda a_1)$ and λ, and $\tilde{\theta}_{(1)}(\lambda)$ satisfies the properties 1, 2, and 4 in Theorem 2.1.

3. Optimization of Penalty in the GC_p Criterion

3.1. Double optimization of θ and λ

In the previous section, we considered the model selection criterion for selecting θ, which can be regarded as an estimator of PMSE$[\hat{\boldsymbol{Y}}_\theta]$. By minimizing the estimator of PMSE of $\hat{\boldsymbol{Y}}_\theta$, we expect to reduce the PMSE of $\hat{\boldsymbol{Y}}_\theta$. However, since the optimal ridge parameter will be changed by the data, it is important to reduce not the PMSE of $\hat{\boldsymbol{Y}}_\theta$ but rather the PMSE of $\hat{\boldsymbol{Y}}_{\hat{\theta}}$, i.e., the predictor of $\boldsymbol{Y}$ after optimizing θ. In this section, we consider optimizing λ using PMSE$[\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}]$, where $\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)} = \boldsymbol{1}_n\hat{\boldsymbol{\mu}}' + \boldsymbol{X}\hat{\boldsymbol{\Xi}}_{\hat{\theta}(\lambda)}$.

Without a loss of generality, we can assume that the covariance matrix of $\boldsymbol{y}_i$ is $\boldsymbol{I}_p$ in the PMSE$[\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}]$. Therefore, from Efron (2004), we obtain PMSE$[\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}]$, which is a function of λ, as follows:

$$\mathrm{PMSE}[\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}] = E_{\boldsymbol{Y}}[\mathrm{tr}(\boldsymbol{W}_{\hat{\theta}(\lambda)}\boldsymbol{\Sigma}^{-1})] + 2E_{\boldsymbol{Y}}\left[\sum_{i=1}^n\sum_{j=1}^p \frac{\partial(\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)})_{ij}}{\partial(\boldsymbol{Y})_{ij}}\right],$$

where $(\boldsymbol{A})_{ij}$ are the (i,j)th elements of $\boldsymbol{A}$. Since $\hat{\theta}(\lambda)$ depends on $(\boldsymbol{Y})_{ij}$, we can see that

$$\frac{\partial(\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)})_{ij}}{\partial(\boldsymbol{Y})_{ij}} = \left.\frac{\partial(\hat{\boldsymbol{Y}}_\theta)_{ij}}{\partial(\boldsymbol{Y})_{ij}}\right|_{\theta=\hat{\theta}(\lambda)} + \left.\frac{\partial(\hat{\boldsymbol{Y}}_\theta)_{ij}}{\partial\theta}\right|_{\theta=\hat{\theta}(\lambda)} \frac{\partial\hat{\theta}(\lambda)}{\partial(\boldsymbol{Y})_{ij}}. \tag{3.1}$$

The first term of the above equation is calculated as

$$\begin{aligned}\left.\frac{\partial(\hat{\boldsymbol{Y}}_\theta)_{ij}}{\partial(\boldsymbol{Y})_{ij}}\right|_{\theta=\hat{\theta}(\lambda)} &= \frac{\partial(\boldsymbol{1}_n\hat{\boldsymbol{\mu}}')_{ij}}{\partial(\boldsymbol{Y})_{ij}} + \left.\frac{\partial(\boldsymbol{X}\boldsymbol{M}_\theta^{-1}\boldsymbol{X}'\boldsymbol{Y})_{ij}}{\partial(\boldsymbol{Y})_{ij}}\right|_{\theta=\hat{\theta}(\lambda)} \\ &= \frac{\partial(\boldsymbol{1}_n\hat{\boldsymbol{\mu}}')_{ij}}{\partial(\boldsymbol{Y})_{ij}} + (\boldsymbol{X}\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-1}\boldsymbol{X}')_{ii}.\end{aligned}$$

Note that $\sum_{i=1}^n\sum_{j=1}^p \partial(\boldsymbol{1}_n\hat{\boldsymbol{\mu}}')_{ij}/\partial(\boldsymbol{Y})_{ij} = p$ and $\sum_{i=1}^n\sum_{j=1}^p(\boldsymbol{X}\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-1}\boldsymbol{X}')_{ii} = p\mathrm{tr}(\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-1}\boldsymbol{M}_0)$.

Next, we consider obtaining the second term of (3.1). Note that

$$\frac{\partial(\hat{\boldsymbol{Y}}_\theta)_{ij}}{\partial\theta} = \frac{\partial(\boldsymbol{X}\boldsymbol{M}_\theta^{-1}\boldsymbol{X}'\boldsymbol{Y})_{ij}}{\partial\theta} = -(\boldsymbol{X}\boldsymbol{M}_\theta^{-2}\boldsymbol{X}'\boldsymbol{Y})_{ij}.$$

Hence, we derive

$$\begin{aligned}\mathrm{PMSE}[\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}] &= E_{\boldsymbol{Y}}[\mathrm{tr}(\boldsymbol{W}_{\hat{\theta}(\lambda)}\boldsymbol{\Sigma}^{-1})] + 2p\{E_{\boldsymbol{Y}}[\mathrm{tr}(\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-1}\boldsymbol{M}_0)] + 1\} \\ &\quad - 2E_{\boldsymbol{Y}}\left[\sum_{i=1}^n\sum_{j=1}^p(\boldsymbol{X}\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-2}\boldsymbol{X}'\boldsymbol{Y})_{ij}\frac{\partial\hat{\theta}(\lambda)}{\partial(\boldsymbol{Y})_{ij}}\right].\end{aligned} \tag{3.2}$$

Based on this result, we need only obtain $\partial\hat{\theta}(\lambda)/\partial(\boldsymbol{Y})_{ij}$ in order to calculate PMSE$[\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}]$. This derivative leads to the following theorem (The proofs are given in Appendix A.3.):

Theorem 3.1. *The PMSE of $\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}$ is expressed as*

$$\mathrm{PMSE}[\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}] = E_{\boldsymbol{Y}}[\mathrm{tr}(\boldsymbol{W}_{\hat{\theta}(\lambda)}\boldsymbol{\Sigma}^{-1}) + 2p\{\mathrm{tr}(\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-1}\boldsymbol{M}_0) + 1\} + 4B(\hat{\theta}(\lambda))], \tag{3.3}$$

where

$$B(\theta) = \frac{\lambda\theta\mathrm{tr}(\boldsymbol{V}\boldsymbol{M}_\theta^{-5}\boldsymbol{M}_0)}{\lambda\mathrm{tr}(\boldsymbol{V}\boldsymbol{M}_\theta^{-3}) - 3\lambda\theta\mathrm{tr}(\boldsymbol{V}\boldsymbol{M}_\theta^{-4}) + 2p\mathrm{tr}(\boldsymbol{M}_\theta^{-3}\boldsymbol{M}_0)}.$$

By neglecting the terms that are independent of λ, we define the penalty selection criteria for optimizing λ as follows:

Definition 3.1. *The penalty selection criteria to optimize λ are defined as*

$$C_p^{\#}(\lambda) = \mathrm{tr}(\boldsymbol{W}_{\hat{\theta}(\lambda)}\boldsymbol{S}^{-1}) + 2p\mathrm{tr}(\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-1}\boldsymbol{M}_0) + 4B(\hat{\theta}(\lambda)),$$
$$MC_p^{\#}(\lambda) = c_M\mathrm{tr}(\boldsymbol{W}_{\hat{\theta}(\lambda)}\boldsymbol{S}^{-1}) + 2p\mathrm{tr}(\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-1}\boldsymbol{M}_0) + 4B(\hat{\theta}(\lambda)),$$

where $\hat{\theta}(\lambda)$ is given by (2.2) and $c_M = 1 - (p+1)/(n-k-1)$.

Here, note that $C_p^{\#}(\lambda)$ is obtained by substituting $\boldsymbol{S}^{-1}$ for $\boldsymbol{\Sigma}^{-1}$ when we neglect the terms that are independent of λ in (3.3). However, there exists a bias because $\boldsymbol{S}^{-1}$ is not an unbiased estimator of $\boldsymbol{\Sigma}^{-1}$ (see, e.g., Siotani, Hayakawa, and Fujikoshi (1985)). Based on the results reported by Yanagihara and Satoh (2010), we will correct the bias of $E_{\boldsymbol{Y}}[\mathrm{tr}(\boldsymbol{W}_{\hat{\theta}(\lambda)}\boldsymbol{S}^{-1})]$ to $E_{\boldsymbol{Y}}[\mathrm{tr}(\boldsymbol{W}_{\hat{\theta}(\lambda)}\boldsymbol{\Sigma}^{-1})]$ under fixed $\hat{\theta}(\lambda)$. Finally, we define $MC_p^{\#}(\lambda)$ by neglecting terms that are independent of θ and λ. Using these criteria, λ and θ are optimized as follows:

$$\hat{\lambda}_{\mathrm{C}}^{\#} = \arg\min_{\lambda\in[0,\infty]} C_p^{\#}(\lambda) \text{ and } \hat{\theta}(\hat{\lambda}_{\mathrm{C}}^{\#}) = \arg\min_{\theta\in[0,\infty]} GC_p(\theta, \hat{\lambda}_{\mathrm{C}}^{\#}),$$
$$\hat{\lambda}_{\mathrm{M}}^{\#} = \arg\min_{\lambda\in[0,\infty]} MC_p^{\#}(\lambda) \text{ and } \hat{\theta}(\hat{\lambda}_{\mathrm{M}}^{\#}) = \arg\min_{\theta\in[0,\infty]} GC_p(\theta, \hat{\lambda}_{\mathrm{M}}^{\#}).$$

These optimization methods are similar to those reported by Ye (1998) and Shen and Ye (2002).

3.2. Optimization of λ with approximated $\hat{\theta}(\lambda)$

In the previous subsection, we proposed penalty selection criteria for selecting λ. These criteria are made from the optimal θ obtained by minimizing the GC_p criterion. This indicates that we need to repeat the optimization of θ until obtaining the optimal λ. Hence, a number of computational tasks are required for such an optimization. In this subsection, we try to reduce the number of computational tasks by using the approximated $\hat{\theta}(\lambda)$, which is given by (2.6). Thus, we propose the penalty selection criterion when the approximated $\hat{\theta}(\lambda)$ is used. As such, we calculate $\partial\tilde{\theta}_{(L)}(\lambda)/\partial(\boldsymbol{Y})_{ij}$. The following lemma is useful for obtaining such a derivative (The proof is provided in Appendix A.4.):

Lemma 3.1. *For any ℓ, the first derivative of a_ℓ with respect to $(\boldsymbol{Y})_{ij}$ is calculated as*

$$\frac{\partial a_\ell}{\partial(\boldsymbol{Y})_{ij}} = 2n^\ell \left(\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{X}\boldsymbol{M}_0^{-\ell-2}\boldsymbol{X}' \left(\boldsymbol{I}_n - \frac{1}{n-k-1}\boldsymbol{Y}\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{H} \right) \right)_{ji},$$

where $\boldsymbol{H} = \boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n'/n - \boldsymbol{X}\boldsymbol{M}_0^{-1}\boldsymbol{X}'$.

By using this lemma and (2.6), we obtain the following theorem:

Theorem 3.2. *The PMSE of $\hat{\boldsymbol{Y}}_{\tilde{\theta}_{(L)}(\lambda)}$ is expressed as*

$$\mathrm{PMSE}[\hat{\boldsymbol{Y}}_{\tilde{\theta}_{(L)}(\lambda)}] = E_{\boldsymbol{Y}}[\mathrm{tr}(\boldsymbol{W}_{\tilde{\theta}_{(L)}(\lambda)}\boldsymbol{\Sigma}^{-1}) + 2p\{\mathrm{tr}(\boldsymbol{M}_{\tilde{\theta}_{(L)}(\lambda)}^{-1}\boldsymbol{M}_0) + 1\}] + 2E_{\boldsymbol{Y}}[B'(\tilde{\theta}_{(L)}(\lambda))],$$

where

$$\begin{aligned}
&B'(\tilde{\theta}_{(L)}(\lambda)) \\
&= \frac{2n}{a_1}\tilde{\theta}_{(1)}(\lambda)\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(L)}(\lambda)}^{-2}\boldsymbol{V}) \\
&+ \frac{1}{\lambda a_1^2}\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(L)}(\lambda)}^{-2}\boldsymbol{V})\sum_{\ell=0}^{L-1}\frac{1}{n^{\ell-1}}(-1)^{\ell+1}(\ell+1)\tilde{\theta}_{(L-1)}^{\ell}(\lambda)\{\lambda(\ell+2)a_{\ell+1}\tilde{\theta}_{(L-1)}(\lambda) - 2pb_{\ell+1}\} \\
&- \frac{1}{2\lambda a_1}\sum_{\ell=0}^{L-1}\frac{1}{n^\ell}(-1)^{\ell+1}(\ell+1)\tilde{\theta}_{(L-1)}^{\ell-1}\{\lambda(\ell+1)(\ell+2)a_{\ell+1}\tilde{\theta}_{(L-1)}(\lambda) - 2p\ell b_{\ell+1}\} \\
&\qquad \times \sum_{i=1}^{n}\sum_{j=1}^{p}(\boldsymbol{X}\boldsymbol{M}_{\tilde{\theta}_{(L)}(\lambda)}^{-2}\boldsymbol{X}'\boldsymbol{Y})_{ij}\frac{\partial\tilde{\theta}_{(L-1)}}{\partial(\boldsymbol{Y})_{ij}} \\
&- \frac{n}{a_1}\sum_{\ell=0}^{L-1}(-1)^{\ell+1}(\ell+1)(\ell+2)\tilde{\theta}_{(L-1)}^{\ell+1}\mathrm{tr}(\boldsymbol{M}_0^{-(\ell+2)}\boldsymbol{M}_{\tilde{\theta}_{(L)}(\lambda)}^{-2}\boldsymbol{V}).
\end{aligned}$$

The proof of this theorem is presented in Appendix A.5. When $\tilde{\theta}_{(1)}(\lambda)$ is used, we obtain

$$\begin{aligned}
\mathrm{PMSE}[\hat{\boldsymbol{Y}}_{\tilde{\theta}_{(1)}(\lambda)}] &= E_{\boldsymbol{Y}}[\mathrm{tr}(\boldsymbol{W}_{\tilde{\theta}_{(1)}(\lambda)}\boldsymbol{\Sigma}^{-1}) + 2p\{\mathrm{tr}(\boldsymbol{M}_{\tilde{\theta}_{(1)}(\lambda)}^{-1}\boldsymbol{M}_0) + 1\}] \\
&\quad + E_{\boldsymbol{Y}}\left[\frac{4n\tilde{\theta}_{(1)}(\lambda)}{a_1}\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(1)}(\lambda)}^{-2}\boldsymbol{V})\right].
\end{aligned}$$

Thus, by neglecting terms that are independent of λ, the penalty selection criteria with $\tilde{\theta}_{(1)}(\lambda)$ are defined as follows:

Definition 3.2. *Penalty selection criteria to optimize λ when $\tilde{\theta}_{(1)}(\lambda)$ is used are defined as follows:*

$$\begin{aligned}
C_p^{(1)}(\lambda) &= \mathrm{tr}(\boldsymbol{W}_{\tilde{\theta}_{(1)}(\lambda)}\boldsymbol{S}^{-1}) + 2p\mathrm{tr}(\boldsymbol{M}_{\tilde{\theta}_{(1)}(\lambda)}^{-1}\boldsymbol{M}_0) + \frac{4n}{a_1}\tilde{\theta}_{(1)}(\lambda)\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(1)}(\lambda)}^{-2}\boldsymbol{V}), \\
MC_p^{(1)}(\lambda) &= c_M\mathrm{tr}(\boldsymbol{W}_{\tilde{\theta}_{(1)}(\lambda)}\boldsymbol{S}^{-1}) + 2p\mathrm{tr}(\boldsymbol{M}_{\tilde{\theta}_{(1)}(\lambda)}^{-1}\boldsymbol{M}_0) + \frac{4n}{a_1}\tilde{\theta}_{(1)}(\lambda)\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(1)}(\lambda)}^{-2}\boldsymbol{V}),
\end{aligned}$$

where $\tilde{\theta}_{(1)}(\lambda) = pb_1/(\lambda a_1)$ and $c_M = 1 - (p+1)/(n-k-1)$.

