Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 11(3): 6866–6909.
DOI: 10.3934/math.2026283
Received: 14 December 2025
Revised: 19 February 2026
Accepted: 27 February 2026
Published: 17 March 2026

Research article

Nonasymptotic oracle inequalities and alternating direction method of multipliers algorithm for adaptive lasso penalized multiplicative regression

Mingzhen Wan[1] **and Wei Chen**[2,*]

[1] Suzhou Institute of Technology, Jiangsu University of Science and Technology, Zhangjiagang 215600, China

[2] School of Zhangjiagang, Jiangsu University of Science and Technology, Zhangjiagang 215600, China

* **Correspondence:** Email: chenweixiyang@sina.com; Tel: +8615365358156.

Abstract: In this paper, we mainly consider the regularized estimation problem of parameters in the multiplicative regression model, where the response variable is always positive. Hao, Lin, and Zhao, *Comput. Stat. Data An.*, **103** (2016) investigated an adaptive variable selection method via the least product relative error (LPRE) criterion and lasso-type penalty with fixed or diverging number of covariates and showed the resultant estimator achieves the oracle property. However, the alternating direction method of multipliers (ADMM) algorithm proposed by the authors is based on the least square approximation of the LPRE loss function, where a well-behaved initial estimator must be determined in advance, and the convergence is not validated. Through careful introduction of auxiliary variables and a three-block reformulation, our ADMM algorithm eliminates sensitivity to initial values while ensuring convergence. In addition, by virtue of the symmetric Bregman (SB) divergence and natural extensions of compatibility and weak cone invertibility factors, we establish nonasymptotic oracle inequalities for the ℓ_1 estimation error and prediction error measured by the SB divergence of the lasso penalized LPRE estimator. The proposed method is shown to be very efficient owing to the fact that almost each derived subproblem has a closed-form solution. Extensive simulation studies are conducted to evaluate the finite-sample performance of the proposal. Finally, a real data set is analyzed to illustrate the practical utility of our proposed method.

Keywords: adaptive Lasso; alternating direction method of multipliers; convergence analysis; least product relative error; multiplicative regression model; oracle inequalities; variable selection
Mathematics Subject Classification: 62F07, 62G07

1. Introduction

In many practical applications, the response variables only take positive values, such as the price of a financial asset or survival time. For modeling the relationship between the positive response and a set of explanatory variables, one can take a logarithmic transformation for the response, then the popular linear mean regression or quantile regression could be employed for the transformed data. Due to the fact that a linear relationship in the transformed model doesn't hold in the original one, the analysis results based on the transformed data had to be transformed back to the original measurement scale. Moreover, this may result in inconsistency when one is solely interested in the conditional expectation estimation of the response given the explanatory variables, rather than the regression parameter itself. It should be noted that the least square or the least absolute deviation criterion are both based on the absolute error. However, in some fields, such as the analysis of heteroscedastic data, which are particularly common in economic, reliability control, survival analysis, epidemiological, and other social studies, the relative error is more attractive and of more concern. Some related examples can be found in [3].

The relative error occurring in the literature has two types: one is the ratio of the absolute error relative to the target value, the other is relative to the predicted value. Historically, the first type is widely used to construct the loss function in the early literature and particularly preferred by some researchers; see [12, 16, 17]. As argued in [3], making statistical inference solely based on the first kind of relative error can be quite inadequate when the unknown target value is large and the predictor is relatively small. A descriptive example is provided in remark 1 of [3]. Therefore, an appealing and reasonable loss function should be one which takes into account both types of relative errors. Inspired by this idea, [3] suggested the following multiplicative regression (MR) model,

$$\tilde{Y} = \exp(X^{\top}\beta)\epsilon, \tag{1.1}$$

where $\tilde{Y}$ and ϵ denote the positive response variable and random error, respectively, and X is the p-dimensional vector of covariates associated with the unknown regression parameter vector β. Besides, [3] proposed the least absolute relative error (LARE) criterion for model (1.1), which is defined as the sum of these two types of relative errors, and enjoys the scale-free property.

However, the LARE loss function is nonsmooth and the asymptotic covariance matrix involves unknown error density, which would bring some numerical difficulties in computing the estimator of regression coefficient and greatly limit the use of this approach. To overcome these drawbacks, [4] introduced a general class of criteria, among which the least product relative error (LPRE) criterion gets special emphasis, to estimate the regression parameter. The LPRE loss function is composed of the product of both relative errors. What's more, this function is infinitely differentiable and strictly convex in the regression coefficient, and is still symmetric in the target and its predictor. All these merits make the LPRE criterion particularly attractive both practically and theoretically. Since then, many works have contributed to statistical inferences for model (1.1) and its various extensions; see [5, 10, 26–28].

It is worth to mention that most of the papers above are only concerned about the parameter estimation issue, where the number of covariates is very small; namely, the covariates are low-dimensional. When the dimensionality of explanatory variables X in model (1.1) is high, even larger than the sample size, direct applications of those methods usually result in non-ignorable bias, bad prediction, and even misleading conclusions. For high dimensional data, a common assumption is that the regression coefficient is sparse. It means that only a few explanatory variables have an influence on the response. Namely, the true regression coefficient has only a few nonzero coordinates, with the rest being zero. However, which covariates are not important is unknown in advance. To ensure both accurate estimation and accurate prediction, variable selection or model selection seems to be necessary and particularly valuable.

In the literature, a large body of research regarding various regularization methods have been developed and mainly built on absolute error criteria. To the best of our knowledge, there exist only a few efforts devoted to the variable selection for MR models based on the relative error criteria. Explicitly speaking, [9] proposed a variable selection method by combining the LPRE criterion and the adaptive lasso penalty and established the oracle property of the proposed estimator with fixed or diverging number of covariates in the model (1.1). [5] extended model (1.1) to a partially linear multiplicative regression model and proposed a class of parameter estimation and variable selection approaches using a locally weighted LARE loss function. [23] considered an adaptive weighting scheme-based regularized estimation for model (1.1) via the LARE criterion, where the dimension of model is allowed to increase with the sample size. [14] developed an adaptive group variable selection method based on the LARE and LPRE criterion. Recently, [6] investigated the nonconvex penalized LPRE variable selection method for model (1.1) with ultra-high dimensional data. Although these authors proved that their estimators enjoy some desirable large-sample properties, such as consistent parameter estimation and variable selection, as well as asymptotic normality, deriving finite-sample upper bounds on the estimation and prediction error is still desirable and worthwhile; for more details, see [11, 19, 21, 26]. It is well known that oracle inequalities are rendered to be an important and useful theoretical instrument to asses the statistical performance of an estimator. However, to the best of our knowledge, there is no study on the oracle inequalities in the context of relative error criterion. In this paper, we will fulfill this gap. By virtue of the symmetric Bregman divergence and natural extensions of compatibility and weak cone invertibility factors, we establish nonasymptotic oracle inequalities for the ℓ_1 estimation error and prediction error, measured by the SB divergence of the lasso-penalized LPRE estimator.

Although the multiplicative regression model is very attractive and the existing variable selection methods have been proved to enjoy many desirable theorematic properties, the practical usage of these methods doesn't match the expectation. A major reason is that there is no efficient, fast, and convergence-guaranteed algorithm to solve the minimization problem therein. Recall that the LARE loss function is nonsmooth; direct optimization can be conducted by some non-smooth algorithms, which don't work quickly and can be prohibitive, when the number of variables is large. To balance the computational cost and estimation accuracy, [23] suggested borrowing the least squares approximation in [22] into the LARE loss function. Then some efficient algorithms, such as the LARS by [7] could be implemented. Surprisingly, when minimizing the lasso-penalized LPRE-based objective function, [9] still resorted to the quadratic approximation to replace the original LPRE loss function part. Their simulation studies demonstrated that these strategies seem to be feasible, but the number of covariates used was very small, namely, $p = 8, 12, 16$. A similar phenomenon still happens in [14, 23]. Note that [9] claimed that they applied the alternating direction method of multipliers (ADMM) algorithm to solve the least-square-type minimization problem. It is well known that the ADMM approach is very flexible and versatile in dealing with many large-scale statistical estimation problems, where the objective function is written as the sum of two convex functions and probably subject to linear constraints, see [1, 2, 13, 15].

When p is ultra-high dimensional, [6] extended [9]'s ADMM approach and conducted a small-scale simulation study, where p was set to be 200 or 400, and 200 replications were done. All these settings are greatly different from those in the ADMM literature. One possible reason is that this algorithm is rather time-consuming. After checking all of the iteration steps therein, one finds that the convergence property of their algorithm is not verified, and that the ADMM approach is implemented for the approximated loss function, where a good initial value is very critical to acquire good performance. Choosing such an initial value is usually not hard work for low-dimensional data, but may be expensive and troublesome for high-dimensional data. This further complicates the choice of the tuning parameter, where the demand of computation becomes large. Is it necessary and indispensable to make a quadratic approximation before applying the ADMM approach? If it is not so, how to develop a more efficient ADMM algorithm in the true sense becomes a urgent issue. Once such an algorithm is obtained, finding a special initial value is avoided, because the convergence usually doesn't depend on the initial value. Luckily, through introducing some auxiliary variables, a three-block ADMM algorithm is developed for optimizing the lasso-type penalized least product relative error objective function, and its convergence is established. What's more, this work is conducive to the broad application of the existing statistical inference methods for model (1.1).

The main contributions of this paper are summarized as follows. First, we first investigate the oracle inequalities in the context of relative error criterion and establish nonasymptotic oracle inequalities for the ℓ_1 estimation error and prediction error measured by the SB divergence of the lasso-penalized LPRE estimator, which is very valuable for choosing a proper tuning parameter. Second, a generic ADMM algorithm is developed to solve the minimization problem without making a quadratic approximation for the LPRE loss, where a lasso-type penalty regularized LPRE estimator is obtained, and the main subproblems at each iteration either have closed-form solutions or can be efficiently solved by some simple algorithms. Third, the convergence of the proposed algorithm is validated. Fourth, there is no need to provide a well-defined initial estimator of the true regression coefficients in our algorithm. Fifth, a Bayesian information criterion-type criterions, which are data-driven, are proposed to select the tuning parameter.