Similar to $MC_p^{\#}(\lambda)$, $MC_p^{(1)}(\lambda)$ can be regarded as simple bias-corrected $C_p^{(1)}(\lambda)$. At least, when $\tilde{\theta}_{(1)}(\lambda) = 0$, $MC_p^{(1)}(\lambda)$ completely corrects the bias of $C_p^{(1)}(\lambda)$. If we use a $\tilde{\theta}_{(L)}(\lambda)$ other than $\tilde{\theta}_{(1)}(\lambda)$, the penalty selection criteria becomes more complicated as the number of L increases. As an example, we describe the penalty selection criteria for $\tilde{\theta}_{(2)}(\lambda)$ in Appendix A.6. From the viewpoint of an application, $C_p^{(1)}(\lambda)$ and $MC_p^{(1)}(\lambda)$ are useful because these are the simplest criteria among all L. When we use $C_p^{(1)}(\lambda)$ and $MC_p^{(1)}(\lambda)$, the optimal θ and λ are given as follows:

$$\begin{aligned}
\hat{\lambda}_{\mathrm{C}}^{(1)} &= \arg\min_{\lambda\in[0,\infty]} C_p^{(1)}(\lambda) \text{ and } \hat{\theta}(\hat{\lambda}_{\mathrm{C}}^{(1)}) = \tilde{\theta}_{(1)}(\hat{\lambda}_{\mathrm{C}}^{(1)}), \\
\hat{\lambda}_{\mathrm{M}}^{(1)} &= \arg\min_{\lambda\in[0,\infty]} MC_p^{(1)}(\lambda) \text{ and } \hat{\theta}(\hat{\lambda}_{\mathrm{M}}^{(1)}) = \tilde{\theta}_{(1)}(\hat{\lambda}_{\mathrm{M}}^{(1)}).
\end{aligned}$$

3.3. Asymptotic optimization for λ

In previous subsections, we proposed the penalty selection criteria. When such criteria are used to optimize λ, we must perform an iterative procedure. In this subsection, we consider the non-iterative optimization of λ. This requires the calculation of an asymptotic optimal λ, which minimizes an asymptotic expansion of $\mathrm{PMSE}[\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}]$ among $\lambda \in [0,\infty]$. The following theorem gives such an asymptotic optimal value of λ (The proof is provided in Appendix A.7.):

Theorem 3.3. *An asymptotic optimal λ^* minimizes $\mathrm{PMSE}[\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}]$ asymptotically is given by*

$$\frac{1}{\lambda^*} = \frac{E_{\boldsymbol{Y}}[a_1^{-1}]}{E_{\boldsymbol{Y}}[a_1^*/a_1^2]} - \frac{2E_{\boldsymbol{Y}}[a_2/a_1^2]}{pb_1 E_{\boldsymbol{Y}}[a_1^*/a_1^2]},$$

where $\boldsymbol{V}^* = \boldsymbol{X}'\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}'\boldsymbol{X}$ and $a_j^* = n^j \mathrm{tr}(\boldsymbol{V}^*\boldsymbol{M}_0^{-(j+2)})$.

By replacing a_1^* with a_1, we estimate λ^* as follows:

$$\hat{\lambda}_0 = \left\{1 - \frac{2\mathrm{tr}(\boldsymbol{M}_0^{-4}\boldsymbol{V})}{p\mathrm{tr}(\boldsymbol{M}_0^{-3}\boldsymbol{V})\mathrm{tr}(\boldsymbol{M}_0^{-1})}\right\}^{-1}. \tag{3.4}$$

Note that $E_{\boldsymbol{Y}}[a_1] = c_{\mathrm{M}}^{-1}E_{\boldsymbol{Y}}[a_1^*]$ holds. Hence, we can estimate $E_{\boldsymbol{Y}}[a_1^*]$ as $c_{\mathrm{M}}a_1$. This implies new estimator of λ^* given by $\hat{\lambda}_{\mathrm{M}} = c_{\mathrm{M}}\hat{\lambda}_0$. When we use $\hat{\lambda}_0$ and $\hat{\lambda}_{\mathrm{M}}$, optimal θ is given by

$$\begin{aligned}
\hat{\theta}(\hat{\lambda}_0) &= \arg\min_{\theta\in[0,\infty]} GC_p(\theta, \hat{\lambda}_0) \text{ and } \hat{\lambda}_0 \text{ is in (3.4)}, \\
\hat{\theta}(\hat{\lambda}_{\mathrm{M}}) &= \arg\min_{\theta\in[0,\infty]} GC_p(\theta, \hat{\lambda}_{\mathrm{M}}) \text{ and } \hat{\lambda}_{\mathrm{M}} = c_{\mathrm{M}}\hat{\lambda}_0.
\end{aligned}$$

4. Numerical Study

In this section, we conduct numerical studies to compare the PMSEs of predictors of $\boldsymbol{Y}$ consisting of the ridge regression estimators with optimized ridge and penalty parameters. Let $\boldsymbol{R}_q$ and $\boldsymbol{\Delta}_q(\rho)$ be $q \times q$ matrices defined by $\boldsymbol{R}_q = \mathrm{diag}(1, \ldots, q)$ and $(\boldsymbol{\Delta}_q(\rho))_{ij} = \rho^{|i-j|}$. The explanatory matrix $\boldsymbol{X}$ was generated from $\boldsymbol{X} = \boldsymbol{W}\boldsymbol{\Psi}^{1/2}$, where $\boldsymbol{\Psi} = \boldsymbol{R}_k^{1/2}\boldsymbol{\Delta}_k(\rho_x)\boldsymbol{R}_k^{1/2}$, and $\boldsymbol{W}$ is an $n \times k$ matrix, the elements of which were generated independently from the uniform distribution on $(-1, 1)$. The $k \times p$ unknown regression coefficient matrix $\boldsymbol{\Xi}$ was defined by $\boldsymbol{\Xi} = \delta \boldsymbol{F} \boldsymbol{\Xi}_i$, where δ is a constant, and $\boldsymbol{F}$ is defined as $\boldsymbol{F} = \mathrm{diag}(\boldsymbol{1}_\kappa, \boldsymbol{0}_{k-\kappa})$, which is a $k \times k$ matrix, and $\boldsymbol{\Xi}_i$ is defined as the first five rows of $\boldsymbol{\Xi}_0$ when $k = 5$, $\boldsymbol{\Xi}_0$ when $k = 10$, and $\boldsymbol{\Xi}_1$ when $k = 15$.

$$\boldsymbol{\Xi}_0 = \begin{pmatrix} 0.8501 & 0.6571 & 0.2159 \\ -0.2753 & -0.2432 & -0.1187 \\ -0.3193 & -0.2926 & -0.1671 \\ 0.2754 & 0.2608 & 0.1766 \\ 0.2693 & 0.2164 & 0.2066 \\ -0.0676 & -0.0663 & -0.0561 \\ 0.2239 & 0.2197 & 0.1880 \\ -0.0352 & -0.0346 & -0.0305 \\ 0.3240 & 0.3199 & 0.2868 \\ -0.3747 & -0.3727 & -0.3554 \end{pmatrix}, \quad \boldsymbol{\Xi}_1 = \begin{pmatrix} 1.3794 & 0.0645 & 0.0330 \\ -0.0766 & -0.0241 & -0.0143 \\ -0.2618 & -0.1396 & -0.0951 \\ -0.4619 & -0.2589 & -0.1798 \\ 0.2381 & 0.1488 & 0.1082 \\ 0.2140 & 0.1463 & 0.1112 \\ 0.3002 & 0.2364 & 0.1950 \\ 0.1155 & 0.0953 & 0.0812 \\ -0.2774 & -0.2395 & -0.2091 \\ 0.3392 & 0.3072 & 0.2807 \\ 0.0016 & 0.0107 & 0.0100 \\ 0.0438 & 0.0408 & 0.0381 \\ -0.3187 & -0.3039 & -0.2904 \\ 0.0529 & 0.0510 & 0.0493 \\ 0.2505 & 0.2451 & 0.2399 \end{pmatrix}.$$

Here, δ controls the scale of the regression coefficient matrix, and $\boldsymbol{F}$ controls the number of non-zero regression coefficients via κ (the dimension of the true model). The values of the elements of $\boldsymbol{\Xi}_0$ and $\boldsymbol{\Xi}_1$, which is an essential regression coefficient matrix, are the same as in Lawless (1981). Simulated data values $\boldsymbol{Y}$ were generated by $N_{n\times 3}(\boldsymbol{X}\boldsymbol{\Xi}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)$ repeatedly under several selections of n, k, κ, δ, ρ_y, and ρ_x, where $\boldsymbol{\Sigma} = \boldsymbol{R}_3^{1/2}\boldsymbol{\Delta}_3(\rho_y)\boldsymbol{R}_3^{1/2}$, and the number of repetition was 1000. At each repetition, we evaluated $r(\boldsymbol{X}\boldsymbol{\Xi}, \hat{\boldsymbol{Y}}_{\hat{\theta}}) = \mathrm{tr}\{(\boldsymbol{X}\boldsymbol{\Xi} - \hat{\boldsymbol{Y}}_\theta)'(\boldsymbol{X}\boldsymbol{\Xi} - \hat{\boldsymbol{Y}}_\theta)\boldsymbol{\Sigma}^{-1}\}$, where $\hat{\boldsymbol{Y}}_{\hat{\theta}} = \boldsymbol{1}_n\hat{\boldsymbol{\mu}}' + \boldsymbol{X}\hat{\boldsymbol{\Xi}}_{\hat{\theta}}$, which is the predicted value of $\boldsymbol{Y}$ obtained from each method. The average of $np + r(\boldsymbol{X}\boldsymbol{\Xi}, \hat{\boldsymbol{Y}}_{\hat{\theta}})$ across 1000 repetitions was regarded as the PMSE of $\hat{\boldsymbol{Y}}_{\hat{\theta}}$. In the simulation, a standardized $\boldsymbol{X}$ was used to estimate the regression coefficients.

Recall that $GC_p(\theta, \lambda)$ is defined in (2.1). Here, λ and θ are optimized by the following methods:

Method 1: $\hat{\theta} = \arg\min_{\theta\in[0,\infty]} GC_p(\theta, \hat{\lambda})$ and $\hat{\lambda} = \hat{\lambda}_0$, where $\hat{\lambda}_0$ is defined in (3.4).

Method 2: $\hat{\theta} = \arg\min_{\theta\in[0,\infty]} GC_p(\theta, \hat{\lambda})$ and $\hat{\lambda} = \hat{\lambda}_{\mathrm{M}} = c_{\mathrm{M}}\hat{\lambda}_0$, where $c_{\mathrm{M}} = 1 - (p+1)/(n-k-1)$.

Method 3: $\hat{\theta} = \arg\min_{\theta\in[0,\infty]} GC_p(\theta, \hat{\lambda})$ and $\hat{\lambda} = \arg\min_{\lambda\in[0,\infty]} C_p^{\#}(\lambda)$, where $C_p^{\#}(\lambda)$ is given by Definition 3.1.

Method 4: $\hat{\theta}=\arg\min_{\theta\in[0,\infty]} GC_p(\theta, \hat{\lambda})$ and $\hat{\lambda}=\arg\min_{\lambda\in[0,\infty]} MC_p^{\#}(\lambda)$, where $MC_p^{\#}(\lambda)$ is given by Definition 3.1.

Method 5: $\hat{\theta} = \tilde{\theta}_{(1)}(\hat{\lambda})$ and $\hat{\lambda} = \arg\min_{\lambda\in[0,\infty]} C_p^{(1)}(\lambda)$, where $\tilde{\theta}_{(1)}(\lambda)$ and $C_p^{(1)}(\lambda)$ are defined in (2.6) and by Definition 3.2.

Method 6: $\hat{\theta} = \tilde{\theta}_{(1)}(\hat{\lambda})$ and $\hat{\lambda} = \arg\min_{\lambda\in[0,\infty]} MC_p^{(1)}(\lambda)$, where $MC_p^{(1)}(\lambda)$ is given by Definition 3.2.

For the purpose of comparison with the proposed methods, we prepare conventional optimization methods, which are obtained using the following methods:

Method 7: $\hat{\theta}_{\mathrm{C}} = \arg\min_{\theta\in[0,\infty]} C_p(\theta) = \arg\min_{\theta\in[0,\infty]} GC_p(\theta, 1)$.

Method 8: $\hat{\theta}_{\mathrm{M}} = \arg\min_{\theta\in[0,\infty]} MC_p(\theta) = \arg\min_{\theta\in[0,\infty]} GC_p(\theta, c_{\mathrm{M}})$, where $c_{\mathrm{M}} = 1 - (p+1)/(n-k-1)$.