The remainder of this paper is structured as follows. We first introduce the penalized LPRE estimator via a general class of adaptive penalty functions in Section 2 with a focus on the adaptive lasso penalty. In Section 3, we establish nonasymptotic oracle inequalities for the ℓ_1 estimation error and prediction error measured by the SB divergence of the lasso-penalized LPRE estimator. In Section 4 we develop a generic ADMM algorithm for solving the lasso-regularized LPRE objective function in detail, and establish the convergence of the proposed algorithm. Besides, some data-driven approaches to tuning parameter selection are provided. Extensive simulation studies are conducted to evaluate the proposed method and compare with several alternatives in Section 5. To illustrate the practical usage of our method, one real example is analyzed in Section 6. Finally, some promising discussions in Section 7 conclude the paper.

2. Models and estimation methods

2.1. Models, data, and LPRE estimation

Let $(\tilde{Y}_i, X_i)$, $i = 1, ..., n$ be independent and identical replicates of $(\tilde{Y}, X)$ that satisfies the MR model (1.1). In other words, the data comes from the following model:

$$\tilde{Y}_i = \exp(X_i^\top \beta)\epsilon_i.$$

Following [4], the LPRE estimator is defined as the minimizer of

$$L_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left\{\left|\frac{\tilde{Y}_i - \exp(X_i^\top\beta)}{\tilde{Y}_i}\right| \times \left|\frac{\tilde{Y}_i - \exp(X_i^\top\beta)}{\exp(X_i\beta)}\right|\right\} \tag{2.1}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[\tilde{Y}_i \exp(-X_i^\top\beta) + \tilde{Y}_i^{-1}\exp(X_i^\top\beta) - 2\right]. \tag{2.2}$$

Without loss of generality, we still denote the expression (2.2) without including the constant term-2 by $L_n(\beta)$. The convexity and smoothness of $L_n(\beta)$ facilitate the computational simplicity and guarantee solution uniqueness.

2.2. Regularized LPRE estimation

When the number of variables, p, is large and the regression coefficient is sparse, variable selection is desirable for gaining satisfactory estimation and prediction performance. Let $Pen(.)$ denote a nonnegative penalty function defined on the interval $[0, \infty)$. The regularized LPRE estimator can be obtained by minimizing the following objective function:

$$Q_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left[\tilde{Y}_i \exp(-X_i^\top\beta) + \tilde{Y}_i^{-1}\exp(X_i^\top\beta)\right] + \lambda\sum_{j=1}^{p} Pen(|\beta_j|), \tag{2.3}$$

where $\lambda > 0$ denotes the tuning parameter. In the literature, several different choices of $Pen(.)$ are investigated. Explicitly speaking, [9] considered the adaptive lasso penalty, where $Pen(|\beta_j|) = w_j|\beta_j|$, and $w_j, j = 1, ..., p$ are some pre-specified positive numbers. [6] considered nonconvex penalty function like the smoothly clipped absolute deviation (SCAD) and minimax concave penalty (MCP) penalties. To employ the standard convex optimization methods, the locally linear approximation of the penalty function is adopted in the subsequent numerical computation. All algorithms used in the above papers had to find a good initial estimator in advance. What's worse, whether these algorithms are convergent is still unknown. To some extent, their numerical studies show that the iteration process is potentially slow, especially in high-dimensional contexts and when the initial value is not good enough. In turn, when the data-driven methods are adopted to choose a proper tuning parameter, the entire computational burden is overwhelming.

On the other hand, in [6], the SCAD- and MCP-penalized LPRE estimator are finally computed by minimizing an adaptive lasso-penalized LPRE objective function, where the weights are determined by the first derivative of the penalty function at the true value; see page 3717 therein. If one rewrites the regression coefficient and the corresponding covariate vector, the adaptive lasso estimator can be transformed back to a standard lasso with weights in adaptive lasso method being unity. For this purpose, we can rewrite $Q_n(\beta)$ in (2.3) by

$$\frac{1}{n}\sum_{i=1}^{n}\left[\tilde{Y}_i \exp(-(W^{-1}X_i)^\top(W\beta)) + \tilde{Y}_i^{-1}\exp((W^{-1}X_i)^\top(W\beta)\right] + \lambda\sum_{j=1}^{p}|w_j\beta_j|,$$

where $W = diag(w_1, ..., w_p)$. Then, we can treat $W\beta$ as the new parameter and $W^{-1}X_i$ as the i-th covariate vector. A standard lasso problem arises. In light of these findings, if an efficient algorithm for calculating lasso penalized LPRE estimator is developed, the rest of the estimators can also be obtained theoretically. Therefore, in the next sections, we mainly consider the following optimization problem:

$$\min_{\beta\in\mathbb{R}^p}\frac{1}{n}\sum_{i=1}^{n}\left[\tilde{Y}_i\exp(-X_i^\top\beta) + \tilde{Y}_i^{-1}\exp(X_i^\top\beta)\right] + \lambda\sum_{j=1}^{p}|\beta_j|. \tag{2.4}$$

2.3. The Karush-Kuhn-Tucker conditions

The objective function $L_n(\beta)$, i.e., the first part of $Q_n(\beta)$ in (2.3), is continuously differential and convex, which facilitates the numerical calculation and theoretical analysis of the minimizer. By the subdifferentiation technique in the convex optimization theory, the lasso-penalized LPRE estimator can be characterized by the Karush-Kuhn-Tucker (KKT) conditions. In other words, a vector $\hat{\beta} = (\hat{\beta}_1, ..., \hat{\beta}_p)^\top$ is a solution to (2.4) if and only if the following KKT conditions hold,

$$\begin{cases} \dot{l}_j(\hat{\beta}) = -\lambda sgn(\hat{\beta}_j), & if\ \hat{\beta}_j \neq 0; \\ |\dot{l}_j(\hat{\beta})| \leq \lambda, & if\ \hat{\beta}_j = 0. \end{cases} \tag{2.5}$$

where $sgn(.)$ is the ordinary sign function, $\dot{l}_j(\beta) = \partial L_n(\beta)/\partial\beta_j$ denotes the jth partial derivative of $L_n(\beta)$ on β_j, and $\dot{l}(\beta) = (\dot{l}_1(\beta), ..., \dot{l}_p(\beta))^\top$ denotes the gradient of $L_n(\beta)$. The necessity and sufficiency of (2.5) can be proved in a standard manner. In addition, the KKT conditions have similar structure TO those in the linear regression model, which enables us to employ similar techniques to derive the upper bounds of the estimation error and prediction error, although the loss function for the multiplicative regression model has a more complex structure.

3. Oracle inequalities

In this section, we derive the oracle inequalities for the estimation error of lasso in the multiplicative regression model. Recall that β^o is the true vector of regression coefficients. Let $\mathcal{T}$ and $\mathcal{T}^c$ be the set of index of β^o being nonzero and zero, respectively, i.e., $\mathcal{T} = \{j : \beta_j^o \neq 0\}$ and $\mathcal{T}^c = \{j : \beta_j^o = 0\}$. For any p-dimensional vector β and a set $C \subseteq \{1, 2, ..., p\}$, write the subvector of β of components in C by β_C.

Based on the KKT conditions (2.5), we first develop a basic inequality involving the SB divergence and ℓ_1 estimation error in support $\mathcal{T}$ of β^o and its complement. The SB divergence is defined as

$$D^s(\hat{\beta},\beta) = (\hat{\beta} - \beta)^\top(\dot{l}(\hat{\beta}) - \dot{l}(\beta)),$$

which can be regarded as symmetric, partial Kullback-Leibler distance between the gradient at $\hat{\beta}$ and β. Thus, $D^s(\hat{\beta},\beta)$ can be viewed as a reasonable measure of prediction performance. The following inequality, given in Lemma 3.1, provides lower and upper bounds for the SB divergence and serves as an important tool for establishing the desirable oracle inequalities.

Lemma 3.1. *Let $\hat{\beta}$ be the lasso estimator in (2.4) and $\Delta = \hat{\beta} - \beta^o$. Define $z^* = \|\dot{l}(\beta^o)\|_\infty$. Then we have*

$$(\lambda - z^*)\|\Delta_{\mathcal{T}^c}\|_1 \leq D^s(\hat{\beta}, \beta^o) + (\lambda - z^*)\|\Delta_{\mathcal{T}^c}\|_1 \leq (\lambda + z^*)\|\Delta_{\mathcal{T}}\|_1.$$

Proof. Note that $L_n(\beta)$ is a convex and differential function, so $L_n(\hat{\beta}) - L_n(\beta^o) - (\hat{\beta} - \beta^o)^\top \dot{l}(\beta^o) \geq 0$ and $L_n(\beta^o) - L_n(\hat{\beta}) - (\beta^o - \hat{\beta})^\top \dot{l}(\hat{\beta}) \geq 0$. It naturally yields that $D^s(\hat{\beta}, \beta^o) \geq 0$, so the first inequality holds. By definition of $D^s(\hat{\beta}, \beta^o)$, we get

$$D^s(\hat{\beta}, \beta^o) = \Delta^\top(\dot{l}(\hat{\beta}) - \dot{l}(\beta^o)) = \sum_{j \in \mathcal{T}} \Delta_j \dot{l}_j(\hat{\beta}) + \sum_{j \in \mathcal{T}^c} \Delta_j \dot{l}_j(\hat{\beta}) + \Delta^\top(-\dot{l}(\beta^o)).$$

For $j \in \mathcal{T}^c$, $\Delta_j = \hat{\beta}_j - \beta_j^o = \hat{\beta}_j$.

$$\begin{aligned}\sum_{j \in \mathcal{T}^c} \Delta_j \dot{l}_j(\hat{\beta}) \equiv \sum_{j \in \mathcal{T}^c} \hat{\beta}_j \dot{l}_j(\hat{\beta}) &= \sum_{j \in \mathcal{T}^c, \hat{\beta}_j \neq 0} \hat{\beta}_j(-\lambda sgn(\hat{\beta}_j)) + \sum_{j \in \mathcal{T}^c, \hat{\beta}_j = 0} 0 \cdot \dot{l}_j(\hat{\beta}) \\ &\equiv -\lambda \sum_{j \in \mathcal{T}^c} \hat{\beta}_j sgn(\hat{\beta}_j) \\ &= -\lambda \sum_{j \in \mathcal{T}^c} |\hat{\beta}_j| = -\lambda \|\Delta_{\mathcal{T}^c}\|_1,\end{aligned}$$

where the second identity is obtained by using the KKT condition. Besides, for $j \in \mathcal{T}$, by KKT condition, $|\dot{l}_j(\hat{\beta})| \leq \lambda$.

$$|\sum_{j \in \mathcal{T}} \Delta_j \dot{l}_j(\hat{\beta})| \leq \sum_{j \in \mathcal{T}} |\Delta_j||\dot{l}_j(\hat{\beta})| \leq \sum_{j \in \mathcal{T}} \lambda |\Delta_j| = \lambda \|\Delta_{\mathcal{T}}\|_1.$$

$$|\Delta^\top(-\dot{l}(\beta^o))| \leq \|\Delta\|_1 \|\dot{l}(\beta^o))\|_\infty = z^* \|\Delta\|_1 = z^*(\|\Delta_{\mathcal{T}}\|_1 + \|\Delta_{\mathcal{T}^c}\|_1).$$

In summary, we have

$$D^s(\hat{\beta}, \beta^o) \leq \lambda \|\Delta_{\mathcal{T}}\|_1 - \lambda \|\Delta_{\mathcal{T}^c}\|_1 + z^*(\|\Delta_{\mathcal{T}}\|_1 + \|\Delta_{\mathcal{T}^c}\|_1).$$

Thus, the second inequality follows after some elementary algebraic calculations. □

If $z^* \leq \frac{\xi-1}{\xi+1}\lambda$ for some $\xi > 1$, then $\lambda - z^* \geq \frac{2}{\xi+1}\lambda$ and $\lambda + z^* \leq \frac{2\xi}{\xi+1}\lambda$. Combined with Lemma 3.1, it holds that

$$\|\Delta_{\mathcal{T}^c}\|_1 \leq \xi \|\Delta_{\mathcal{T}}\|_1.$$

Furthermore, in the event $\mathfrak{A} = \{z^* \leq \frac{\xi-1}{\xi+1}\lambda\}$, the estimation error $\Delta \in \mathfrak{C}$, where $\mathfrak{C} = \mathfrak{C}(\xi, \mathcal{T}) = \{b \in \mathbb{R}^p : b_{\mathcal{T}^c}\|_1 \leq \xi \|b_{\mathcal{T}}\|_1\}$. Note that $\mathfrak{C}$ is a cone set and $\mathbb{P}(\mathfrak{A}) \leq \mathbb{P}(\mathfrak{C})$. Similar results also occurred in the context of other regression models; see [11, 26].

Recall that for the ordinary least square estimator, the Gram matrix $\Sigma_n = \frac{1}{n}\sum_{i=1}^n X_i X_i^\top$ is assumed to be positive definite. However, Σ_n would be singular when $p > n$. Therefore, in the sparse, high-dimensional linear model via lasso, [19] proposed the restricted eigenvalue (RE) condition under the sparse cone set $\mathfrak{C}$, i.e., for a given nonnegative definite matrix Σ,

$$RE(\xi, \mathcal{T}; \Sigma) = \inf_{0 \neq b \in \mathfrak{C}(\xi, \mathcal{T})} \frac{(b^\top \Sigma b)^{1/2}}{\|b\|_2} > 0. \tag{3.1}$$

Besides, [20, 24] introduced the compatibility factor (CF)

$$CF(\xi, \mathcal{T}; \Sigma) = \inf_{0\neq b\in\mathfrak{C}(\xi,\mathcal{T})} \frac{d_{\mathcal{T}}(b^\top \Sigma b)^{1/2}}{\|b_{\mathcal{T}}\|_1} > 0. \tag{3.2}$$

and the weak cone invertibility factor (CIF)

$$CIF_q(\xi, \mathcal{T}; \Sigma) = \inf_{0\neq b\in\mathfrak{C}(\xi,\mathcal{T})} \frac{d_{\mathcal{T}}^{1/q}(b^\top \Sigma b)}{\|b_{\mathcal{T}}\|_1 \|b\|_q} > 0 \tag{3.3}$$

for some $q > 1$. Choose $q = 2$; CF and CIF are closely related to RE, but usually yield somewhat sharper oracle inequalities than the RE due to the fact $b_{\mathcal{T}}\|_1 \leq d_{\mathcal{T}}^{1/2}\|b\|_2$.

Let $\Sigma = \ddot{l}(\beta^o) = \partial^2 L_n(\beta^o)/\partial\beta\partial\beta^\top$; then we write $RE(\xi, \mathcal{T}) = RE(\xi, \mathcal{T}; \ddot{l}(\beta^o))$, $CF(\xi, \mathcal{T}) = CF(\xi, \mathcal{T}; \ddot{l}(\beta^o))$, and $CIF(\xi, \mathcal{T}) = CIF(\xi, \mathcal{T}; \ddot{l}(\beta^o))$. As argued in [21], it is possible to have $CF(\xi, \mathcal{T}) \gg RE(\xi, \mathcal{T})$.

Based on these concepts, we then introduce the second lemma, which explicitly controls the SB divergence and Hessian matrix of the LPRE loss function $L_n(\beta)$ in a neighborhood of β^o. For any two symmetric matrices A and B, $A \leq B$ means that $B - A$ is nonnegative definite.

Lemma 3.2. *Let* $\eta_b = \max_{1\leq i\leq n}\{\|X_i\|_2\}\|b\|_2$. *Then, it holds that*

$$e^{-\eta_b} b^\top \ddot{l}(\beta^o) b \leq D^s(\beta^o + b, \beta^o) \leq e^{\eta_b} b^\top \ddot{l}(\beta^o) b.$$

Moreover, $e^{-2\eta_b}\ddot{l}(\beta^o) \leq \ddot{l}(\beta^o + b) \leq e^{2\eta_b}\ddot{l}(\beta^o)$.

Proof. By Taylor expansion of $L_n(\beta)$ at β^o and $\beta^o + b$, respectively, we obtain

$$L_n(\beta^o + b) - L_n(\beta^o) = \dot{l}(\beta^o)^\top b + \frac{1}{2} b^\top \ddot{l}(\tilde{\beta}) b,$$
$$L_n(\beta^o) - L_n(\beta^o + b) = \dot{l}(\beta^o + b)^\top(-b) + \frac{1}{2} b^\top \ddot{l}(\bar{\beta}) b.$$

Taking the sum for both sides yields

$$0 = -b^\top(\dot{l}(\beta^o + b) - \dot{l}(\beta^o)) + \frac{1}{2} b^\top(\ddot{l}(\tilde{\beta}) + \ddot{l}(\bar{\beta}))b,$$

i.e., $D^s(\beta^o + b, \beta^o) = b^\top(\dot{l}(\beta^o + b) - \dot{l}(\beta^o)) = \frac{1}{2} b^\top(\ddot{l}(\tilde{\beta}) + \ddot{l}(\bar{\beta}))b$, where $\tilde{\beta}$ and $\bar{\beta}$ are lying between β^o and $\beta^o + b$. Note that

$$\begin{aligned}\ddot{l}(\bar{\beta}) &= \frac{1}{n}\sum_{i=1}^n \left[\tilde{Y}_i \exp(-\bar{\beta}^\top X_i) + \tilde{Y}_i^{-1}\exp(\bar{\beta}^\top X_i)\right] X_i X_i^\top \\ &= \frac{1}{n}\sum_{i=1}^n \left[\epsilon_i \exp(-(\bar{\beta} - \beta^o)^\top X_i) + \epsilon_i^{-1}\exp((\bar{\beta} - \beta^o)^\top X_i)\right] X_i X_i^\top,\end{aligned}$$

and $|(\bar{\beta} - \beta^o)^\top X_i| \leq \|\bar{\beta} - \beta^o\|_2\|X_i\|_2 \leq \|b\|_2\|X_i\|_2 \leq \max_{1\leq i\leq n}\|X_i\|_2\|b\|_2 \equiv \eta_b > 0$. We have, for every $i = 1, ..., n$, $\exp(-\eta_b) \leq \exp(\pm(\bar{\beta} - \beta^o)^\top X_i) \leq \exp(\eta_b)$. Similarly, $\exp(-\eta_b) \leq \exp(\pm(\tilde{\beta} - \beta^o)^\top X_i) \leq \exp(\eta_b)$. Combining $\ddot{l}(\beta^o) = \frac{1}{n}\sum_{i=1}^n(\epsilon_i + \epsilon_i^{-1})X_i X_i^\top$,

$$\exp(-\eta_b)\ddot{l}(\beta^o) \leq \frac{1}{2}(\ddot{l}(\tilde{\beta}) + \ddot{l}(\bar{\beta})) \leq \exp(\eta_b)\ddot{l}(\beta^o),$$

Thus, the first inequality is proved. Based on this result, the last inequality can be derived naturally. □

Based on these lemmas and the CF and CIF factors mentionedabove, we can establish the main theorem as follows.