In Methods 3 through 8, the fminsearch function in Matlab is used to find the minimizer of the penalty selection criterion or model selection criterion. In the fminsearch function, the Nelder-Mead simplex method (see, e.g., Lagarias *et al.*, 1998) is used to search the value that minimizes the function. When Methods 1 through 4 are used, an optimal θ is searched using the fminsearch function. We can see that computational speeds of Methods 1 and 2 are the same as those of Methods 5 and 6. Furthermore, the computational speeds of Methods 5 and 6 are almost the same as those of Methods 7 and 8 because these four methods optimize one parameter. It is easy to predict that the computational speeds of Methods 3 and 4 are slower than the other methods because Methods 3 and 4 optimize two parameters simultaneously.

In this paper, we proposed Methods 1 through 6 as referred to above, and these methods can be regarded as the estimation methods for the optimal λ. To obtain the optimal λ, called λ^{**}, which minimizes the PMSE, we divided $[0, 2]$ into 100 parts and used each point. Then we compute $r(\boldsymbol{X\Xi}, \hat{\boldsymbol{Y}}_{\hat{\theta}})$ for each point in each repetition. After 1000 repetitions, we compute the averages of these values for each point which are regarded as the main term of PMSE of $\hat{\boldsymbol{Y}}_{\hat{\theta}}$. By comparing the average values, the λ^{**} is obtained. For comparing $\hat{\lambda}$ which is estimated λ by using each method in above Methods 1 through 6, we show the Figure 1 in some situations which shows the box plots of each $\hat{\lambda}$ in 1000 repetitions. The horizontal line means the λ^{**}. Tables 1 through 4 show the averages of $(\hat{\lambda} - \lambda^{**})^2$ across 1000 repetitions, which is referred as the mean squared error (MSE) of $\hat{\lambda}$, for each method.

In the Theorem 2.2, we derived the expansion for $\hat{\theta}(\lambda)$ and we suggested to use the first term of the expansion which is referred as $\tilde{\theta}_{(1)}(\lambda)$. To compare the $\hat{\theta}(\lambda)$ and $\tilde{\theta}_{(1)}(\lambda)$, we show the scatter plots in some situations when we fix λ as 1 or 2 in Figure 2. In each scatter plot, the line means the 45-degree line which means the line of $\hat{\theta}(\lambda) = \tilde{\theta}_{(1)}(\lambda)$. When the scatter plot close up this line, $\tilde{\theta}_{(1)}(\lambda)$ closes to $\hat{\theta}(\lambda)$.

Tables 7 through 12 show the simulation results obtained for PMSE$[\hat{\boldsymbol{Y}}_{\hat{\theta}}]/\{p(n+k+1)\} \times 100$ for the cases in which $(k, n) = (5, 30)$, $(5, 50)$, $(10, 30)$, $(10, 50)$, $(15, 30)$, and $(15, 50)$, respectively, where $p(n + k + 1)$ is the PMSE of the predictor of $\boldsymbol{Y}$ derived using the LS estimators. We note $p = 3$ in numerical studies. In the tables, bold font indicates the minimized PMSE, and italic font indicates the second-smallest PMSE.

Please insert figures 1 and 2 around here

Please insert tables 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 around here

From the figure 1 shows the box plots for $\hat{\lambda}$ for each method, we can see that the dispersion of $\hat{\lambda}$ and the differences between $\hat{\lambda}$ and λ^{**}. Methods 2, 4 and 6 are always smaller value than Methods 1, 3 and 5. The dispersion of Methods 2, 4 and 6 are smaller than Methods 1, 3 and 5. This facts mean that the correction of each method make the optimized value and dispersion smaller. We note that the dispersions of Methods 1 and 2 are smaller than other methods. When ρ_y and ρ_x are small, our optimization method is nearly equal to λ^{**}. From the tables 1 through 6, we can see that the numerical evaluation for each method. When k and δ are zeros, Method 6 is the best and Method 5 is the second best. Methods 2 and 4 are the best and the second best when κ and δ is small. When κ is equal to $k = 10$ and δ is large, Method 6 is the best method. On the other hand, when $k = \kappa = 15$ and δ is large, Method 6 or 2 is the best in ρ_y is small or large. Consequently, Method 6 and 5 was, on average, the best and the second best method except $k = 15$. When $k = 15$, Method 5 was the best. Hence we recommend using Method 6 to optimize the penalty parameter λ.

From the figure 2 shows the scatter plot for $\tilde{\theta}_{(1)}(\lambda)$ and $\hat{\theta}(\lambda)$, we can see the dispersion in each situation. We note that λ become large, the difference between $\hat{\theta}(\lambda)$ and $\tilde{\theta}_{(1)}(\lambda)$ becomes small. Also when ρ_y or n become large, the difference between $\hat{\theta}(\lambda)$ and $\tilde{\theta}_{(1)}(\lambda)$ becomes small. On the other hand, the difference between $\hat{\theta}(\lambda)$ and $\tilde{\theta}_{(1)}(\lambda)$ becomes large ρ_x is large. In almost case, $\tilde{\theta}_{(1)}(\lambda)$ is smaller than $\hat{\theta}(\lambda)$. This fact is corresponding the result in the Theorem 2.2 since $\hat{\theta}(\lambda) = \tilde{\theta}_{(1)}(\lambda) + O(n^{-1})$. When the ρ_y or δ become large, the dispersion of $\hat{\theta}(\lambda)$ becomes small. The each value of $\hat{\theta}(\lambda)$ and $\tilde{\theta}_{(1)}(\lambda)$ become small when ρ_y, ρ_x, δ or λ becomes large.

Based on the simulations, we can see that all of the methods improved the PMSEs of the LS estimators in almost all cases. All of the methods greatly improved the PMSE when n becomes small or k becomes large. Moreover, the improvement in the PMSE of the proposed method increases as ρ_y decreases. The improvement in the PMSE when $\kappa \neq 0$ and $\delta \neq 0$ of the proposed method increases as ρ_x increases. Comparison of several methods reveals that Methods 2 and 4 were better than Methods 1 and 3, respectively, in almost all cases when ρ_x is large. When k and ρ_x become large, Methods 5 and 6 provide a greater improvement in PMSE than Methods 3 and 4. When k becomes small and n becomes large, Methods 4 and 6 improve the PMSE more than Methods 2 and 4 in most cases. Occasionally, Method 7 improves the PMSE more than Method 8, especially when κ and δ become large. Consequently, Method 6 was, on average, the best method. In particular, it strongly improved the PMSE when δ and κ are small. Based on these results, we recommend using Method 6 to optimize the multivariate ridge regression.

Acknowledgments

The author wish express my deepest gratitude to Dr. Hirokazu Yanagihara of Hiroshima University for his valuable advice and encouragement and introducing me to various fields of mathematical statistics. In addition, I would like to thank Professor Hirofumi Wakaki for previous comments and Professor Yasunori Fujikoshi of Hiroshima University for providing helpful comments and suggestions. Also, I thank to Dr. Kengo Kato of Hiroshima University for his encouragement and comments. Futhemore, I sincere thank Professor Megu Ohtaki, Dr. Kenichi Satoh, and Dr. Tetsuji Tonda of Hiroshima University for their encouragement and valuable comments and introducing me to various fields of mathematical statistics. I also would like to thank Dr. Tomoyuki Akita, Dr. Shinobu Fujii, and Dr. Kunihiko Taniguchi of Hiroshima University for their advice and encouragement. Special thanks are due to Professor Takashi Kanda of Hiroshima Institute of Technology for his encouragement and his recommendation to attend Hiroshima University.

Appendix

A.1. Proof of Theorem 2.1

In this subsection, we prove Theorem 2.1, which shows the properties of $\hat{\theta}(\lambda)$. Using d_j

and $\boldsymbol{z}_j$ in (2.4), we can write $\mathrm{tr}(\boldsymbol{W}_\theta \boldsymbol{S}^{-1})$ and $\mathrm{tr}(\boldsymbol{M}_\theta^{-1}\boldsymbol{M}_0)$ in (2.1) as

$$g(\theta) = \mathrm{tr}(\boldsymbol{W}_\theta \boldsymbol{S}^{-1}) = \mathrm{tr}(\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{S}^{-1}) - 2\sum_{j=1}^{k} \frac{\|\boldsymbol{z}_j\|^2}{d_j+\theta} + \sum_{j=1}^{k} \frac{\|\boldsymbol{z}_j\|^2 d_j}{(d_j+\theta)^2} - n\hat{\boldsymbol{\mu}}'\boldsymbol{S}^{-1}\hat{\boldsymbol{\mu}},$$

$$h(\theta) = \mathrm{tr}(\boldsymbol{M}_\theta^{-1}\boldsymbol{M}_0) = \sum_{j=1}^{k} \frac{d_j}{d_j+\theta}.$$

Since $d_j > 0$ and $\theta \geq 0$, we have

$$\dot{g}(\theta) = \frac{\partial g(\theta)}{\partial \theta} = 2\theta \sum_{j=1}^{k} \frac{\|\boldsymbol{z}_j\|^2}{(d_j+\theta)^3} \geq 0, \tag{A.1}$$

with equality if and only if $\theta = 0$ or $\theta \to \infty$, and

$$\dot{h}(\theta) = \frac{\partial h(\theta)}{\partial \theta} = -\sum_{j=1}^{k} \frac{d_j}{(d_j+\theta)^2} \leq 0, \tag{A.2}$$

with equality if and only if $\theta \to \infty$. Therefore, $g(\theta)$ and $h(\theta)$ are strictly monotonic increasing and decreasing functions of $\theta \in [0, \infty]$, respectively. Since $GC_p(\theta, \lambda) = \lambda g(\theta) + 2ph(\theta)$, these results imply that

$$\hat{\theta}(\infty) = \arg\min_{\theta \in [0,\infty]} GC_p(\theta, \infty) = \arg\min_{\theta \in [0,\infty]} \mathrm{tr}(\boldsymbol{W}_\theta \boldsymbol{S}^{-1}) = 0,$$
$$\hat{\theta}(0) = \arg\min_{\theta \in [0,\infty]} GC_p(\theta, 0) = \arg\min_{\theta \in [0,\infty]} \mathrm{tr}(\boldsymbol{M}_\theta^{-1}\boldsymbol{M}_0) = \infty.$$

On the other hand, from (A.1) and (A.2), we derive

$$\frac{\partial GC_p(\theta, \lambda)}{\partial \theta} = \dot{GC}_p(\theta, \lambda) = 2\sum_{j=1}^{k} \frac{pd_j^2(\theta r_{\lambda,j} - 1)}{(d_j+\theta)^3} = \sum_{j=1}^{k} \phi_j(\theta|\lambda),$$

where $r_{\lambda,j}$ is given by (2.5). Note that $\phi_j(\theta|\lambda) \leq 0$ when $\theta \in [0, (r_{\lambda,1}^+)^{-1}]$. Therefore, $\dot{GC}_p(\theta, \lambda) < 0$ when $\theta \in [0, (r_{\lambda,1}^+)^{-1}]$. These imply that $GC_p(\theta, \lambda)$ is a monotonic decreasing function with respect to $\theta \in [0, (r_{\lambda,1}^+)^{-1}]$. Thus, $\hat{\theta}(\lambda) > (r_{\lambda,1}^+)^{-1}$. On the other hand, if $\max_{j=1,\ldots,k} r_{\lambda,j} \leq 0$ is satisfied, $\phi_j(\theta|\lambda) \leq 0$ holds for any θ. This fact means $GC_p(\theta, \lambda)$ is a monotonic decreasing function with respect to $\theta \in [0, \infty]$. Hence, we can see that $\hat{\theta}(\lambda) = \infty$ when $\max_{j=1,\ldots,k} r_{\lambda,j} \leq 0$. Since $\max_{j=1,\ldots,k} r_{\lambda,j} \leq 0$ holds when $\lambda < \min_{j=1,\ldots,k} pd_j/\|\boldsymbol{z}_j\|^2$. Thus, $\hat{\theta}(\lambda) = \infty$ when $\lambda < \min_{j=1,\ldots,k} pd_j/\|\boldsymbol{z}_j\|^2$ is satisfied.

Using Equation (2.3) and $\dot{GC}_p(\theta, \lambda) = \lambda \dot{g}(\theta) + 2p\dot{h}(\theta)$, we have

$$\frac{\partial}{\partial \lambda} \dot{GC}_p(\hat{\theta}(\lambda), \lambda) = \dot{g}(\hat{\theta}(\lambda)) + \frac{\partial \hat{\theta}(\lambda)}{\partial \lambda} \ddot{GC}_p(\hat{\theta}(\lambda), \lambda) = 0, \tag{A.3}$$

where $\ddot{GC}_p(\theta, \lambda) = \partial^2 GC_p(\theta, \lambda)/(\partial \theta^2)$. Since $\hat{\theta}(\lambda)$ satisfies (2.2), i.e., $\hat{\theta}(\lambda)$ is the minimizer of $GC_p(\theta, \lambda)$, $GC_p(\theta, \lambda)$ is a convex function around the neighborhood of $\hat{\theta}(\lambda)$. Hence, we have $\ddot{GC}_p(\hat{\theta}(\lambda), \lambda) > 0$. Using this result and Equation (A.3), we obtain

$$\frac{\partial \hat{\theta}(\lambda)}{\partial \lambda} = -\frac{\dot{g}(\hat{\theta}(\lambda))}{\ddot{GC}_p(\hat{\theta}(\lambda), \lambda)}.$$

We derive $\dot{g}(\hat{\theta}(\lambda)) \geq 0$ because $\dot{g}(\theta)$ is a strictly monotonic decreasing function of $\theta \in [0, \infty]$. Hence, $\partial\hat{\theta}(\lambda)/(\partial\lambda) \leq 0$ is obtained. This implies that $\hat{\theta}(\lambda)$ is a monotonic decreasing function with respect to λ.