Theorem 3.1. *Let* $K = \max_{1\le i\le n}\{\|X_i\|_2\}$*,* $\tau = \frac{Kd_{\mathcal{T}}\lambda(\xi+1)}{2CF(\xi,\mathcal{T})^2}$*, and* $\hat{\beta}$ *be defined as in (2.4). Suppose the conditions* (3.2) *and* (3.3) *hold and* $\tau \le e^{-1}$*. Then in the event* $\{z^* \le \frac{\xi-1}{\xi+1}\lambda\}$*,*

$$D^s(\hat{\beta},\beta^o) \le \frac{4e^{\tilde{t}}\xi^2 d_{\mathcal{T}}\lambda^2}{2CF(\xi,\mathcal{T})^2(\xi+1)^2} \tag{3.4}$$

$$\|\hat{\beta}-\beta^o\|_1 \le \frac{e^{\tilde{t}}(\xi+1)d_{\mathcal{T}}\lambda}{2CF(\xi,\mathcal{T})^2} \tag{3.5}$$

$$\|\hat{\beta}-\beta^o\|_q \le \frac{2e^{\tilde{t}}\xi d_{\mathcal{T}}^{1/q}\lambda}{CIF(\xi,\mathcal{T})^2}, \tag{3.6}$$

where $\tilde{t}$ *is the smaller solution of* $\tau = te^{-t}$*.*

Proof. Denote $\Delta = \hat{\beta}-\beta^o$ and $b = \Delta/\|\Delta\|_1$, then $\|b\|_1 = 1$ and $\hat{\beta} = \beta^o + \|\Delta\|_1 b$. Since $L_n(\beta)$ is a convex function and differentiable, $b^\top(\dot{l}(\beta^o+tb)-\dot{l}(\beta^o))$ is an increasing function of t on $[0,\|\Delta\|_1]$. Under event $z^* \le \frac{\xi-1}{\xi+1}\lambda$ for some $\xi > 1$ and Lemma 3.1, it holds that

$$\begin{aligned}
b^\top(\dot{l}(\beta^o+tb)-\dot{l}(\beta^o)) &\le b^\top(\dot{l}(\beta^o+\|\Delta\|_1 b)-\dot{l}(\beta^o)) \\
&= \|\Delta\|_1^{-1}D^s(\beta^o+\|\Delta\|_1 b,\beta^o) \\
&\le \frac{2\lambda\xi}{\xi+1}\|b_{\mathcal{T}}\|_1 - \frac{2\lambda}{\xi+1}\|b_{\mathcal{T}^c}\|_1) \\
&= \frac{2\lambda\xi}{\xi+1}\|b_{\mathcal{T}}\|_1 - \frac{2\lambda}{\xi+1}\|b_{\mathcal{T}^c}\|_1) \\
&= \frac{2\lambda(\xi+1)-2\lambda}{\xi+1}\|b_{\mathcal{T}}\|_1 - \frac{2\lambda}{\xi+1}\|b_{\mathcal{T}^c}\|_1) \\
&= 2\lambda\|b_{\mathcal{T}}\|_1 - \frac{2\lambda}{\xi+1}.
\end{aligned}$$

Then,

$$b^\top(\dot{l}(\beta^o+tb)-\dot{l}(\beta^o)) \le \frac{\lambda(\xi+1)}{2}\|b_{\mathcal{T}}\|_1^2, \tag{3.7}$$

where the last inequality follows from the fact that for any two nonnegative numbers y_1, y_2, $2\sqrt{y_1y_2} - y_1 \le y_2$ by setting $y_1 = \frac{2\lambda}{\xi+1}$ and $y_2 = \frac{\lambda(\xi+1)}{2}\|b_{\mathcal{T}}\|_1^2$. Then by the definition of $D^s(.,.)$,

$$D^s(\beta^o+tb,\beta^o) \le \frac{t\lambda(\xi+1)}{2}\|b_{\mathcal{T}}\|_1^2 \tag{3.8}$$

Denote $K = \max_{1\le i\le n}\{\|X_i\|_2\}$, and by Lemma 3.2, it holds that

$$D^s(\beta^o+tb,\beta^o) \ge t^2e^{-\eta_t b}b^\top\ddot{l}(\beta^o)b = t^2e^{-Kt}b^\top\ddot{l}(\beta^o)b.$$

Furthermore, by condition (3.2), we have

$$D^s(\beta^o+tb,\beta^o) \ge t^2e^{-Kt}\frac{CF(\xi,\mathcal{T})^2\|b_{\mathcal{T}}\|_1^2}{d_{\mathcal{T}}}. \tag{3.9}$$

Combining (3.8) and (3.9), we have

$$Kte^{-Kt} \le \frac{Kd_{\mathcal{T}}\lambda(\xi+1)}{2CF(\xi,\mathcal{T})^2} = \tau.$$

We find that any t satisfying (3.8) must satisfy (3.9), which demonstrates that the set of all nonnegative numbers t satisfying (3.9) should be a closed interval $[0, t_0]$. In addition, under the condition $\tau \le e^{-1}$, and noting that $e^{-1} = \max_{t\ge0}\{te^{-t}\}$, it follows that $t_0 \le \tilde{t}$, where $\tilde{t}$ is the smaller solution of $\tau = te^{-t}$. It means that $Kt_0 \le \tilde{t}$. Note that $\|\Delta\|_1 \le t_0$. Therefore, $\|\Delta\|_1 \le K^{-1}\tilde{t} = K^{-1}e^{\tilde{t}}\tau$. As a result, the inequality (3.5) is proved. Using $e^{-Kt} \ge e^{-\tilde{t}}$, (3.8), Lemma 3.1, and the definition of $CF(\xi,\mathcal{T})$, we have

$$\frac{2\xi\lambda}{\xi+1}\|b_{\mathcal{T}}\|_1 \ge D^s(\hat{\beta},\beta^o) \ge e^{-\tilde{t}}\Delta^\top\ddot{l}(\beta^o)\Delta \ge \frac{e^{-\tilde{t}}CF(\xi,\mathcal{T})^2\|b_{\mathcal{T}}\|_1^2}{d_{\mathcal{T}}},$$

which yields the upper bound of $\|b_{\mathcal{T}}\|_1$. Using (3.8) again, the first inequality (3.4) is verified. At last, via the definition of $CIF(\xi,\mathcal{T})$,

$$\frac{d_{\mathcal{T}}^{1/q}(b^\top\ddot{l}(\beta^o)b)}{CIF_q(\xi,\mathcal{T})\|b_{\mathcal{T}}\|_1\|b\|_q} \ge 1.$$

Then based on (3.8) and (3.9), we have

$$\begin{aligned} te^{-\tilde{t}} \le te^{-Kt} \le \frac{te^{-Kt}d_{\mathcal{T}}^{1/q}(b^\top\ddot{l}(\beta^o)b)}{CIF_q(\xi,\mathcal{T})\|b_{\mathcal{T}}\|_1\|b\|_q} &\le \frac{b^\top(\dot{l}(\beta^o+tb)-\dot{l}(\beta^o))d_{\mathcal{T}}^{1/q}}{CIF_q(\xi,\mathcal{T})\|b_{\mathcal{T}}\|_1\|b\|_q} \\ &\le \frac{\lambda(\xi+1)}{2}\frac{d_{\mathcal{T}}^{1/q}}{CIF_q(\xi,\mathcal{T})\|b\|_q}, \end{aligned}$$

where $t = \|\Delta\|_1$. Note that $\|\Delta\|_q = \|\Delta\|_1\|b\|_q = t\|b\|_q$, and thus the final inequality is obtained. □

Remark 1. From Theorem 3.3, one can find that the estimation and prediction error are both dominated by the tuning parameter λ. If one sets λ to be small, the corresponding errors also become small. However, doing so will decrease the probability of the event $\{z^* \le \frac{\xi-1}{\xi+1}\lambda\}$, under which the desirable inequalities hold. Therefore, the value of λ should not be too small. A practical method for choosing a proper λ will be discussed in the next section.

4. ADMM algorithm

In this section, we develop a generic ADMM algorithm to solve the lasso-type penalized LPRE problem efficiently. Note that the minimization problem in (2.4) can be written as

$$\min_{\beta} L_n(\beta) + \lambda\|\beta\|_1.$$

As demonstrated in the later simulation studies, the proposed algorithm works well when p is larger than n. For the traditional ADMM, it mainly aims to minimize a two-block separable objective function, i.e., there exist two primal variables that needs to be optimized; see [1, 2, 13]. Along this line, one can reformulate the minimization problem above as

$$\min_{(\beta,\alpha)} \frac{1}{n}\sum_{i=1}^{n}\left[\tilde{Y}_i\exp(-X_i^\top\beta) + \tilde{Y}_i^{-1}\exp(X_i^\top\beta)\right] + \lambda\|\alpha\|_1, \tag{4.1}$$

subject to $\beta = \alpha$, or

$$\min_{(\beta,r)} \frac{1}{n} \sum_{i=1}^{n} [\exp(-r_i) + \exp(r_i)] + \lambda\|\beta\|_1, \tag{4.2}$$

subject to $Y - X^\top\beta = r$, where $Y = (\log(\tilde{Y}_1), ..., \log(\tilde{Y}_n))^\top$ and $r = (r_1, ..., r_n)^\top$.

Whether based on (4.1) or (4.2), there exists at least one minimization subproblem that either doesn't have a closed form or can't be rapidly solved by common methods, such as the Newton-Raphson algorithm. This means that the computation burden involved therein is rather heavy and the efficiency is low, which is particularly undesirable in the $p > n$ case. To avoid this drawback, [9] suggested to approximate the loss function term, namely, the first term in the objective function by a quadratic function first, then to apply the existing regularization algorithm based on the least square criterion. In their method, a well-defined initial estimator of the true regression coefficients plays an important role in numerical performance. What's more, finding such initial values will become very difficult for large-scale problems.

4.1. Algorithm

To develop an applicable ADMM algorithm for practitioners, one alternative is to introduce some auxiliary variables into the original objective function. Then, a multiblock ADMM algorithm is constructed, but the convergence property isn't usually guaranteed. Fortunately, the results in [2] demonstrate that if the auxiliary variables are chosen carefully, the convergence of the multiblock ADMM algorithm is still guaranteed. Inspired by this finding, and by integrating the merits of (4.1) and (4.2), a three-block ADMM algorithm is developed. Explicitly speaking, the minimization problem in (2.4) can be rewritten as

$$\min_{(\beta,\alpha,r)} \frac{1}{n} \sum_{i=1}^{n} [\exp(-r_i) + \exp(r_i)] + \lambda\|\alpha\|_1, \tag{4.3}$$

subject to $Y - X^\top\beta = r$ and $\beta = \alpha$. Then, the augmented Lagrangian function of (4.3) with the Lagrange multipliers $u_1 \in \mathbb{R}^p$ and $u_2 \in \mathbb{R}^p$ is given by

$$S(\beta, \alpha, r, u_1, u_2) = \frac{1}{n} \sum_{i=1}^{n} [\exp(-r_i) + \exp(r_i)] + \lambda\|\alpha\|_1 + \frac{\eta_1}{2}\|X^\top\beta - Y + r + \eta_1^{-1}u_1\|^2 + \frac{\eta_2}{2}\|\beta - \alpha + \eta_2^{-1}u_2\|^2, \tag{4.4}$$

where $\eta_1 > 0$ and $\eta_2 > 0$ are prespecified penalty parameters, and u_1 and u_2 correspond to the linear constraints. While the new objective function involves more variables and seems to be more complicated, some surprisingly simple updating processes arise. After rearranging the terms in (4.4), omitting some constant terms, and given the values of parameters at the kth iteration step by $\beta^{(k)}$, $\alpha^{(k)}$, $r^{(k)}$, $u_1^{(k)}$, and $u_2^{(k)}$, the ADMM updates of the primal variables β, α, r, u_1, and u_2 are obtained in the iterative process. They can be written

$$\beta^{(k+1)} = \arg\min_{\beta} \frac{\eta_1}{2}\|X^\top\beta - Y + r^{(k)} + \eta_1^{-1}u_1^{(k)}\|^2 + \frac{\eta_2}{2}\|\beta - \alpha^{(k)} + \eta_2^{-1}u_2^{(k)}\|^2, \tag{4.5}$$

$$\alpha^{(k+1)} = \arg\min_{\alpha} \lambda\|\alpha\|_1 + \frac{\eta_2}{2}\|\alpha - \beta^{(k+1)} - \eta_2^{-1}u_2^{(k)}\|^2, \tag{4.6}$$

$$r^{(k+1)} = \arg\min_{r} \frac{1}{n}\sum_{i=1}^{n}[\exp(-r_i)+\exp(r_i)] + \frac{\eta_1}{2}\|r + X^\top\beta^{(k+1)} - Y + \eta_1^{-1}u_1^{(k)}\|^2, \tag{4.7}$$

$$u_1^{(k+1)} = u_1^{(k)} + \eta_1(X^\top\beta^{(k+1)} - Y + r^{(k+1)}), \tag{4.8}$$

$$u_2^{(k+1)} = u_2^{(k)} + \eta_2(\beta^{(k+1)} - \alpha^{(k+1)}). \tag{4.9}$$

For the proposed three-block ADMM algorithm, the main subproblems at each iteration either have closed-form solutions or can be efficiently solved by some simple algorithms. In the following, we describe these results in detail.

First, we consider the subproblem of updating β in (4.5). Note that the minimization problem is quadratic, so the solution to it can be calculated by

$$\beta^{(k+1)} = (\eta_1 X^\top X + \eta_2 I_p)^{-1}(\eta_1 X^\top y_1 + \eta_2 y_2), \tag{4.10}$$

where I_p denotes the p-dimensional identity matrix, $y_1 = r^{(k)} + \eta_1^{-1}u_1^{(k)} - Y$, and $y_2 = \eta_2^{-1}u_2^{(k)} - \alpha^{(k)}$. Due to the fact that $\eta_1 X^\top X + \eta_2 I_p$ is completely determined by the design matrix X, η_1, and η_2, its inversion always exists and can be precalculated before the iteration, which avoids the additional computational burden. In practice, the inversion can be readily solved by Cholesky decomposition when $p < n$. When $p > n$, the Woodbury identity can be applied to efficiently compute the inversion matrix by

$$(\eta_1 X^\top X + \eta_2 I_p)^{-1} = \eta_2^{-1}(I_p - \eta_1\eta_2^{-1}X^\top(I_n + \eta_1\eta_2^{-1}XX^\top)^{-1}X).$$

Especially, when $\eta_1 = \eta_2$, we have $\eta_1\eta_2^{-1} = 1$.

Second, for the subproblem of updating α in (4.6), it is known as the soft-thresholding operator and has a closed form.

Third, we consider the subproblem of updating r in (4.7). Obviously, this objective function is not common, but is continuously differentiable, convex, and separable. These features allow us to solve the solution to it in a component-wise manner. Denote the ith element in $Y - X^\top\beta^{(k+1)} - \eta_1^{-1}u_1^{(k)}$ and $r^{(k+1)}$ by b_i and $r_i^{(k+1)}$, respectively. Then, it yields that

$$r_i^{(k+1)} = \arg\min_{r_i\in\mathbb{R}} \frac{1}{n}(\exp(-r_i)+\exp(r_i)) + \frac{\eta_1}{2}(r_i - b_i)^2. \tag{4.11}$$

Some elementary algebraic calculations indicate that $r_i^{(k+1)}$ lies between 0 and b_i if $b_i > 0$ and conversely if $b_i < 0$. Besides, the Newton-Raphson algorithm can be employed here. In fact, write the objective function in 4.11 by $f(r_i)$, then its first and second derivative function are calculated by

$$f'(r_i) = \frac{1}{n}(-\exp(-r_i)+\exp(r_i)) + \eta_1(r_i - b_i),$$

and

$$f''(r_i) = \frac{1}{n}(\exp(-r_i)+\exp(r_i)) + \eta_1.$$

Given the current iteration value $r_i^{(k+1,l)}$, update it at the $l+1-th$ by

$$r_i^{(k+1,l+1)} = r_i^{(k+1,l)} - f'(r_i^{(k+1,l)})/f''(r_i^{(k+1,l)})$$

until convergence is achieved. Here, we stop the iteration when the number of iterations is larger than 5000 or the absolute error $|r_i^{(k+1,l+1)} - r_i^{(k+1,l)}| < 0.0001$. Our simulation studies demonstrate that the convergence can be achieved after a few iterations with initial values being $b_i/2$.

Remark 2. For the parameters η_1 and η_2, there are several different choices in the literature. In this paper, we set $\eta_1 = \eta_2 = 5.5$. The following simulation studies show that this strategy works well. Other alternatives are discussed in the final section, and how to choose the best values of η_1 and η_2 remains an open problem.

4.2. Convergence analysis

Theorem 4.1. *For any $\eta_1 > 0$ and $\eta_2 > 0$, the sequence $(\beta^{(k)}, \alpha^{(k)}, r^{(k)}, u_1^{(k)}, u_2^{(k)})$ generated by the proposed ADMM algorithm with an arbitrary initial value $(\beta^{(0)}, \alpha^{(0)}, r^{(0)}, u_1^{(0)}, u_2^{(0)})$ satisfying the constraints in the problem converges to $(\beta^{(\infty)}, \alpha^{(\infty)}, r^{(\infty)}, u_1^{(\infty)}, u_2^{(\infty)})$, where $(\beta^{(\infty)}, \alpha^{(\infty)}, r^{(\infty)})$ is a solution point of (4.4).*

Proof. Rewrite the optimization problem in (4.4) as

$$\min_{(\beta,\alpha,r)} \frac{1}{n} \sum_{i=1}^{n} [\exp(-r_i) + \exp(r_i)] + \lambda\|\alpha\|_1,$$

subject to $A_1\beta + A_2\alpha + A_3 r = b$, where $A_1 = (X, I_p)^\top$, $A_2 = (O_n, -I_p)^\top$, $A_3 = (I_n, O_p)^\top$, $b = (Y^\top, 0_p^\top)^\top$, and I_p and O_p denote the p-order identity matrix and zero matrix, respectively. Note that the objective function is the sum of three closed convex functions with respect to β, α, and r, and $A_2^\top A_3 = 0$, which is the condition sufficient to ensure convergence of the proposed algorithm presented in case 1 on page 60 of [2]. By directly applying Theorem 2.4 on page 66 of [2] and following the standard routine in the literature, the result in Theorem 1 above can be shown. □

4.3. Choice of tuning parameter

To implement the algorithm, the tuning parameter λ had to be determined in advance, which plays a significant role in the regularization estimation. In the literature, some criterions have been proposed to select the smoothing parameter λ, such as the Bayesian information criterion (BIC), generalized approximation cross-validation criterion and generalized cross-validation criterion. As done in [9], we select the tuning parameter λ via minimizing the following BIC-type criterions as follows,

$$BIC(\lambda) = \log(L_n(\hat{\beta})) + df_\lambda \frac{\log(n)}{n} C_n, \tag{4.12}$$

where $C_n > 0$, and df_λ denotes the model degrees of freedom, defined by the cardinality of $\{j : \hat{\beta}_j \neq 0\}$. In particular, the set of possible values of $\log(\lambda)$ are determined by the equally spaced grids between -5 and 15, where 50 grids are considered. Of course, using the KKT conditions, an upper bound, which leads all the regression parameters to be zero, can be obtained. However, this data-driven approach performs badly, because the resultant value is rather large, which makes the range of λ very wide and the selection process rather time-consuming. When p is fixed or diverging but smaller than the sample size n, we set $C_n = 1$, but $C_n = \log\log p$ when $p > n$. Later numerical studies prove this method works well.

4.4. Stopping rule of the algorithm and other details

In the literature of ADMM algorithm, one popular stopping rule is based on the calculations of primal error and dual error. The other is based on the convergence analysis mentioned above. Here we adopt the latter, which is also implemented in [13]. When

$$\frac{\|\beta^{(k+1)} - \beta^{(k)}\|}{\max(1, \|\beta^{(k)}\|)} < 0.0001,$$

or the number of iterations exceeds a certain number, typically 5000, the ADMM algorithm stops. In addition, a zero vector is adopted as the initial value.

As for the updation of $u_1^{(k+1)}$ and $u_2^{(k+1)}$, in our simulation studies and real data analysis, we adjust the iteration by $u_1^{(k+1)} = u_1^{(k)} + 1.414(X^\top\beta^{(k+1)} - Y + r^{(k+1)})$ and $u_2^{(k+1)} = u_2^{(k)} + 1.414(\beta^{(k+1)} - \alpha^{(k+1)})$, which is inspired by the strategy in [2]. As for the weights used in the adaptive lasso methods, we utilize the inverse of the sum of 0.000001 and 3/2 powered absolute value of the lasso-penalized estimator. In addition, the lasso-penalized estimator is adopted as the initial value when using the ADMM algorithm to compute the adaptive lasso-penalized LPRE estimator.