A.2. Proof of Theorem 2.2

In this subsection, we present the proof of Theorem 2.2, which describes the expansion of $\hat{\theta}(\lambda)$. In order to prove this theorem, we expand the GC_p criterion in (2.1) under fixed λ. Recall that $\boldsymbol{X}'\boldsymbol{1}_n = \boldsymbol{0}_k$, $(n-k-1)\boldsymbol{S} = \boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n'/n - \boldsymbol{X}\boldsymbol{M}_0^{-1}\boldsymbol{X}')\boldsymbol{Y}$, and $\boldsymbol{Q}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{Q} = \boldsymbol{D}$. Hence, we derive

$$\begin{aligned}
\boldsymbol{W}_\theta \boldsymbol{S}^{-1} &= \boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n'/n - \boldsymbol{X}\boldsymbol{M}_\theta^{-1}\boldsymbol{X}')^2\boldsymbol{Y}\boldsymbol{S}^{-1} \\
&= (n-k-1)\boldsymbol{I}_p + \boldsymbol{Y}'\boldsymbol{X}(\boldsymbol{M}_0^{-1} - 2\boldsymbol{M}_\theta^{-1} + \boldsymbol{M}_\theta^{-1}\boldsymbol{M}_0\boldsymbol{M}_\theta^{-1})\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{S}^{-1} \\
&= (n-k-1)\boldsymbol{I}_p + \boldsymbol{Y}'\boldsymbol{X}\boldsymbol{Q}\boldsymbol{D}^{-1/2}\{\boldsymbol{I}_k - \boldsymbol{D}(\boldsymbol{D}+\theta\boldsymbol{I}_k)^{-1}\}^2\boldsymbol{D}^{-1/2}\boldsymbol{Q}'\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{S}^{-1}.
\end{aligned}$$

Based on this result, the GC_p criterion is expressed as

$$GC_p(\theta, \lambda) = \lambda(n-k-1)p + \lambda\theta^2\mathrm{tr}\{\boldsymbol{D}^{-1/2}\boldsymbol{Q}'\boldsymbol{V}\boldsymbol{Q}\boldsymbol{D}^{-1/2}(\boldsymbol{D}+\theta\boldsymbol{I}_k)^{-2}\} + 2p\{\boldsymbol{D}(\boldsymbol{D}+\theta\boldsymbol{I}_k)^{-1}\}.$$

Letting $t_j = (\boldsymbol{D}^{-1/2}\boldsymbol{Q}'\boldsymbol{V}\boldsymbol{Q}\boldsymbol{D}^{-1/2})_{jj}$, we obtain

$$GC_p(\theta, \lambda) = \lambda(n-k-1)p + \sum_{i=1}^{k}\left\{\lambda\left(1+\frac{\theta}{d_i}\right)^{-2}\frac{\theta^2 t_i}{d_i^2} + 2p\left(1+\frac{\theta}{d_i}\right)^{-1}\right\}.$$

By Taylor expansion around $\theta = 0$, we have

$$\begin{aligned}
GC_p(\theta, \lambda) &= \lambda(n-k-1)p + \lambda\sum_{i=1}^{k}\sum_{\ell=1}^{\infty}\frac{(-1)^{\ell+1}\ell\theta^{\ell+1}}{d_i^{\ell+1}}t_i + 2p\sum_{i=1}^{k}\left(1 - \sum_{\ell=1}^{\infty}\frac{(-1)^{\ell+1}}{d_i^\ell}\theta^\ell\right) \\
&= (\lambda(n-k-1)+2k)p + \sum_{\ell=1}^{\infty}\left\{\lambda(-1)^{\ell+1}\ell\theta^{\ell+1}\sum_{i=1}^{k}\frac{t_i}{d_i^{\ell+1}} - 2p(-1)^{\ell+1}\theta^\ell\sum_{i=1}^{k}\frac{1}{d_i^\ell}\right\} \\
&= (\lambda(n-k-1)+2k)p + \sum_{\ell=1}^{\infty}(-1)^{\ell+1}\theta^\ell\{\lambda\ell\theta\mathrm{tr}(\boldsymbol{V}\boldsymbol{M}_0^{-(\ell+2)}) - 2p\mathrm{tr}(\boldsymbol{M}_0^{-\ell})\}.
\end{aligned}$$

Recall that $a_j = n^j\mathrm{tr}(\boldsymbol{V}\boldsymbol{M}_0^{-(j+2)})$ and $b_j = n^j\mathrm{tr}(\boldsymbol{M}_0^{-j})$. It follows that

$$GC_p(\theta, \lambda) = (\lambda(n-k-1)+2k)p + \lim_{L\to\infty}\sum_{\ell=1}^{L}\frac{(-1)^{\ell+1}\theta^\ell}{n^\ell}\{\lambda\ell\theta a_\ell - 2pb_\ell\}.$$

Then, the following equation is derived:

$$\frac{\partial}{\partial\theta}GC_p(\theta,\lambda)=\lim_{L\to\infty}\sum_{\ell=1}^{L}\frac{(-1)^{\ell+1}\ell\theta^{\ell-1}}{n^{\ell}}\{\lambda(\ell+1)a_\ell\theta-2pb_\ell\}.$$

Using the above equation and $\hat{\theta}(\lambda)$ satisfying (2.3), we obtain the equation in Theorem 2.2.

A.3. Proof of Theorem 3.1

In this subsection, we prove Theorem 3.1, which shows the risk function with respect to λ. Recall that $g(\theta)=\mathrm{tr}(\boldsymbol{W}_\theta\boldsymbol{S}^{-1})$, $h(\theta)=\mathrm{tr}(\boldsymbol{M}_\theta^{-1}\boldsymbol{M}_0)$ and $GC_p(\theta,\lambda)=\lambda g(\theta)+2ph(\theta)$. Since $\hat{\theta}(\lambda)$ satisfies (2.3), we obtain

$$\begin{aligned}0&=\frac{\partial}{\partial(\boldsymbol{Y})_{ij}}\left(\lambda\left.\frac{\partial g(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}(\lambda)}+2p\left.\frac{\partial h(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}(\lambda)}\right)\\&=\frac{\partial\hat{\theta}(\lambda)}{\partial(\boldsymbol{Y})_{ij}}\left\{\lambda\ddot{g}(\hat{\theta}(\lambda))+2p\ddot{h}(\hat{\theta}(\lambda))\right\}+\lambda\frac{\partial\dot{g}(\hat{\theta}(\lambda))}{\partial(\boldsymbol{Y})_{ij}},\end{aligned}$$

where $\dot{g}(\hat{\theta}(\lambda))=\partial g(\theta)/(\partial\theta)|_{\theta=\hat{\theta}(\lambda)}$, $\ddot{g}(\hat{\theta}(\lambda))=\partial^2 g(\theta)/(\partial\theta)^2|_{\theta=\hat{\theta}(\lambda)}$, and $\dot{h}(\hat{\theta}(\lambda))=\partial h(\theta)/(\partial\theta)$ $|_{\theta=\hat{\theta}(\lambda)}$. Thus, we obtain

$$\frac{\partial\hat{\theta}(\lambda)}{\partial(\boldsymbol{Y})_{ij}}=-\frac{\lambda(\partial\dot{g}(\hat{\theta}(\lambda)))/(\partial(\boldsymbol{Y})_{ij})}{G\ddot{C}_p(\hat{\theta}(\lambda),\lambda)},$$

because, from Appendix A.1, $G\ddot{C}_p(\hat{\theta}(\lambda),\lambda)=\lambda\ddot{g}(\hat{\theta}(\lambda))+2p\ddot{h}(\hat{\theta}(\lambda))>0$. By simple calculation, we have $\dot{g}(\hat{\theta}(\lambda))=2\hat{\theta}(\lambda)\mathrm{tr}(\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-3}\boldsymbol{V})$. As in the proof of Lemma 3.1, which is given in Appendix A.4, we obtain $\partial\mathrm{tr}(\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-3}\boldsymbol{V})/(\partial(\boldsymbol{Y})_{ij})=2(\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{X}\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-3}\boldsymbol{X}'\{\boldsymbol{I}_n-\boldsymbol{Y}\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{H}/(n-k-1)\})_{ji}$. Hence, we derive

$$\sum_{i=1}^{n}\sum_{j=1}^{p}(\boldsymbol{X}\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-2}\boldsymbol{X}'\boldsymbol{Y})_{ij}\frac{\partial\hat{\theta}(\lambda)}{\partial(\boldsymbol{Y})_{ij}}=-\frac{4\lambda\hat{\theta}(\lambda)\mathrm{tr}(\boldsymbol{V}\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-5}\boldsymbol{M}_0)}{G\ddot{C}_p(\hat{\theta}(\lambda),\lambda)},$$

because $\boldsymbol{X}'\boldsymbol{1}_n=\boldsymbol{0}_k$ and $\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-3}\boldsymbol{M}_0\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-2}=\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-5}\boldsymbol{M}_0$. By simple calculation, we have $G\dot{C}_p(\theta,\lambda)=\lambda\dot{g}(\theta)+2p\dot{h}(\theta)=2\{\lambda\theta\mathrm{tr}(\boldsymbol{M}_\theta^{-3}\boldsymbol{V})-p\mathrm{tr}(\boldsymbol{M}_\theta^{-2}\boldsymbol{M}_0)\}$ and

$$G\ddot{C}_p(\hat{\theta}(\lambda),\lambda)=2\{\lambda\mathrm{tr}(\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-3}\boldsymbol{V})-3\lambda\hat{\theta}(\lambda)\mathrm{tr}(\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-4}\boldsymbol{V})+2p\mathrm{tr}(\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-3}\boldsymbol{M}_0)\}.$$

Thus, the theorem is proved.

A.4. Proof of Lemma 3.1

In this subsection, we prove Lemma 3.1, which shows the derivative of a_ℓ for any ℓ. Since $a_\ell=n^\ell\mathrm{tr}(\boldsymbol{V}\boldsymbol{M}_0^{-(\ell+2)})$, $\boldsymbol{M}_0=\boldsymbol{X}'\boldsymbol{X}$, and $\boldsymbol{V}=\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{X}$, we need only obtain the derivative of $\boldsymbol{V}$. We can see that

$$\begin{aligned}\frac{\partial \boldsymbol{S}^{-1}}{\partial(\boldsymbol{Y})_{ij}} &= -\boldsymbol{S}^{-1}\frac{\partial \boldsymbol{S}}{\partial(\boldsymbol{Y})_{ij}}\boldsymbol{S}^{-1} \\ &= -\frac{1}{n-k-1}\boldsymbol{S}^{-1}(\boldsymbol{e}_{j\cdot p}\boldsymbol{e}_{i\cdot n}'\boldsymbol{H}\boldsymbol{Y}+\boldsymbol{Y}'\boldsymbol{H}\boldsymbol{e}_{i\cdot n}\boldsymbol{e}_{j\cdot p}')\boldsymbol{S}^{-1},\end{aligned}$$

where $\boldsymbol{e}_{i\cdot n}$ is the n-dimensional vector, the ith element of which is one and other elements of which are zeros. Thus, we obtain

$$\begin{aligned}\frac{\partial \boldsymbol{V}}{\partial(\boldsymbol{Y})_{ij}} &= \boldsymbol{X}'\boldsymbol{e}_{i\cdot n}\boldsymbol{e}_{j\cdot p}'\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{X}+\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{S}^{-1}\boldsymbol{e}_{j\cdot p}\boldsymbol{e}_{i\cdot n}'\boldsymbol{X} \\ &\quad -\frac{1}{n-k-1}\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{S}^{-1}(\boldsymbol{e}_{j\cdot p}\boldsymbol{e}_{i\cdot n}'\boldsymbol{H}\boldsymbol{Y}+\boldsymbol{Y}'\boldsymbol{H}\boldsymbol{e}_{i\cdot n}\boldsymbol{e}_{j\cdot p}')\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{X}.\end{aligned}$$

From $\partial a_\ell/(\partial(\boldsymbol{Y})_{ij}) = n^\ell \mathrm{tr}\{\boldsymbol{M}_0^{-(\ell+2)}(\partial\boldsymbol{V})/(\partial(\boldsymbol{Y})_{ij})\}$, we derive this lemma.

A.5. Proof of Theorem 3.2

From (3.2), we obtain

$$\begin{aligned}\mathrm{PMSE}[\hat{\boldsymbol{Y}}_{\tilde{\theta}_{(L)}(\lambda)}] &= E_{\boldsymbol{Y}}[\mathrm{tr}(\boldsymbol{W}_{\tilde{\theta}_{(L)}(\lambda)}\boldsymbol{\Sigma}^{-1})+2p\{\mathrm{tr}(\boldsymbol{M}_{\tilde{\theta}_{(L)}(\lambda)}^{-1}\boldsymbol{M}_0)+1\}] \\ &\quad -2E_{\boldsymbol{Y}}\left[\sum_{i=1}^{n}\sum_{j=1}^{p}(\boldsymbol{X}\boldsymbol{M}_{\tilde{\theta}_{(L)}(\lambda)}^{-2}\boldsymbol{X}'\boldsymbol{Y})_{ij}\frac{\partial\tilde{\theta}_{(L)}(\lambda)}{\partial(\boldsymbol{Y})_{ij}}\right].\end{aligned}$$

Hence, we need only calculate $\partial\tilde{\theta}_{(L)}(\lambda)/(\partial(\boldsymbol{Y})_{ij})$. Using Theorem 2.2, we derive the derivative as follows:

$$\begin{aligned}\frac{\partial\tilde{\theta}_{(L)}(\lambda)}{\partial(\boldsymbol{Y})_{ij}} &= \frac{\partial\tilde{\theta}_{(1)}(\lambda)}{\partial(\boldsymbol{Y})_{ij}} \\ &\quad +\frac{\partial}{\partial(\boldsymbol{Y})_{ij}}\left[\frac{1}{2\lambda a_1}\sum_{\ell=0}^{L-1}\frac{1}{n^\ell}(-1)^{\ell+1}(\ell+1)\ \tilde{\theta}_{(L-1)}^{\ell}(\lambda)\{\lambda(\ell+2)a_{\ell+1}\tilde{\theta}_{(L-1)}(\lambda)-2pb_{\ell+1}\}\right].\end{aligned}$$

Recall that $\tilde{\theta}_{(0)}(\lambda)=0$. From Lemma 3.1, we have

$$\begin{aligned}\frac{\partial\tilde{\theta}_{(1)}(\lambda)}{\partial(\boldsymbol{Y})_{ij}} &= -\frac{pb_1}{\lambda a_1^2}\frac{\partial a_1}{\partial(\boldsymbol{Y})_{ij}} \\ &= -\frac{2n\tilde{\theta}_{(1)}(\lambda)}{a_1}\left(\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{X}\boldsymbol{M}_0^{-3}\boldsymbol{X}'\left(\boldsymbol{I}_n-\frac{1}{n-k-1}\boldsymbol{Y}\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{H}\right)\right)_{ji},\end{aligned}$$

and

$$\frac{\partial}{\partial(\boldsymbol{Y})_{ij}}\left[\frac{1}{2\lambda a_1}\sum_{\ell=0}^{L-1}\frac{1}{n^\ell}(-1)^{\ell+1}(\ell+1)\tilde{\theta}^\ell_{(L-1)}(\lambda)\{\lambda(\ell+2)a_{\ell+1}\tilde{\theta}_{(L-1)}(\lambda)-2pb_{\ell+1}\}\right]$$
$$=-\frac{n}{\lambda a_1^2}\left(\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{X}\boldsymbol{M}_0^{-3}\boldsymbol{X}'\left(\boldsymbol{I}_n-\frac{\boldsymbol{Y}\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{H}}{n-k-1}\right)\right)_{ji}$$
$$\times\sum_{\ell=0}^{L-1}\frac{1}{n^\ell}(-1)^{\ell+1}(\ell+1)\tilde{\theta}^\ell_{(L-1)}(\lambda)\{\lambda(\ell+2)a_{\ell+1}\tilde{\theta}_{(L-1)}(\lambda)-2pb_{\ell+1}\}$$
$$+\frac{1}{2\lambda a_1}\sum_{\ell=0}^{L-1}\frac{1}{n^\ell}(-1)^{\ell+1}(\ell+1)\tilde{\theta}^{\ell-1}_{(L-1)}\{\lambda(\ell+1)(\ell+2)a_{\ell+1}\tilde{\theta}_{(L-1)}(\lambda)-2p\ell b_{\ell+1}\}\frac{\partial\tilde{\theta}_{(L-1)}}{\partial(\boldsymbol{Y})_{ij}}$$
$$+\frac{n}{a_1}\sum_{\ell=0}^{L-1}(-1)^{\ell+1}(\ell+1)(\ell+2)\tilde{\theta}^{\ell+1}_{(L-1)}\left(\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{X}\boldsymbol{M}_0^{-(\ell+3)}\boldsymbol{X}'\left(\boldsymbol{I}_n-\frac{1}{n-k-1}\boldsymbol{Y}\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{H}\right)\right)_{ji}.$$