5. Simulation studies

In this section, extensive simulation studies are conducted to evaluate the finite-sample performance of the proposed method and compare them with the least squares (LS) and the least absolute deviation (LAD) criteria for log-linear models. For the true regression coefficients β_0 in model (1.1), we consider two settings: One is $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0_{p-6}^\top)^\top$, and the other is $\beta_0 = (4, 3, 2, 0.5, 1, 0.5, 0_{p-6}^\top)^\top$. The dimensions of β_0, and p, are set to be 20 and 80 for the $n > p$ case, and 200 and 400 for the $n < p$ case. The covariate X is generated from a p-variate normal distribution with mean zero and covariance matrix $\Sigma = (\sigma_{jk})$ with $\sigma_{jk} = \rho^{|j-k|}$ and $\rho = 0$, 0.5, -0.5, respectively. For the random error, three kinds of distributions are considered, which are $\log\epsilon \sim Normal(0, 1)$, $\log\epsilon \sim Uniform(-2, 2)$, and $\epsilon \sim f_0(t)$, $f_0(t) = c_1 \exp(-t - t^{-1} - \log(t) + 2)$ for $t > 0$ and 0 for otherwise, where c_1 is a normalization constant. The three kinds of distributions are denoted by LN, LU, and f_0, respectively. It is worth noting that for the third error distribution, the LPRE estimation is efficient. The sample size n is 200 and 400 for the $n > p$ case and 100 for the $n < p$ case. All the results below are based on 1000 replications and implemented by the software R using an Intel(R) Core i3-4150 CPU at 3.5GHz with 16 GB RAM. The main R codes are attached in the supplementary materials.

For convenience, we report our method with lasso and adaptive lasso penalty as PR-L and PR-aL, respectively. Similarly, we denote the LS and LAD methods based estimators by LS-L, LS-aL, LAD-L, and LAD-aL, respectively. The LS-type estimators are implemented by R package `glmnet`, LAD-type estimators are obtained by R packages `quanreg` and `rqPen`.

To assess the performances of these estimators, the following quantities are calculated, the proportion of correctly fitted (CF) and over-fitted (OF) models among 1000 simulations; the false positive rate (FPR) and the false negative rate (FNR), meaning the portion of occasions on which the model selected contains some zero components and nonzero components, respectively; the average (M), median (ME), and sample standard deviation (SD) of true positives (K1) and true negatives (K0); the median of estimation errors (MSE), defined by $\|\hat{\beta} - \beta_0\|^2$; and the median of relative estimation errors (MREE), defined as $(\hat{\beta}-\beta_0)^\top X^\top X(\hat{\beta}-\beta_0)$. Additionally, we also record the average number of the estimated nonzero components with their absolute bias relative to the true values less than 0.1 and zero components with their absolute values less than 0.02, as well as maximum values of the absolute bias lying in the nonzero and zero groups, denoted by N1, N0, Mbias1, and Mbias0, respectively.

As suggested by the reviewers, we conduct s small simulation study on the behavior of different values of $\eta_1 = \eta_2$. The results are shown in Table 1. From it, we can find that $\eta_1 = \eta_2 = 5.5$ is an appropriate choice.

Table 1. The median of estimation errors (MSE), and the median of relative estimation errors (MREE) of the lasso- and adaptive lasso-penalized LPRE estimators, denoted by LR-L and LR-aL, respectively, and the average total computation time (TIME, in seconds) for both estimation methods under n=200,p=20, $\beta_0 = (4, 3, 2, 0.5, 1, 0.5, 0, ...0)^\top$, $\epsilon \sim f_0(t)$ for different $\eta_1 = \eta_2$ values.

$\eta_1 = \eta_2$	0.1	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5
LR-L											
MSE	0.039	0.039	0.042	0.044	0.047	0.046	0.045	0.051	0.046	0.050	0.048
MREE	0.037	0.036	0.040	0.041	0.043	0.045	0.042	0.050	0.043	0.049	0.044
LR-aL											
MSE	0.010	0.010	0.011	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012
MREE	0.010	0.011	0.011	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012
$\eta_1 = \eta_2$	15	25	50	100	200						
LR-L											
MSE	0.048	0.050	0.055	0.056	0.051						
MREE	0.044	0.047	0.054	0.053	0.047						
LR-aL											
MSE	0.012	0.012	0.013	0.013	0.013						
MREE	0.012	0.012	0.012	0.013	0.013						
$\eta_1 = \eta_2$	0.1	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5
TIME	2.503	2.784	6.235	7.939	8.497	9.340	9.995	10.635	11.214	11.666	12.092
$\eta_1 = \eta_2$	15	25	50	100	200						
TIME	13.948	15.665	16.825	16.042	11.414						

Figures 1–3 depict the histograms, boxplots, and QQ-plots of the adaptive lasso-penalized LPRE estimators for nonzero regression parameters under one case, respectively, where $n = 100$, $p = 200$, $\epsilon \sim f_0(t)$, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0_{p-6}^\top)^\top$. Tables 2–5 present the variable selection results and estimation results under n=200, p=80, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0, ...0)^\top$. In addition, Tables S1–S16 in the supplementary materials summarize all the numerical results. From the numerical experiment, we draw the following findings:

- In view of the numerical computation time per replication for various model settings, where six estimators are calculated and determined by minimizing the BIC-type information criterion with at least 50 different values of the tuning parameter; on average it takes from 9.1 to 15.4 seconds for cases with $p = 20, n = 200$, from 17.2 to 28.3 seconds for cases with $p = 80$ and $n = 200$, from 16.7 to 24.6 seconds for cases with $p = 20$ and $n = 400$, and from 25.1 to 39.7 seconds for cases with $p = 80$ and $n = 400$. In other words, for a given value of the tuning parameter, all six estimators can be obtained after running less than 0.4 seconds on average. This means that our algorithm has the capacity to compute the regularization paths effectively. When there exist some small values in the true value of the regression coefficient, for any method, the computation will need slightly more time.

- From the perspective of variable selection, as presented in Tables S1–S8 in the supplementary materials, adaptive lasso-type estimators perform much better than the lasso-type alternatives in terms of accuracy. The former tends to achieve more correctly-fitted models and lower false positive rates. In the following, we limit ourselves to the adaptive lasso-type estimators. When the random error follows the normal distribution, the LS method is regarded as the most efficient, and indeed, this is shown in the tables. One can find that the false negative rate of the LS-based estimator is the lowest and nearly equal to zero. The number of selected positives and negatives are almost approaching those of the true model. Compared with it, the LPRE and LAD methods behave slightly worse. Especially, the LAD method tends to produce an over-fitted model, and the LPRE method is more likely to select the true model. When the random error is generated from the density f_0, the LPRE method would be the most efficient. At this time, some contrary observations are made. LPRE-based estimators can almost select the true model in all 1000 replications, where all CF values are approaching to one, and both NPR and NFR are approaching zero. However, the performance of LAD-based estimators isn't the worst, because the LAD method is relatively robust.
- In view of the performance of parameter estimation in Tables 4–5, and Tables S9–S16 in the supplementary materials, findings similar to those indicated in the variable selection part are observed, except that the performance of the lasso-penalized LAD-based estimators is the worst. Adaptive lasso-type LPRE estimators present the smallest Mbias0, MSE, and MREE compared to the lasso-type alternatives in most scenarios. Besides, whether the random error follows the normal distribution or the other two distributions, LPRE-based adaptive lasso estimators provide the best estimates of the nonzero components and distinguish others significantly according to the N1 values. Of course, this point is rather prominent when the random error follows from the density f_0, where Mbias0 values are exactly zero, which means that in all 1000 replications under each case, PR-aL can identify the zero components correctly.
- As the sample size n increases, but is still less than p, the accuracy of both variable selection and parameter estimation are improved significantly for all methods. For a fixed sample size, when the dimension p grows, the accuracy tends to decrease slightly. An interesting finding is that for given n, p, and ϵ, the performance under $\rho = -0.5$ is the worst, and $\rho = 0.5$ cases usually produce the best results. $\rho = 0$ corresponds to the independent cases that produce the results in between.
- Tables S7, S8, S15, and S16 in the supplementary materials present the results when $p > n$, i.e. $n = 100, p = 200$. The total numerical computation time is shorter than time used in $n = 400, p = 80$ case for each model setting. Precisely speaking, the ratios of the two times on average vary from 0.53 seconds to 0.78 seconds. One can find that the performance of LPRE-based adaptive lasso method still is the most powerful to identify the important variables and remove those non-effective ones in most cases.

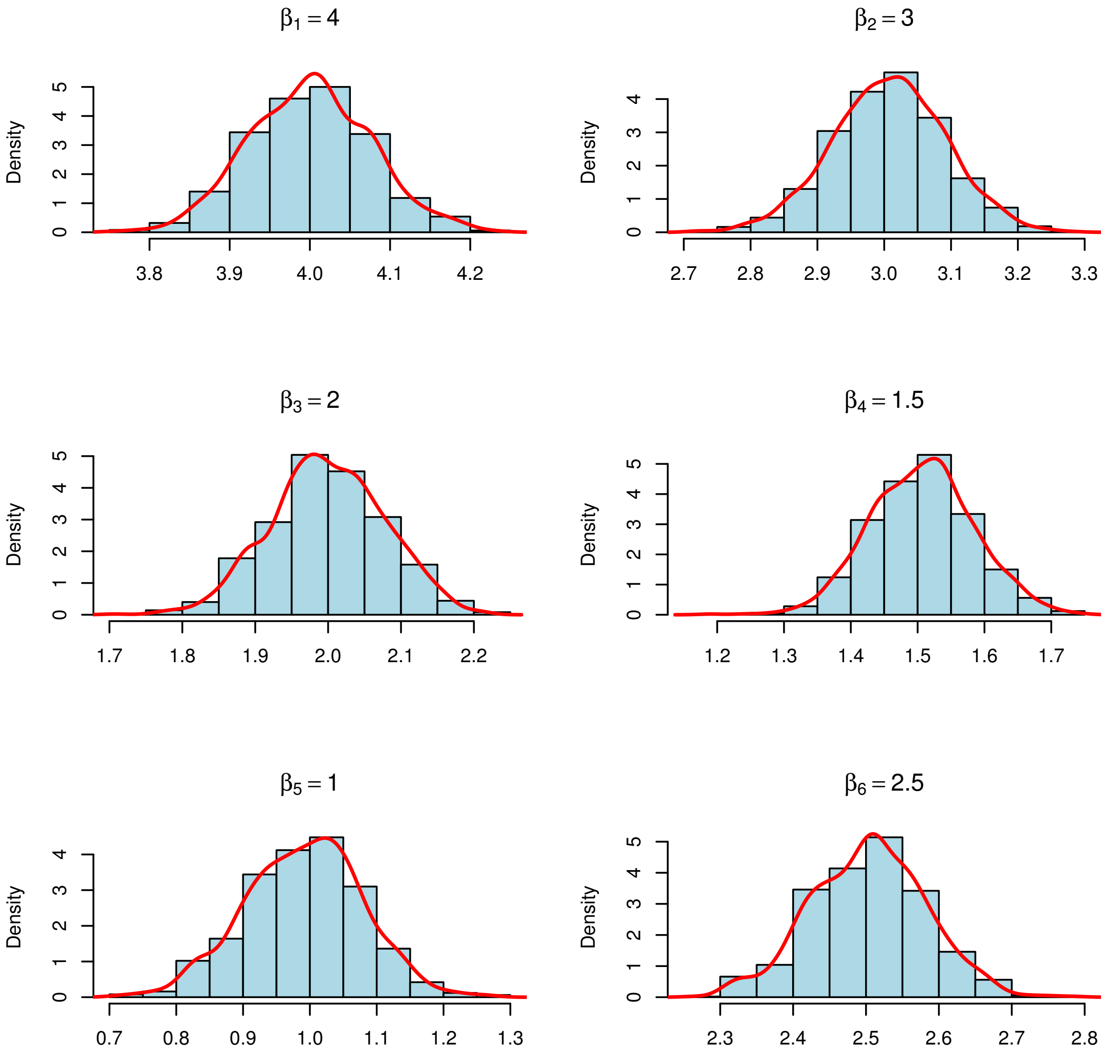

Figure 1. Histograms of the adaptive lasso-penalized LPRE estimators for nonzero regression parameters under one case, where $n = 100$, $p = 200$, $\epsilon \sim f_0(t)$, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0_{p-6}^\top)^\top$.

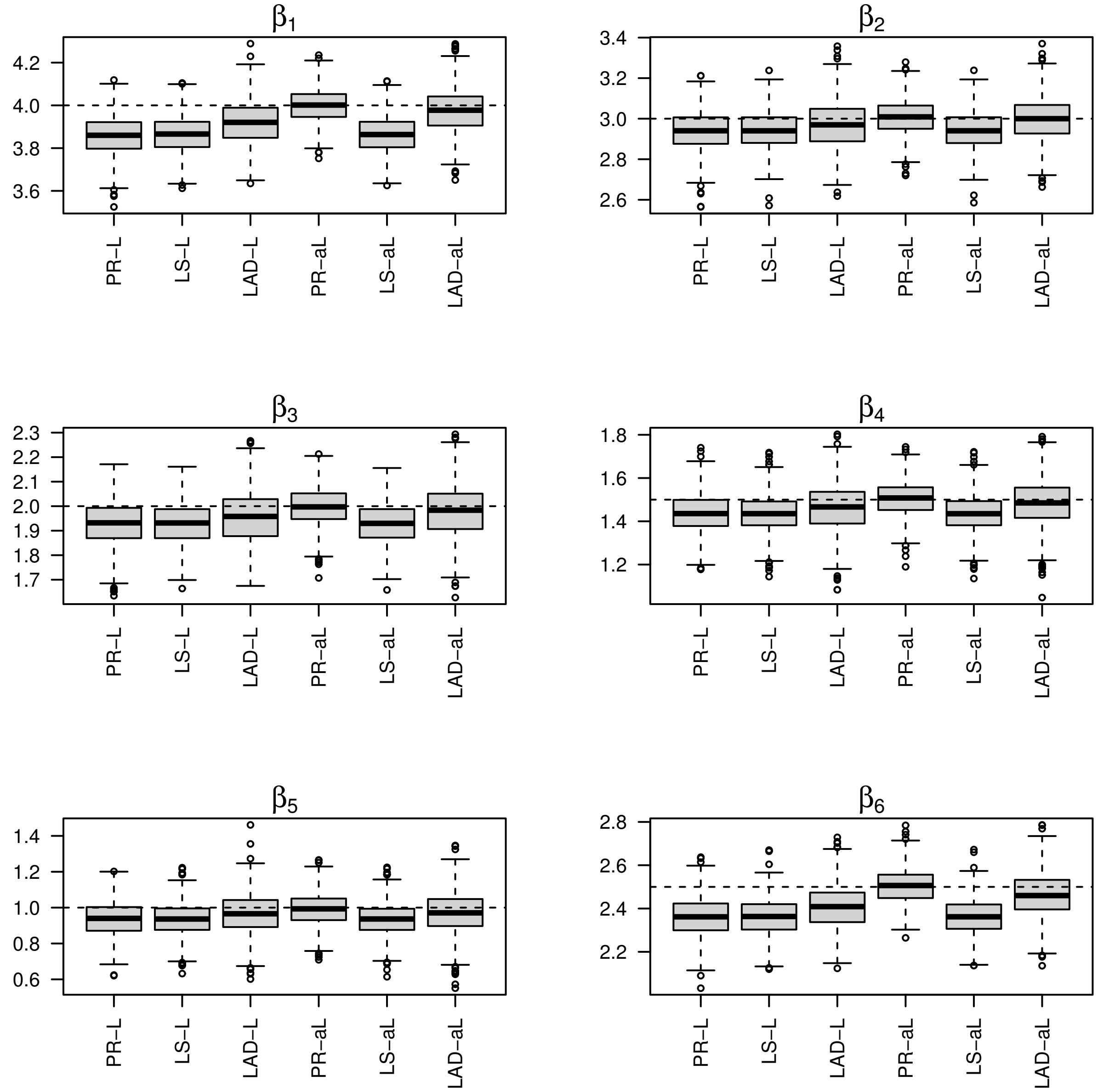

Figure 2. Boxplots of all six estimation methods under one case, where $n = 100$, $p = 200$, $\epsilon \sim f_0(t)$, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0_{p-6}^{\top})^{\top}$. PR-L and PR-aL denote our method with a lasso and adaptive lasso penalty, respectively. Similarly, we denote the LS and LAD methods based estimators by LS-L, LS-aL, LAD-L, and LAD-aL, respectively.

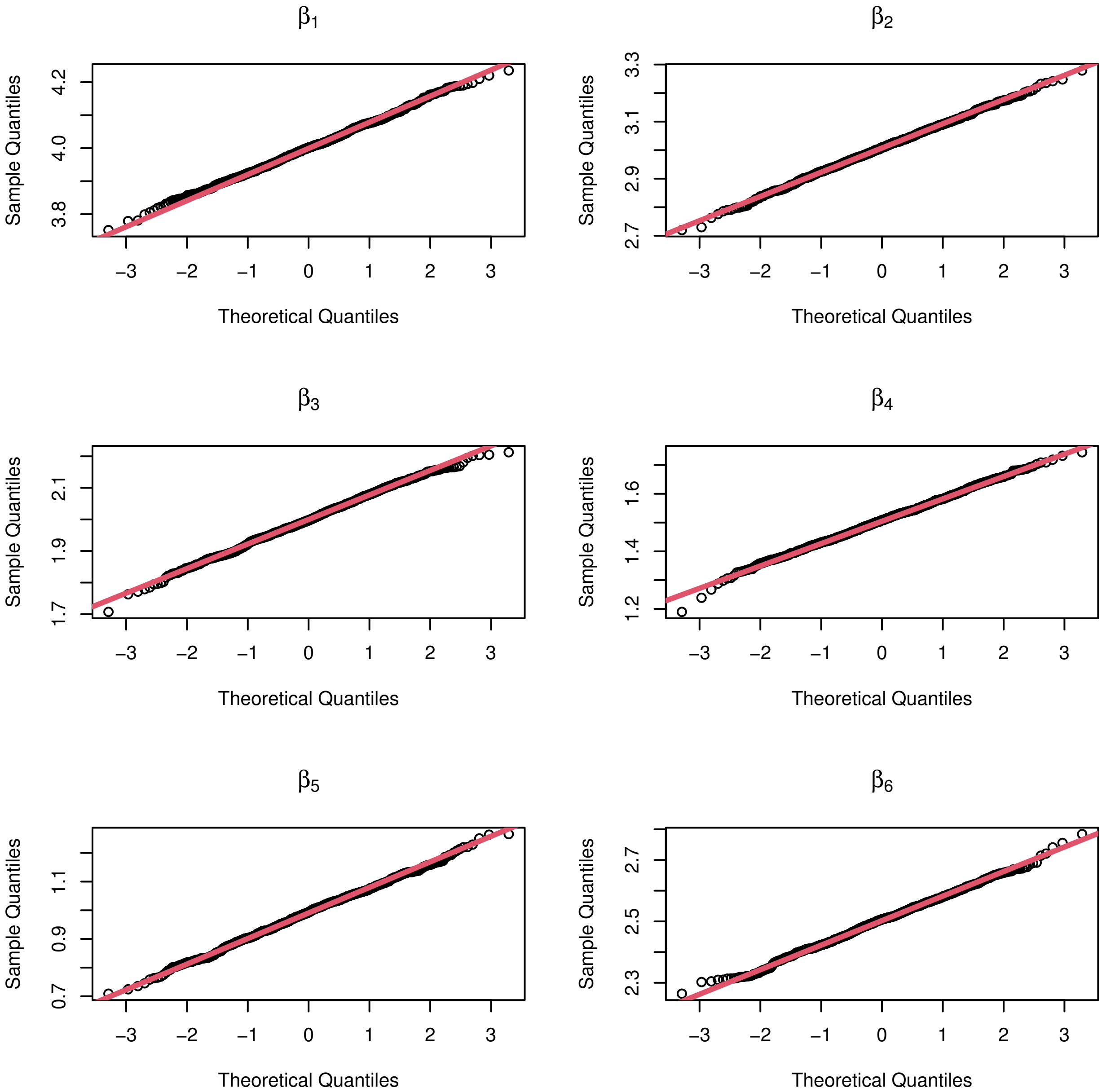

Figure 3. QQ-plots of the adaptive lasso-penalized LPRE estimators for nonzero regression parameters under one case, where $n = 100$, $p = 200$, $\epsilon \sim f_0(t)$, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0_{p-6}^{\top})^{\top}$.