Thus, we obtain

$$\sum_{i=1}^n\sum_{j=1}^p(\boldsymbol{X}\boldsymbol{M}^{-2}_{\tilde{\theta}_{(L)}(\lambda)}\boldsymbol{X}'\boldsymbol{Y})_{ij}$$
$$\times\frac{\partial}{\partial(\boldsymbol{Y})_{ij}}\left[\frac{1}{2\lambda a_1}\sum_{\ell=0}^{L-1}\frac{1}{n^\ell}(-1)^{\ell+1}(\ell+1)\ \tilde{\theta}^\ell_{(L-1)}(\lambda)\{\lambda(\ell+2)a_{\ell+1}\tilde{\theta}_{(L-1)}(\lambda)-2pb_{\ell+1}\}\right]$$
$$=-\frac{n}{\lambda a_1^2}\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}^{-2}_{\tilde{\theta}_{(L)}(\lambda)}\boldsymbol{V})\sum_{\ell=0}^{L-1}\frac{1}{n^\ell}(-1)^{\ell+1}(\ell+1)\tilde{\theta}^\ell_{(L-1)}(\lambda)\{\lambda(\ell+2)a_{\ell+1}\tilde{\theta}_{(L-1)}(\lambda)-2pb_{\ell+1}\}$$
$$+\frac{1}{2\lambda a_1}\sum_{\ell=0}^{L-1}\frac{1}{n^\ell}(-1)^{\ell+1}(\ell+1)\tilde{\theta}^{\ell-1}_{(L-1)}\{\lambda(\ell+1)(\ell+2)a_{\ell+1}\tilde{\theta}_{(L-1)}(\lambda)-2p\ell b_{\ell+1}\}$$
$$\times\sum_{i=1}^n\sum_{j=1}^p(\boldsymbol{X}\boldsymbol{M}^{-2}_{\tilde{\theta}_{(L)}(\lambda)}\boldsymbol{X}'\boldsymbol{Y})_{ij}\frac{\partial\tilde{\theta}_{(L-1)}}{\partial(\boldsymbol{Y})_{ij}}$$
$$+\frac{n}{a_1}\sum_{\ell=0}^{L-1}(-1)^{\ell+1}(\ell+1)(\ell+2)\tilde{\theta}^{\ell+1}_{(L-1)}\mathrm{tr}(\boldsymbol{M}_0^{-(\ell+2)}\boldsymbol{M}^{-2}_{\tilde{\theta}_{(L)}(\lambda)}\boldsymbol{V}),$$

and

$$\sum_{i=1}^n\sum_{j=1}^p(\boldsymbol{X}\boldsymbol{M}^{-2}_{\tilde{\theta}_{(L)}(\lambda)}\boldsymbol{X}'\boldsymbol{Y})_{ij}\frac{\partial\tilde{\theta}_{(1)}(\lambda)}{\partial(\boldsymbol{Y})_{ij}}=-\frac{2n\tilde{\theta}_{(1)}(\lambda)}{a_1}\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}^{-2}_{\tilde{\theta}_{(L)}(\lambda)}\boldsymbol{V}).$$

We derive this theorem by substituting these results into (3.2).

A.6. The criteria for optimizing λ when we use $\tilde{\theta}_{(2)}(\lambda)$

In this subsection, we calculate the criterion for optimizing λ when we use $\tilde{\theta}_{(2)}(\lambda)$. From (2.6), we obtain

$$\tilde{\theta}_{(2)}(\lambda)=\tilde{\theta}_{(1)}(\lambda)+\frac{1}{n\lambda a_1}\tilde{\theta}_{(1)}(\lambda)\{3\lambda a_2\tilde{\theta}_{(1)}(\lambda)-2pb_2\}$$
$$=\tilde{\theta}_{(1)}(\lambda)+\frac{1}{n}\tilde{\theta}^2_{(1)}(\lambda)\left\{3\frac{a_2}{a_1}-2\frac{b_2}{b_1}\right\}.$$

Since $b_\ell = n^\ell \mathrm{tr}(\boldsymbol{M}_0^{-\ell})$ does not depend on $(\boldsymbol{Y})_{ij}$, we derive

$$\begin{aligned}
\frac{\partial\tilde{\theta}_{(2)}(\lambda)}{\partial(\boldsymbol{Y})_{ij}} &= \frac{\partial\tilde{\theta}_{(1)}(\lambda)}{\partial(\boldsymbol{Y})_{ij}} + \frac{1}{n}\frac{\partial}{\partial(\boldsymbol{Y})_{ij}}\left\{\tilde{\theta}_{(1)}^2(\lambda)\left(3\frac{a_2}{a_1} - 2\frac{b_2}{b_1}\right)\right\} \\
&= \frac{\partial\tilde{\theta}_{(1)}(\lambda)}{\partial(\boldsymbol{Y})_{ij}} + \frac{1}{n}\left\{\frac{\partial\tilde{\theta}_{(1)}^2(\lambda)}{\partial(\boldsymbol{Y})_{ij}}\left(3\frac{a_2}{a_1} - 2\frac{b_2}{b_1}\right) + 3\tilde{\theta}_{(1)}^2(\lambda)\frac{\partial\ a_2/a_1}{\partial(\boldsymbol{Y})_{ij}}\right\} \\
&= \frac{\partial\tilde{\theta}_{(1)}(\lambda)}{\partial(\boldsymbol{Y})_{ij}}\left\{1 + \frac{2\tilde{\theta}_{(1)}(\lambda)}{n}\left(3\frac{a_2}{a_1} - 2\frac{b_2}{b_1}\right)\right\} + \frac{3\tilde{\theta}_{(1)}^2(\lambda)}{n}\frac{\partial a_2/a_1}{\partial(\boldsymbol{Y})_{ij}}.
\end{aligned}$$

Hence, the third term of (3.1) is obtained using the result of $\partial\tilde{\theta}_{(1)}(\lambda)/(\partial(\boldsymbol{Y})_{ij})$ as follows:

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{p}(\boldsymbol{X}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{X}'\boldsymbol{Y})_{ij}\frac{\partial\tilde{\theta}_{(2)}(\lambda)}{\partial(\boldsymbol{Y})_{ij}} \\
&= -\frac{2n\tilde{\theta}_{(1)}(\lambda)}{a_1}\left\{1 + \frac{2\tilde{\theta}_{(1)}(\lambda)}{n}\left(3\frac{a_2}{a_1} - 2\frac{b_2}{b_1}\right)\right\}\mathrm{tr}\left(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}\right) \\
&\quad + \frac{3\tilde{\theta}_{(1)}^2(\lambda)}{na_1^2}\sum_{i=1}^{n}\sum_{j=1}^{p}(\boldsymbol{X}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{X}'\boldsymbol{Y})_{ij}\left(a_1\frac{\partial a_2}{\partial(\boldsymbol{Y})_{ij}} - a_2\frac{\partial a_1}{\partial(\boldsymbol{Y})_{ij}}\right).
\end{aligned}$$

From Lemma 3.1, we obtain

$$\begin{aligned}
&\frac{3\tilde{\theta}_{(1)}^2(\lambda)}{na_1^2}\sum_{i=1}^{n}\sum_{j=1}^{p}(\boldsymbol{X}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{X}'\boldsymbol{Y})_{ij}\left(a_1\frac{\partial a_2}{\partial(\boldsymbol{Y})_{ij}} - a_2\frac{\partial a_1}{\partial(\boldsymbol{Y})_{ij}}\right) \\
&= \frac{6\tilde{\theta}_{(1)}^2(\lambda)}{a_1^2}\sum_{i=1}^{n}\sum_{j=1}^{p}(\boldsymbol{X}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{X}'\boldsymbol{Y})_{ij} \\
&\quad \times\left(\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{X}\boldsymbol{M}_0^{-3}\left(na_1\boldsymbol{M}_0^{-1} - a_2\boldsymbol{I}_k\right)\boldsymbol{X}'\left(\boldsymbol{I}_n - \frac{1}{n-k-1}\boldsymbol{Y}\boldsymbol{S}^{-1}\boldsymbol{Y}'\boldsymbol{H}\right)\right)_{ji} \\
&= \frac{6\tilde{\theta}_{(1)}^2(\lambda)}{a_1^2}\mathrm{tr}\{\boldsymbol{X}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}\boldsymbol{M}_0^{-3}(na_1\boldsymbol{M}_0^{-1} - a_2\boldsymbol{I}_k)\boldsymbol{X}'\} \\
&= \frac{6\tilde{\theta}_{(1)}^2(\lambda)}{a_1^2}\mathrm{tr}\{\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}(na_1\boldsymbol{M}_0^{-1} - a_2\boldsymbol{I}_k)\}.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&-\sum_{i=1}^{n}\sum_{j=1}^{p}(\boldsymbol{X}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{X}'\boldsymbol{Y})_{ij}\frac{\partial\tilde{\theta}_{(2)}(\lambda)}{\partial(\boldsymbol{Y})_{ij}} \\
&= \frac{2n\tilde{\theta}_{(1)}(\lambda)}{a_1}\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}) + \frac{4\tilde{\theta}_{(1)}^2(\lambda)}{a_1}\left(3\frac{a_2}{a_1} - 2\frac{b_2}{b_1}\right)\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}) \\
&\quad + \frac{6\tilde{\theta}_{(1)}^2(\lambda)}{a_1^2}\mathrm{tr}\{\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}(a_2\boldsymbol{I}_k - na_1\boldsymbol{M}_0^{-1})\} \\
&= \frac{2n\tilde{\theta}_{(1)}(\lambda)}{a_1}\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}) + \frac{18\tilde{\theta}_{(1)}^2(\lambda)a_2}{a_1^2}\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}) \\
&\quad - \frac{8\tilde{\theta}_{(1)}^2(\lambda)b_2}{a_1b_1}\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}) - \frac{6n\tilde{\theta}_{(1)}^2(\lambda)}{a_1}\mathrm{tr}(\boldsymbol{M}_0^{-3}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}).
\end{aligned}$$

When we use Theorem 3.2, we obtain the same result. Using this result, we obtain the C_p type criteria for optimizing λ are defined as

$$
\begin{aligned}
C_p^{(2)}(\lambda) &= \mathrm{tr}(\boldsymbol{W}_{\tilde{\theta}_{(1)}(\lambda)}\boldsymbol{S}^{-1}) + 2p\mathrm{tr}(\boldsymbol{M}_{\tilde{\theta}_{(2)}\lambda}^{-1}\boldsymbol{M}_0) \\
&\quad + \frac{4n\tilde{\theta}_{(1)}(\lambda)}{a_1}\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}) + \frac{36\tilde{\theta}_{(1)}^2(\lambda)a_2}{a_1^2}\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}) \\
&\quad - \frac{16\tilde{\theta}_{(1)}^2(\lambda)b_2}{a_1 b_1}\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}) - \frac{12n\tilde{\theta}_{(1)}^2(\lambda)}{a_1}\mathrm{tr}(\boldsymbol{M}_0^{-3}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}),
\end{aligned}
$$

and

$$
\begin{aligned}
MC_p^{(2)}(\lambda) &= c_{\mathrm{M}}\mathrm{tr}(\boldsymbol{W}_{\tilde{\theta}_{(1)}(\lambda)}\boldsymbol{S}^{-1}) + 2p\mathrm{tr}(\boldsymbol{M}_{\tilde{\theta}_{(2)}\lambda}^{-1}\boldsymbol{M}_0) \\
&\quad + \frac{4n\tilde{\theta}_{(1)}(\lambda)}{a_1}\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}) + \frac{36\tilde{\theta}_{(1)}^2(\lambda)a_2}{a_1^2}\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}) \\
&\quad - \frac{16\tilde{\theta}_{(1)}^2(\lambda)b_2}{a_1 b_1}\mathrm{tr}(\boldsymbol{M}_0^{-2}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}) - \frac{12n\tilde{\theta}_{(1)}^2(\lambda)}{a_1}\mathrm{tr}(\boldsymbol{M}_0^{-3}\boldsymbol{M}_{\tilde{\theta}_{(2)}(\lambda)}^{-2}\boldsymbol{V}).
\end{aligned}
$$

A.7. Proof of Theorem 3.3

In this subsection, we show an asymptotic expansion $\mathrm{PMSE}[\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}]$ of and calculation to obtain $\hat{\lambda}_0$. Since $\mathrm{PMSE}[\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}]$ is obtained as (3.3), we consider expanding each term for obtaining $\hat{\lambda}_0$. We obtain

$$
\begin{aligned}
\boldsymbol{W}_{\hat{\theta}(\lambda)}\boldsymbol{\Sigma}^{-1} &= \boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n'/n - \boldsymbol{X}\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-1}\boldsymbol{X}')^2\boldsymbol{Y}\boldsymbol{\Sigma}^{-1} \\
&= (n-k-1)\boldsymbol{S}\boldsymbol{\Sigma}^{-1} + \boldsymbol{Y}'\boldsymbol{X}(\boldsymbol{M}_0^{-1} - 2\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-1} + \boldsymbol{M}_{\hat{\theta}(\lambda)}^{-1}\boldsymbol{M}_0\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-1})\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{\Sigma}^{-1} \\
&= (n-k-1)\boldsymbol{S}\boldsymbol{\Sigma}^{-1} + \boldsymbol{Y}'\boldsymbol{X}\boldsymbol{Q}\boldsymbol{D}^{-1/2}(\boldsymbol{I}_k - \boldsymbol{D}\{\boldsymbol{D} + \hat{\theta}(\lambda)\boldsymbol{I}_k)^{-1}\}^2\boldsymbol{D}^{-1/2}\boldsymbol{Q}'\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{\Sigma}^{-1},
\end{aligned}
$$

because $\boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n'/n - \boldsymbol{X}\boldsymbol{M}_0^{-1}\boldsymbol{X}')\boldsymbol{Y} = (n-k-1)\boldsymbol{S}$, $\boldsymbol{X}'\boldsymbol{1}_n = \boldsymbol{0}_k$ and $\boldsymbol{Q}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{Q} = \boldsymbol{D}$. Hence, we obtain

$$
\mathrm{tr}(\boldsymbol{W}_{\hat{\theta}(\lambda)}\boldsymbol{\Sigma}^{-1}) = (n-k-1)\mathrm{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}) + \theta^2\mathrm{tr}\{(\boldsymbol{D} + \theta\boldsymbol{I}_k)^{-2}\boldsymbol{D}^{-1}\boldsymbol{Q}'\boldsymbol{V}^*\boldsymbol{Q}\}.
$$

Since $\boldsymbol{S}$ is an unbiased estimator of $\boldsymbol{\Sigma}$, we have

$$
E_{\boldsymbol{Y}}[\mathrm{tr}(\boldsymbol{W}_{\hat{\theta}(\lambda)}\boldsymbol{\Sigma}^{-1})] = (n-k-1)p + E_{\boldsymbol{Y}}\left[\sum_{j=1}^{k}\left(\frac{\hat{\theta}(\lambda)}{d_j + \hat{\theta}(\lambda)}\right)^2\frac{(\boldsymbol{Q}'\boldsymbol{V}^*\boldsymbol{Q})_{jj}}{d_j}\right].
$$

Then, since $d_i = O(n)$ and $V^* = O_p(n^2)$, we can expand the above equation as follows:

$$
\begin{aligned}
E_{\boldsymbol{Y}}[\mathrm{tr}(\boldsymbol{W}_{\hat{\theta}(\lambda)}\boldsymbol{\Sigma}^{-1})] &= (n-k-1)p + E_{\boldsymbol{Y}}\left[\sum_{j=1}^{k}\frac{\hat{\theta}^2(\lambda)}{d_j^3}(\boldsymbol{Q}'\boldsymbol{V}^*\boldsymbol{Q})_{jj} + O_p(n^{-2})\right] \\
&= (n-k-1)p + E_{\boldsymbol{Y}}\left[\frac{a_1^*\hat{\theta}^2(\lambda)}{n} + O_p(n^{-2})\right].
\end{aligned}
$$

From simple calculation by Taylor expansion and noting that $a_j = O_p(1)$ and $b_j = O(1)$, we derive

$$
\begin{aligned}
\mathrm{tr}(\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-1}\boldsymbol{M}_0) &= k - \frac{b_1\hat{\theta}(\lambda)}{n} + O_p(n^{-2}), \\
\lambda\hat{\theta}(\lambda)\mathrm{tr}(\boldsymbol{V}\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-5}\boldsymbol{M}_0) &= \frac{\lambda\hat{\theta}(\lambda)a_2}{n^2} + O_p(n^{-3}), \\
\lambda\mathrm{tr}(\boldsymbol{V}\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-3}) &= \frac{\lambda a_1}{n} + O_p(n^{-2}), \\
\lambda\theta\mathrm{tr}(\boldsymbol{V}\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-4}) &= \frac{\lambda\hat{\theta}(\lambda)a_2}{n^2} + O_p(n^{-3}), \\
\mathrm{tr}(\boldsymbol{M}_{\hat{\theta}(\lambda)}^{-3}\boldsymbol{M}_0) &= \frac{b_2}{n^2} + O_p(n^{-3}).
\end{aligned}
$$

By substituting these results into $\mathrm{PMSE}[\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}]$ in (3.3), we obtain the asymptotic expansion of $\mathrm{PMSE}[\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}]$, as follows:

$$
\mathrm{PMSE}[\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}] = (n+k+1)p + E_{\boldsymbol{Y}}\left[\frac{a_1^*\hat{\theta}^2(\lambda)}{n} - \frac{2pb_1\hat{\theta}(\lambda)}{n} + \frac{4a_2\hat{\theta}(\lambda)}{na_1} + O_p(n^{-2})\right].
$$

From (2.6), which is proved in Appendix A.2, $\hat{\theta}(\lambda) = pb_1/(\lambda a_1) + O_p(n^{-1})$. Hence, we consider minimizing the following approximated PMSE:

$$
\mathrm{PMSE}[\hat{\boldsymbol{Y}}_{\hat{\theta}(\lambda)}] = (n+k+1)p + E_{\boldsymbol{Y}}\left[\frac{pb_1}{na_1}\left(\frac{pb_1a_1^*}{\lambda^2a_1} - \frac{2pb_1}{\lambda} + \frac{4a_2}{\lambda a_1}\right)\right] + O(n^{-2}).
$$

Hence we obtain the asymptotic optimal λ^*, which minimizes the second term of the above equation as in Theorem 3.3.

References

[1] A. C. Atkinson, A note on the generalized information criterion for choice of a model, Biometrika, **67** (1980), 413-418.

[2] P. J. Brown and J. V. Zidek, Adaptive multivariate ridge regression, Ann. Statist., **8** (1980), 64–74.

[3] B. Efron, The estimation of prediction error: covariance penalties and cross-validation, J. Amer. Statist. Assoc., **99** (2004), 619–632.

[4] S. J. V. Dien, S. Iwatani, Y. Usuda and K. Matsui, Theoretical analysis of amino acid-producing *Eschenrichia coli* using a stoixhiometrix model and multivariate linear regression, J. Biosci. Bioeng., **102** (2006), 34–40.

[5] Y. Fujikoshi and K. Satoh, Modified AIC and C_p in multivariate linear regression, Biometrika, **84** (1997), 707–716.

[6] Y. Fujikoshi, H. Yanagihara and H. Wakaki, Bias corrections of some criteria for selecting multivariate linear models in a general nonnormal case, Amer. J. Math. Management Sci., **25** (2005), 221-258.

[7] Y. Haitovsky, On multivariate ridge regression, Biometrika, **74** (1987), 563–570.

[8] A. E. Hoerl and R. W. Kennard, Ridge regression: biased estimation for nonorthogonal problems, Technometrics, **12**(1970), 55–67.

[9] J. C. Lagarias, J. A. Reeds, M. H. Wright and P. E. Wright, Convergence properties of the Nelder-Mead simplex method in low dimensions, SIAM J. Optim., **9** (1998), 112–147.

[10] J. F. Lawless, Mean squared error properties of generalized ridge estimators, J. Amer. Statist. Assoc., **76** (1981), 462–466.

[11] C. L. Mallows, Some comments on C_p, Technometrics, **15** (1973), 661–675.

[12] C. L. Mallows, More comments on C_p, Technometrics, **37** (1995), 362–372.

[13] C. Sârbu, C. Onisor, M. Posa, S. Kevresan and K. Kuhajda, Modeling and prediction (correction) of partition coefficients of bile acids and their derivatives by multivariate regression methods, Talanta, **75** (2008), 651–657.

[14] R. Saxén and J. Sundell, ^{137}Cs in freshwater fish in Finland since 1986– a statistical analysis with multivariate linear regression models, J. Environ. Radioactiv., **87** (2006), 62–76.

[15] X. Shen and J. Ye, Adaptive model selection, J. Amer. Statist. Assoc., **97** (2002), 210–221.

[16] M. Siotani, T. Hayakawa and Y. Fujikoshi, Modern Multivariate Statistical Analysis: A Graduate Course and Handbook, American Sciences Press, Columbus, Ohio, 1985.

[17] B. Skagerberg, J. MacGregor and C. Kiparissides, Multivariate data analysis applied to low-density polyethylene reactors, Chemometr. Intell. Lab. Syst., **14** (1992), 341–356.

[18] R. S. Sparks, D. Coutsourides and L. Troskie, The multivariate C_p, Comm. Statist. A – Theory Methods, **12** (1983), 1775–1793.

[19] M. S. Srivastava, Methods of Multivariate Statistics, John Wiley & Sons, New York, 2002.

[20] N. H. Timm, Applied Multivariate Analysis, Springer-Verlag, New York, 2002.

[21] H. Yanagihara and K. Satoh, An unbiased C_p criterion for multivariate ridge regression, J. Multivariate Anal., **101** (2010), 1226–1238.

[22] J. Ye, On measuring and correcting the effects of data mining and model selection, J. Amer. Statist. Assoc., **93** (1998), 120–131.

[23] A. Yoshimoto, H. Yanagihara and Y. Ninomiya, Finding factors affecting a forest stand growth through multivariate linear modeling, J. Jpn. For. Res., **87** (2005), 504–512 (in Japanese).

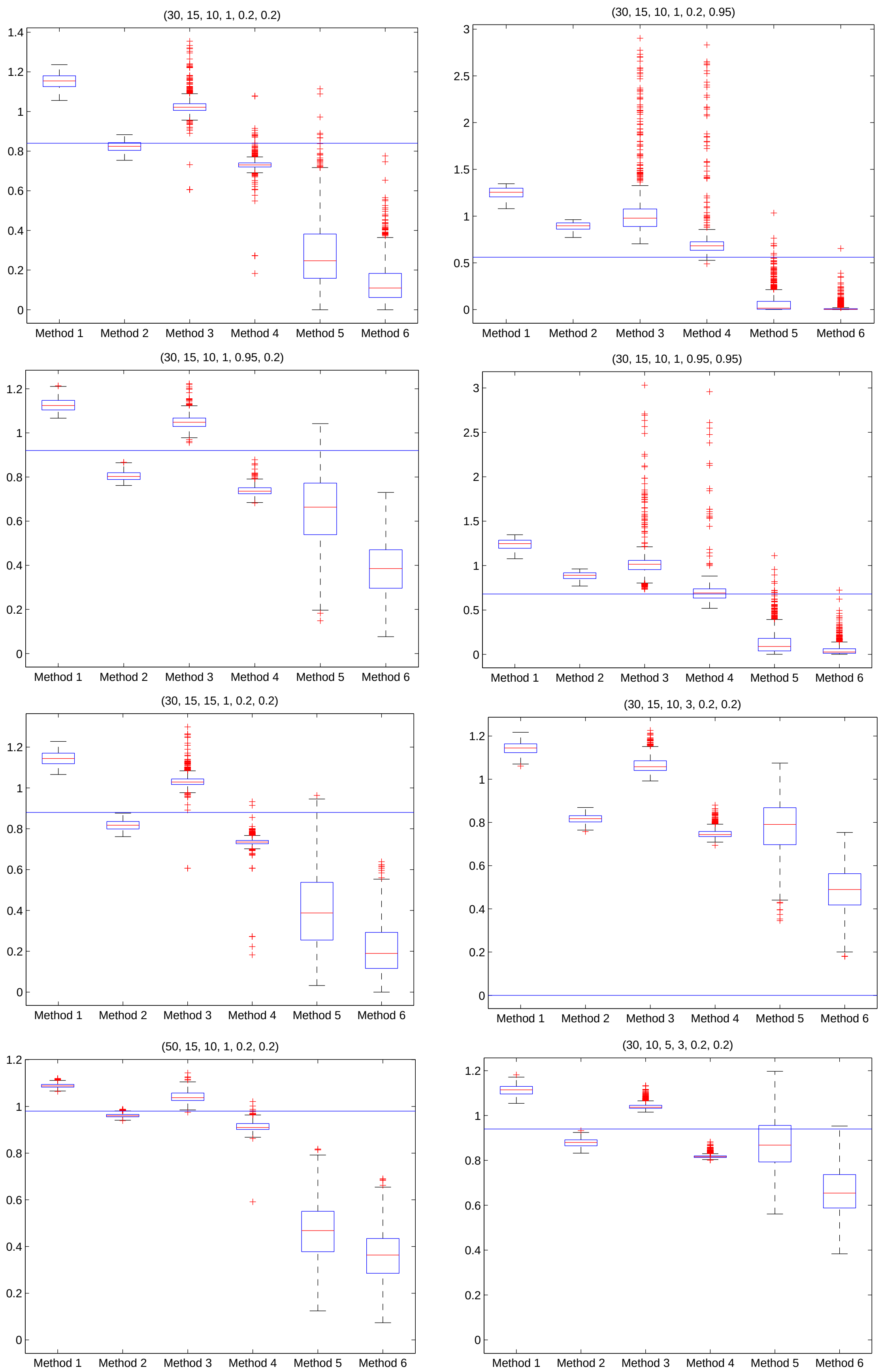

Figure 1. The box plots for $\hat{\lambda}$ in the case of $(n, k, \kappa, \delta, \rho_y, \rho_x)$

Table 1. MSE of $\hat{\lambda}$ in which $(k, n) = (5, 30)$

				Method					
κ	δ	ρ_y	ρ_x	1	2	3	4	5	6
0	0	0.2	0.2	1.470	1.014	1.475	1.145	*0.039*	**0.014**
			0.95	1.644	1.135	2.599	2.151	*0.036*	**0.013**
		0.95	0.2	1.327	0.896	1.239	0.907	*0.033*	**0.013**
			0.95	1.593	1.092	2.482	2.050	*0.035*	**0.012**
5	1	0.2	0.2	0.030	**0.002**	*0.009*	0.015	0.181	0.331
			0.95	0.305	**0.114**	0.424	*0.266*	0.378	0.445
		0.95	0.2	*0.006*	0.016	**0.001**	0.032	0.055	0.167
			0.95	0.340	**0.137**	0.327	*0.169*	0.270	0.338
	3	0.2	0.2	0.008	0.012	*0.004*	0.017	**0.001**	0.030
			0.95	1.577	1.095	1.261	0.815	*0.303*	**0.154**
		0.95	0.2	1.322	0.918	1.293	*0.894*	1.253	**0.861**
			0.95	1.516	1.053	1.359	0.927	*0.677*	**0.397**
Average				0.928	0.624	1.039	0.782	*0.272*	**0.231**

Table 2. MSE of $\hat{\lambda}$ in which $(k, n) = (5, 50)$

				Method					
κ	δ	ρ_y	ρ_x	1	2	3	4	5	6
0	0	0.2	0.2	1.394	1.148	1.504	1.277	*0.027*	**0.015**
			0.95	1.654	1.363	2.645	2.397	*0.022*	**0.012**
		0.95	0.2	1.164	0.941	1.218	0.998	*0.023*	**0.017**
			0.95	1.656	1.364	2.680	2.436	*0.020*	**0.010**
5	1	0.2	0.2	0.010	**0.001**	*0.003*	*0.003*	0.076	0.136
			0.95	*0.288*	**0.176**	0.413	0.310	0.406	0.442
		0.95	0.2	*0.003*	*0.003*	**0.001**	0.008	0.015	0.048
			0.95	0.232	*0.133*	0.187	**0.121**	0.322	0.377
	3	0.2	0.2	**0.000**	0.013	*0.001*	0.016	0.002	0.021
			0.95	1.550	1.281	1.281	1.040	*0.460*	**0.337**
		0.95	0.2	1.289	1.065	1.276	*1.054*	1.253	**1.033**
			0.95	1.483	1.226	1.335	1.100	*0.928*	**0.731**
Average				0.894	0.726	1.045	0.897	*0.296*	**0.265**

Table 3. MSE of $\hat{\lambda}$ in which $(k, n) = (10, 30)$

				Method					
κ	δ	ρ_y	ρ_x	1	2	3	4	5	6
0	0	0.2	0.2	1.293	0.798	0.977	0.773	*0.016*	**0.003**
			0.95	1.375	0.849	1.601	1.261	*0.016*	**0.004**
		0.95	0.2	1.292	0.798	0.930	0.701	*0.017*	**0.004**
			0.95	1.332	0.815	1.539	1.111	*0.017*	**0.004**
5	1	0.2	0.2	0.070	**0.001**	0.028	*0.011*	0.311	0.461
			0.95	0.328	**0.104**	0.375	*0.202*	0.311	0.355
		0.95	0.2	0.056	**0.000**	0.020	*0.007*	0.145	0.284
			0.95	0.303	**0.090**	0.283	*0.127*	0.308	0.361
	3	0.2	0.2	0.030	**0.004**	*0.011*	0.015	0.017	0.088
			0.95	0.181	*0.031*	0.110	**0.029**	0.329	0.438
		0.95	0.2	0.018	0.011	*0.007*	0.022	**0.002**	0.054
			0.95	1.369	0.853	1.130	0.673	*0.241*	**0.106**
10	1	0.2	0.2	0.035	**0.003**	*0.010*	0.019	0.267	0.450
			0.95	0.283	**0.079**	0.277	*0.141*	0.339	0.392
		0.95	0.2	0.043	**0.002**	0.015	*0.014*	0.121	0.277
			0.95	0.303	*0.090*	0.215	**0.071**	0.282	0.348
	3	0.2	0.2	1.299	0.810	1.230	*0.755*	0.936	**0.535**
			0.95	1.411	0.879	1.115	0.657	*0.122*	**0.043**
		0.95	0.2	1.307	0.815	1.259	*0.777*	1.059	**0.624**
			0.95	1.399	0.872	1.159	0.693	*0.233*	**0.094**
Average				0.686	0.395	0.615	0.403	*0.255*	**0.246**

Table 4. MSE of $\hat{\lambda}$ in which $(k, n) = (10, 50)$

				Method					
κ	δ	ρ_y	ρ_x	1	2	3	4	5	6
0	0	0.2	0.2	1.173	0.940	0.977	0.862	*0.016*	**0.007**
			0.95	1.282	1.028	1.562	1.409	*0.008*	**0.004**
		0.95	0.2	1.129	0.902	0.890	0.751	*0.014*	**0.007**
			0.95	1.285	1.031	1.598	1.403	*0.007*	**0.003**
5	1	0.2	0.2	0.012	**0.000**	0.005	*0.004*	0.200	0.287
			0.95	0.203	**0.111**	0.183	*0.126*	0.407	0.437
		0.95	0.2	0.007	**0.001**	*0.003*	0.004	0.058	0.115
			0.95	0.182	*0.096*	0.120	**0.067**	0.372	0.414
	3	0.2	0.2	0.005	*0.002*	0.003	0.003	**0.001**	0.015
			0.95	0.014	**0.000**	*0.013*	0.024	0.492	0.601
		0.95	0.2	0.006	**0.001**	0.004	*0.002*	**0.001**	0.006
			0.95	1.261	1.016	1.138	0.917	*0.455*	**0.328**
10	1	0.2	0.2	0.004	*0.002*	**0.001**	0.008	0.158	0.248
			0.95	0.171	*0.087*	0.102	**0.054**	0.426	0.465
		0.95	0.2	0.007	**0.001**	0.004	*0.003*	0.047	0.104
			0.95	0.123	*0.054*	0.063	**0.022**	0.430	0.487
	3	0.2	0.2	1.224	0.986	1.217	*0.979*	1.066	**0.8444**
			0.95	1.325	1.067	1.161	0.912	*0.229*	**0.147**
		0.95	0.2	1.211	0.976	1.206	*0.971*	1.124	**0.897**
			0.95	1.304	1.050	1.194	0.951	*0.459*	**0.325**
Average				0.596	0.468	0.572	0.474	*0.298*	**0.287**

Table 5. MSE of $\hat{\lambda}$ in which $(k, n) = (15, 30)$

				Method					
κ	δ	ρ_y	ρ_x	1	2	3	4	5	6
0	0	0.2	0.2	1.311	0.660	1.046	0.716	*0.014*	**0.002**
			0.95	1.514	0.762	1.364	0.940	*0.011*	**0.001**
		0.95	0.2	1.217	0.594	0.922	0.540	*0.012*	**0.004**
			0.95	1.462	0.726	1.326	0.861	*0.013*	**0.003**
5	1	0.2	0.2	0.285	*0.042*	0.163	**0.036**	0.242	0.328
			0.95	0.864	0.328	0.717	0.356	**0.080**	*0.096*
		0.95	0.2	0.030	*0.024*	**0.013**	0.048	0.207	0.471
			0.95	0.316	**0.044**	0.201	*0.049*	0.338	0.412
	3	0.2	0.2	*0.012*	0.047	**0.005**	0.069	0.181	0.473
			0.95	0.232	**0.017**	0.141	*0.048*	0.427	0.522
		0.95	0.2	0.004	0.063	*0.003*	0.068	**0.001**	0.111
			0.95	0.139	**0.002**	0.043	*0.018*	0.224	0.390
10	1	0.2	0.2	0.099	**0.001**	0.038	*0.020*	0.336	0.506
			0.95	0.475	*0.111*	0.340	**0.099**	0.253	0.296
		0.95	0.2	0.044	**0.014**	*0.018*	0.033	0.097	0.301
			0.95	0.315	*0.044*	0.169	**0.038**	0.315	0.398
	3	0.2	0.2	1.308	0.667	1.139	*0.562*	0.621	**0.249**
			0.95	1.542	0.787	1.145	0.547	*0.057*	**0.013**
		0.95	0.2	0.002	0.073	**0.001**	0.084	*0.002*	0.118
			0.95	0.114	**0.002**	0.026	*0.024*	0.210	0.376
15	1	0.2	0.2	0.071	**0.005**	0.025	*0.023*	0.260	0.462
			0.95	0.468	*0.109*	0.307	**0.074**	0.244	0.292
		0.95	0.2	0.033	*0.020*	**0.013**	0.040	0.074	0.267
			0.95	1.540	0.786	1.142	0.540	*0.050*	**0.011**
	3	0.2	0.2	1.259	0.642	1.146	*0.571*	0.841	**0.374**
			0.95	1.527	0.779	1.121	0.556	*0.104*	**0.029**
		0.95	0.2	0.006	0.056	*0.004*	0.063	**0.001**	0.083
			0.95	0.121	**0.002**	0.033	*0.018*	0.149	0.305
Average				0.583	0.264	0.450	0.251	**0.192**	*0.246*

Table 6. MSE of $\hat{\lambda}$ in which $(k, n) = (15, 50)$

				Method					
κ	δ	ρ_y	ρ_x	1	2	3	4	5	6
0	0	0.2	0.2	1.166	0.903	0.867	0.753	*0.006*	**0.002**
			0.95	1.256	0.973	1.247	1.059	*0.004*	**0.001**
		0.95	0.2	1.167	0.904	0.816	0.701	*0.007*	**0.003**
			0.95	1.254	0.972	1.163	0.997	*0.004*	**0.002**
5	1	0.2	0.2	*0.023*	**0.001**	*0.023*	0.025	0.520	0.625
			0.95	0.289	*0.163*	0.210	**0.141**	0.318	0.340
		0.95	0.2	*0.002*	0.007	**0.001**	0.009	0.073	0.154
			0.95	0.044	**0.006**	*0.010*	**0.006**	0.555	0.630
	3	0.2	0.2	**0.000**	*0.015*	**0.000**	0.017	0.053	0.126
			0.95	0.035	**0.003**	0.011	*0.007*	0.557	0.645
		0.95	0.2	**0.006**	0.041	**0.006**	0.042	*0.013*	0.059
			0.95	*0.004*	0.005	**0.000**	0.015	0.117	0.205
10	1	0.2	0.2	0.012	**0.000**	*0.004*	0.005	0.278	0.392
			0.95	0.174	0.080	*0.078*	**0.036**	0.441	0.472
		0.95	0.2	*0.004*	*0.004*	**0.003**	0.006	0.034	0.095
			0.95	0.063	*0.014*	0.021	**0.005**	0.452	0.532
	3	0.2	0.2	1.171	0.912	1.158	*0.900*	0.946	**0.714**
			0.95	1.261	0.982	1.069	0.817	*0.126*	0**0.073**
		0.95	0.2	**0.003**	0.034	**0.003**	0.034	*0.008*	0.046
			0.95	*0.004*	0.005	**0.000**	0.014	0.088	0.168
15	1	0.2	0.2	0.006	*0.003*	**0.001**	0.011	0.232	0.345
			0.95	0.142	*0.059*	0.060	**0.020**	0.449	0.494
		0.95	0.2	*0.002*	0.007	**0.001**	0.009	0.030	0.088
			0.95	0.053	*0.010*	0.015	**0.002**	0.442	0.523
	3	0.2	0.2	1.197	0.932	1.184	*0.920*	1.002	**0.761**
			0.95	1.270	0.989	1.056	0.812	*0.192*	**0.122**
		0.95	0.2	0.007	**0.002**	0.007	**0.002**	*0.003*	0.005
			0.95	0.016	**0.000**	0.006	*0.003*	0.051	0.115
Average				0.380	0.287	0.322	*0.263*	**0.250**	0.276

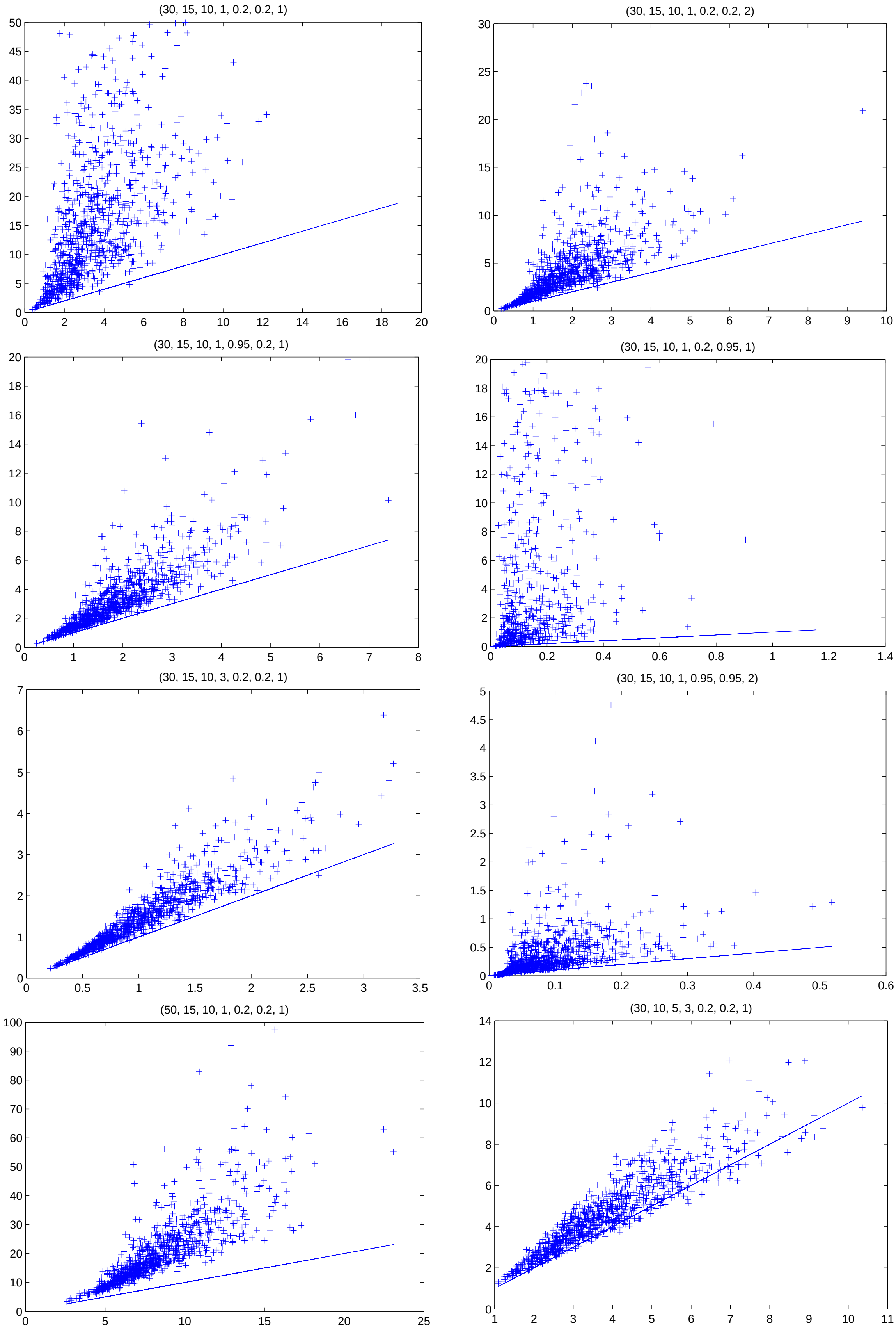

Figure 2. The scatter plots for $\tilde{\theta}_{(1)}(\lambda)$ and $\hat{\theta}(\lambda)$ in the case of $(n, k, \kappa, \delta, \rho_y, \rho_x, \lambda)$

Table 7. Simulation results for the case in which $(k, n) = (5, 30)$

κ	δ	ρ_y	ρ_x	Method 1	2	3	4	5	6	7	8
0	0	0.2	0.2	88.34	87.53	88.00	87.16	87.30	**86.76**	87.44	*86.88*
			0.95	90.24	89.05	90.52	89.32	88.34	**87.46**	88.63	*87.73*
		0.95	0.2	88.49	87.64	88.15	87.28	87.39	**86.82**	87.53	*86.94*
			0.95	90.25	89.03	90.37	89.15	88.28	**87.41**	88.56	*87.67*
5	1	0.2	0.2	*94.80*	**94.72**	94.81	94.90	94.87	95.10	**94.72**	94.91
			0.95	92.17	91.54	92.09	91.48	91.25	**90.96**	91.34	*91.02*
		0.95	0.2	97.34	*97.31*	97.34	97.36	97.37	97.46	**97.29**	97.41
			0.95	93.47	92.94	93.47	93.00	92.74	**92.43**	92.76	*92.45*
	3	0.2	0.2	*98.92*	98.93	*98.92*	98.94	*98.92*	98.95	**98.91**	98.98
			0.95	95.44	95.23	95.69	95.75	95.40	95.43	**95.19**	*95.21*
		0.95	0.2	**99.37**	**99.37**	**99.37**	*99.38*	**99.37**	*99.38*	**99.37**	99.40
			0.95	*97.61*	**97.54**	97.66	97.82	97.72	97.86	**97.54**	97.67
Average				93.87	93.40	93.87	93.46	93.25	**93.00**	93.27	*93.02*

Table 8. Simulation results for the case in which $(k, n) = (5, 50)$

κ	δ	ρ_y	ρ_x	Method 1	2	3	4	5	6	7	8
0	0	0.2	0.2	92.22	91.94	92.06	91.70	91.63	**91.45**	91.73	*91.54*
			0.95	93.29	92.91	93.52	93.08	92.10	**91.80**	92.27	*91.98*
		0.95	0.2	92.22	91.96	92.08	91.72	91.67	**91.49**	91.76	*91.58*
			0.95	93.24	92.83	93.42	93.01	92.05	**91.78**	92.21	*91.93*
5	1	0.2	0.2	*97.46*	**97.45**	*97.46*	97.48	97.52	97.57	*97.46*	97.52
			0.95	94.96	94.79	95.00	94.82	94.54	**94.46**	94.57	*94.48*
		0.95	0.2	**98.75**	*98.76*	*98.76*	98.77	98.77	98.79	98.77	98.81
			0.95	96.19	96.06	96.28	96.20	95.96	95.92	*95.91*	**95.87**
	3	0.2	0.2	**99.57**	**99.57**	**99.57**	**99.57**	**99.57**	**99.57**	**99.57**	*99.58*
			0.95	97.70	*97.67*	97.80	97.84	97.78	97.83	**97.66**	97.70
		0.95	0.2	**99.79**	*99.80*	**99.79**	*99.80*	**99.79**	*99.80*	*99.80*	99.81
			0.95	**98.73**	**98.73**	*98.74*	98.75	98.77	98.80	*98.74*	98.78
Average				96.18	96.04	96.21	96.06	95.84	**95.77**	95.87	*95.80*

Table 9. Simulation results for the case in which $(k, n) = (10, 30)$

				Method							
κ	δ	ρ_y	ρ_x	1	2	3	4	5	6	7	8
0	0	0.2	0.2	79.07	77.26	78.10	76.70	77.69	**76.47**	77.86	*76.56*
			0.95	81.19	78.61	80.54	78.24	78.86	**77.14**	79.13	*77.28*
		0.95	0.2	78.78	77.07	77.90	76.56	77.48	**76.33**	77.64	*76.41*
			0.95	81.01	78.44	80.29	77.72	78.72	**76.89**	79.03	*77.08*
5	1	0.2	0.2	87.14	**86.60**	87.12	87.07	86.75	86.88	*86.72*	86.76
			0.95	83.49	81.63	82.99	81.41	81.83	**80.74**	82.01	*80.86*
		0.95	0.2	91.60	**91.36**	91.45	91.52	91.46	91.66	*91.38*	91.56
			0.95	85.21	83.47	84.66	83.12	83.70	**82.59**	83.87	*82.68*
	3	0.2	0.2	95.74	*95.71*	*95.71*	95.80	95.72	95.85	**95.67**	95.83
			0.95	88.41	87.52	88.34	87.83	87.67	**87.21**	87.73	*87.23*
		0.95	0.2	*97.50*	97.51	*97.50*	97.56	*97.50*	97.56	**97.47**	97.58
			0.95	91.85	*91.67*	91.91	92.63	91.77	92.18	**91.66**	92.02
10	1	0.2	0.2	92.02	**91.74**	91.93	91.93	91.85	92.12	*91.76*	91.98
			0.95	84.62	82.96	84.27	82.91	83.14	**82.18**	83.29	*82.28*
		0.95	0.2	94.53	*94.37*	94.50	94.57	94.47	94.68	**94.36**	94.57
			0.95	86.19	84.74	85.81	84.66	84.91	**84.02**	85.03	*84.10*
	3	0.2	0.2	**98.57**	98.64	*98.59*	98.71	*98.59*	98.72	**98.57**	98.77
			0.95	91.50	**90.99**	91.59	91.67	91.12	91.23	*91.06*	91.12
		0.95	0.2	**99.69**	99.73	*99.70*	99.76	*99.70*	99.76	*99.70*	99.81
			0.95	94.15	**93.83**	94.20	94.37	94.00	94.19	*93.87*	94.02
Average				89.11	88.19	88.85	88.24	88.35	**87.92**	88.39	*87.93*

Table 10. Simulation results for the case in which $(k, n) = (10, 50)$

κ	δ	ρ_y	ρ_x	Method							
				1	2	3	4	5	6	7	8
0	0	0.2	0.2	84.93	84.48	84.60	**84.09**	84.41	**84.09**	84.51	*84.17*
			0.95	86.22	85.43	85.59	84.96	84.99	**84.48**	85.19	*84.63*
		0.95	0.2	84.98	84.50	84.62	*84.13*	84.44	**84.10**	84.54	84.18
			0.95	86.20	85.42	85.58	85.00	85.04	**84.60**	85.20	*84.68*
5	1	0.2	0.2	*92.65*	**92.60**	92.67	92.72	92.66	92.74	**92.60**	92.66
			0.95	88.43	87.94	88.14	87.77	87.70	**87.40**	87.82	*87.48*
		0.95	0.2	*95.43*	*95.43*	95.44	95.47	95.45	95.51	**95.42**	95.50
			0.95	90.03	89.69	89.94	89.68	89.55	**89.31**	89.61	*89.37*
	3	0.2	0.2	**97.90**	*97.92*	**97.90**	97.93	**97.90**	97.93	**97.90**	97.95
			0.95	93.17	**93.09**	93.36	93.52	93.15	93.22	**93.09**	*93.14*
		0.95	0.2	**99.27**	*99.28*	**99.27**	*99.28*	**99.27**	*99.28*	**99.27**	99.30
			0.95	95.51	**95.47**	95.49	95.53	95.54	95.61	*95.48*	95.54
10	1	0.2	0.2	*96.50*	**96.47**	96.53	96.55	96.52	96.57	**96.47**	96.53
			0.95	89.71	89.33	89.58	89.32	89.16	**88.94**	89.23	*89.00*
		0.95	0.2	**97.64**	**97.64**	*97.66*	97.68	*97.66*	97.70	**97.64**	97.70
			0.95	91.07	90.75	91.09	90.91	90.63	**90.48**	90.67	*90.51*
	3	0.2	0.2	**99.44**	**99.44**	**99.44**	**99.44**	**99.44**	**99.44**	**99.44**	*99.45*
			0.95	95.66	**95.64**	95.85	95.97	95.75	95.89	*95.65*	95.77
		0.95	0.2	**99.46**	**99.46**	**99.46**	*99.47*	**99.46**	*99.47*	**99.46**	*99.47*
			0.95	97.31	**97.29**	97.35	97.38	97.36	97.43	*97.30*	97.37
Average				93.08	92.86	92.98	92.84	92.80	**92.71**	92.83	*92.72*

Table 11. Simulation results for the case in which $(k, n) = (15, 30)$

κ	δ	ρ_y	ρ_x	Method							
				1	2	3	4	5	6	7	8
0	0	0.2	0.2	72.83	69.39	71.66	68.77	70.91	**68.51**	71.05	*68.56*
			0.95	76.39	71.49	74.39	70.03	72.67	**69.23**	72.91	*69.43*
		0.95	0.2	73.86	69.79	72.30	69.03	71.62	**68.75**	71.78	*68.81*
			0.95	76.96	71.53	74.71	70.21	72.86	**69.12**	73.14	*69.26*
5	1	0.2	0.2	81.70	79.12	81.18	79.31	80.22	**78.70**	80.33	*78.73*
			0.95	78.31	73.62	76.23	72.84	74.79	**71.99**	75.01	*72.06*
		0.95	0.2	93.65	93.73	93.78	94.48	*93.61*	94.47	**93.50**	94.31
			0.95	83.27	80.58	82.40	80.94	81.18	**80.12**	81.29	*80.16*
	3	0.2	0.2	*95.37*	95.79	95.46	96.48	95.45	96.66	**95.33**	96.48
			0.95	83.89	81.25	83.09	81.71	81.83	**80.83**	81.96	*80.84*
		0.95	0.2	**99.19**	99.30	**99.19**	99.32	*99.20*	99.34	**99.19**	99.40
			0.95	91.36	**90.77**	91.08	91.36	90.87	91.24	*90.84*	91.15
10	1	0.2	0.2	86.74	**85.67**	86.32	86.26	*86.01*	86.17	86.02	86.09
			0.95	80.23	76.29	78.81	75.65	77.28	**74.85**	77.44	*74.89*
		0.95	0.2	94.47	94.45	94.46	94.87	*94.43*	94.95	**94.33**	94.84
			0.95	84.20	81.78	83.37	82.11	82.29	**81.24**	82.40	*81.27*
	3	0.2	0.2	97.61	*97.51*	97.59	97.73	97.57	97.76	**97.48**	97.72
			0.95	86.41	*84.68*	85.85	85.24	85.06	**84.55**	85.13	**84.55**
		0.95	0.2	**99.24**	99.32	*99.25*	99.35	*99.25*	99.35	**99.24**	99.40
			0.95	92.31	**91.76**	91.97	92.10	91.90	92.16	*91.86*	92.08
15	1	0.2	0.2	89.95	**89.49**	89.75	90.12	89.59	90.23	*89.54*	90.11
			0.95	81.17	77.65	79.74	77.00	78.54	**76.39**	78.69	*76.45*
		0.95	0.2	95.13	*95.08*	95.11	95.38	95.09	95.47	**95.00**	95.39
			0.95	84.74	82.63	84.10	83.32	83.04	**82.47**	83.13	*82.49*
	3	0.2	0.2	*97.99*	98.09	98.00	98.23	*97.99*	98.25	**97.95**	98.28
			0.95	88.99	**87.89**	88.33	88.61	*88.07*	88.28	88.09	88.23
		0.95	0.2	*99.16*	99.21	*99.16*	99.23	*99.16*	99.23	**99.15**	99.27
			0.95	92.82	**92.25**	92.46	92.53	92.40	92.56	*92.36*	92.48
Average				87.78	86.08	87.13	86.15	86.53	*85.82*	86.58	**85.81**

Table 12. Simulation results for the case in which $(k, n) = (15, 50)$

κ	δ	ρ_y	ρ_x	Method 1	2	3	4	5	6	7	8
0	0	0.2	0.2	79.00	78.27	78.34	**77.85**	78.34	*77.87*	78.42	77.93
			0.95	80.06	78.92	78.88	78.24	78.74	**78.07**	78.90	*78.17*
		0.95	0.2	78.96	78.24	78.35	**77.78**	78.31	*77.80*	78.39	77.86
			0.95	80.47	79.24	79.11	*78.33*	79.01	**78.24**	79.18	78.34
5	1	0.2	0.2	89.06	**88.80**	89.42	89.64	88.86	88.89	88.84	*88.83*
			0.95	82.51	81.71	81.91	81.31	81.60	**81.14**	81.70	*81.20*
		0.95	0.2	**97.59**	97.65	*97.60*	97.69	97.63	97.76	97.61	97.77
			0.95	88.51	88.15	88.41	88.35	88.14	*88.05*	88.16	**88.04**
	3	0.2	0.2	**98.10**	98.12	**98.10**	98.13	*98.11*	98.16	**98.10**	98.18
			0.95	89.07	**88.81**	89.06	89.18	**88.81**	88.87	*88.82*	88.84
		0.95	0.2	**99.76**	*99.77*	**99.76**	*99.77*	**99.76**	*99.77*	**99.76**	*99.77*
			0.95	95.77	*95.74*	95.77	95.81	95.77	95.83	**95.73**	95.81
10	1	0.2	0.2	93.46	*93.41*	93.49	93.55	93.46	93.60	**93.40**	93.53
			0.95	84.38	83.67	83.97	83.45	83.59	**83.15**	83.68	*83.22*
		0.95	0.2	**98.01**	98.03	*98.02*	98.05	*98.02*	98.07	**98.01**	98.08
			0.95	89.21	88.93	89.27	89.22	88.92	*88.88*	88.93	**88.87**
	3	0.2	0.2	**98.84**	98.86	**98.84**	98.86	*98.85*	98.87	*98.85*	98.89
			0.95	92.05	**91.85**	92.06	92.11	91.89	91.92	*91.86*	91.87
		0.95	0.2	**99.68**	*99.69*	**99.68**	*99.69*	**99.68**	*99.69*	*99.69*	99.70
			0.95	*96.56*	*96.56*	96.57	96.62	96.58	96.65	**96.55**	96.64
15	1	0.2	0.2	*95.10*	**95.05**	95.13	95.17	*95.10*	95.19	**95.05**	95.14
			0.95	85.64	85.11	85.46	85.14	85.04	**84.76**	85.11	*84.80*
		0.95	0.2	**98.37**	98.39	*98.38*	98.41	98.39	98.43	*98.38*	98.44
			0.95	89.86	*89.58*	89.72	89.75	*89.58*	*89.58*	89.59	**89.56**
	3	0.2	0.2	*99.46*	*99.46*	*99.46*	*99.46*	*99.46*	*99.46*	**99.45**	99.48
			0.95	93.82	**93.65**	93.77	93.73	93.69	93.71	**93.65**	*93.66*
		0.95	0.2	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
			0.95	96.76	*96.75*	96.78	96.81	96.77	96.82	**96.74**	96.80
Average				91.79	91.52	91.62	91.50	91.50	**91.40**	91.52	*91.41*