Table 2. Variable selection results under n=200, p=80, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	OF	CF	FPR	FNR	K0-M	K0-ME	K0-SD	K1-M	K1-ME	K1-SD
LN	0	PR-L	0.728	0.272	0.728	0.000	72.553	73.000	1.293	6.000	6.000	0.000
		LS-L	0.631	0.369	0.631	0.000	72.359	73.000	2.166	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	59.713	60.000	3.233	6.000	6.000	0.000
		PR-aL	0.031	0.969	0.031	0.000	73.968	74.000	0.182	6.000	6.000	0.000
		LS-aL	0.624	0.376	0.624	0.000	72.438	73.000	2.016	6.000	6.000	0.000
		LAD-aL	0.567	0.433	0.567	0.000	71.609	73.000	3.858	6.000	6.000	0.000
	0.5	PR-L	0.374	0.626	0.374	0.000	73.499	74.000	0.751	6.000	6.000	0.000
		LS-L	0.211	0.789	0.211	0.000	73.627	74.000	0.964	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	61.825	62.000	2.967	6.000	6.000	0.000
		PR-aL	0.037	0.963	0.037	0.000	73.961	74.000	0.204	6.000	6.000	0.000
		LS-aL	0.218	0.782	0.218	0.000	73.626	74.000	0.969	6.000	6.000	0.000
		LAD-aL	0.398	0.602	0.398	0.000	72.328	74.000	3.131	6.000	6.000	0.000
	-0.5	PR-L	0.984	0.016	0.984	0.000	69.383	70.000	2.383	6.000	6.000	0.000
		LS-L	0.988	0.012	0.988	0.000	68.049	69.000	3.573	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	61.492	62.000	3.046	6.000	6.000	0.000
		PR-aL	0.033	0.967	0.033	0.000	73.962	74.000	0.225	6.000	6.000	0.000
		LS-aL	0.985	0.015	0.985	0.000	67.951	69.000	3.784	6.000	6.000	0.000
		LAD-aL	0.915	0.079	0.921	0.006	68.593	70.000	5.072	5.994	6.000	0.077
f_0	0	PR-L	0.273	0.727	0.273	0.000	73.691	74.000	0.536	6.000	6.000	0.000
		LS-L	0.558	0.442	0.558	0.000	72.874	73.000	1.487	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	62.657	63.000	3.041	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	74.000	74.000	0.000	6.000	6.000	0.000
		LS-aL	0.543	0.457	0.543	0.000	72.905	73.000	1.450	6.000	6.000	0.000
		LAD-aL	0.447	0.553	0.447	0.000	72.421	74.000	2.750	6.000	6.000	0.000
	0.5	PR-L	0.056	0.944	0.056	0.000	73.943	74.000	0.236	6.000	6.000	0.000
		LS-L	0.168	0.832	0.168	0.000	73.697	74.000	0.950	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	63.868	64.000	2.845	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	74.000	74.000	0.000	6.000	6.000	0.000
		LS-aL	0.176	0.824	0.176	0.000	73.690	74.000	0.926	6.000	6.000	0.000
		LAD-aL	0.330	0.670	0.330	0.000	72.874	74.000	2.329	6.000	6.000	0.000
	-0.5	PR-L	0.802	0.198	0.802	0.000	72.379	73.000	1.262	6.000	6.000	0.000
		LS-L	0.971	0.029	0.971	0.000	69.422	70.000	2.905	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	63.743	64.000	2.892	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	74.000	74.000	0.000	6.000	6.000	0.000
		LS-aL	0.968	0.032	0.968	0.000	69.365	70.000	3.035	6.000	6.000	0.000
		LAD-aL	0.724	0.276	0.724	0.000	71.179	72.000	3.416	6.000	6.000	0.000

Table 3. Variable selection results under n=400, p=80, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	OF	CF	FPR	FNR	K0-M	K0-ME	K0-SD	K1-M	K1-ME	K1-SD
LN	0	PR-L	0.635	0.365	0.635	0.000	72.931	73.000	1.086	6.000	6.000	0.000
		LS-L	0.336	0.664	0.336	0.000	73.437	74.000	1.054	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	61.476	62.000	3.260	6.000	6.000	0.000
		PR-aL	0.011	0.989	0.011	0.000	73.988	74.000	0.118	6.000	6.000	0.000
		LS-aL	0.345	0.655	0.345	0.000	73.452	74.000	0.986	6.000	6.000	0.000
		LAD-aL	0.359	0.641	0.359	0.000	72.832	74.000	2.270	6.000	6.000	0.000
	0.5	PR-L	0.302	0.698	0.302	0.000	73.607	74.000	0.669	6.000	6.000	0.000
		LS-L	0.083	0.917	0.083	0.000	73.877	74.000	0.490	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	63.492	64.000	2.949	6.000	6.000	0.000
		PR-aL	0.010	0.990	0.010	0.000	73.989	74.000	0.114	6.000	6.000	0.000
		LS-aL	0.086	0.914	0.086	0.000	73.873	74.000	0.536	6.000	6.000	0.000
		LAD-aL	0.335	0.665	0.335	0.000	72.861	74.000	2.327	6.000	6.000	0.000
	-0.5	PR-L	0.938	0.062	0.938	0.000	70.906	71.000	1.873	6.000	6.000	0.000
		LS-L	0.914	0.086	0.914	0.000	70.925	71.000	2.381	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	63.152	63.000	2.993	6.000	6.000	0.000
		PR-aL	0.009	0.991	0.009	0.000	73.991	74.000	0.094	6.000	6.000	0.000
		LS-aL	0.903	0.097	0.903	0.000	70.950	71.000	2.414	6.000	6.000	0.000
		LAD-aL	0.584	0.416	0.584	0.000	72.003	73.000	2.918	6.000	6.000	0.000
f_0	0	PR-L	0.151	0.849	0.151	0.000	73.833	74.000	0.414	6.000	6.000	0.000
		LS-L	0.245	0.755	0.245	0.000	73.650	74.000	0.728	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	64.871	65.000	2.742	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	74.000	74.000	0.000	6.000	6.000	0.000
		LS-aL	0.250	0.750	0.250	0.000	73.652	74.000	0.710	6.000	6.000	0.000
		LAD-aL	0.319	0.681	0.319	0.000	73.224	74.000	1.485	6.000	6.000	0.000
	0.5	PR-L	0.033	0.967	0.033	0.000	73.967	74.000	0.179	6.000	6.000	0.000
		LS-L	0.050	0.950	0.050	0.000	73.941	74.000	0.279	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	66.192	66.000	2.595	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	74.000	74.000	0.000	6.000	6.000	0.000
		LS-aL	0.048	0.952	0.048	0.000	73.936	74.000	0.349	6.000	6.000	0.000
		LAD-aL	0.262	0.738	0.262	0.000	73.394	74.000	1.255	6.000	6.000	0.000
	-0.5	PR-L	0.624	0.376	0.624	0.000	73.037	73.000	0.958	6.000	6.000	0.000
		LS-L	0.815	0.185	0.815	0.000	72.017	72.000	1.686	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	66.002	66.000	2.509	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	74.000	74.000	0.000	6.000	6.000	0.000
		LS-aL	0.818	0.182	0.818	0.000	71.946	72.000	1.784	6.000	6.000	0.000
		LAD-aL	0.428	0.572	0.428	0.000	72.904	74.000	1.782	6.000	6.000	0.000

Table 4. Estimation results under n=200, p=80, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	N0	N1	Mbias0	Mbias1	MSE	MREE
LN	0	PR-L	72.553	1.439	0.200	0.523	0.188	0.184
		LS-L	72.359	1.122	0.174	0.502	0.211	0.204
		LAD-L	59.713	2.846	0.343	0.438	0.189	0.182
		PR-aL	73.968	4.859	0.300	0.323	0.032	0.030
		LS-aL	72.438	1.110	0.174	0.517	0.208	0.204
		LAD-aL	71.609	4.161	0.281	0.398	0.067	0.064
	0.5	PR-L	73.499	3.357	0.180	0.467	0.092	0.099
		LS-L	73.627	2.875	0.142	0.482	0.116	0.145
		LAD-L	61.825	3.502	0.326	0.439	0.148	0.133
		PR-aL	73.961	4.302	0.286	0.387	0.045	0.029
		LS-aL	73.626	2.894	0.151	0.455	0.113	0.141
		LAD-aL	72.328	3.796	0.383	0.450	0.077	0.054
	-0.5	PR-L	69.383	0.159	0.284	0.857	0.746	0.302
		LS-L	68.049	0.206	0.259	0.752	0.660	0.272
		LAD-L	61.492	0.524	0.312	0.789	0.611	0.293
		PR-aL	73.962	4.194	0.318	0.444	0.049	0.032
		LS-aL	67.951	0.201	0.259	0.785	0.662	0.272
		LAD-aL	68.593	2.907	0.370	1.000	0.162	0.100
f_0	0	PR-L	73.691	2.640	0.134	0.341	0.083	0.073
		LS-L	72.874	2.900	0.137	0.271	0.075	0.066
		LAD-L	62.657	4.230	0.227	0.303	0.078	0.073
		PR-aL	74.000	5.838	0.000	0.163	0.011	0.010
		LS-aL	72.905	2.872	0.131	0.291	0.076	0.066
		LAD-aL	72.421	5.370	0.198	0.265	0.026	0.025
	0.5	PR-L	73.943	4.669	0.112	0.321	0.036	0.043
		LS-L	73.697	4.404	0.106	0.273	0.042	0.055
		LAD-L	63.868	4.655	0.200	0.314	0.065	0.058
		PR-aL	74.000	5.547	0.000	0.217	0.016	0.011
		LS-aL	73.690	4.399	0.101	0.290	0.042	0.056
		LAD-aL	72.874	4.971	0.200	0.292	0.032	0.023
	-0.5	PR-L	72.379	0.183	0.147	0.639	0.412	0.163
		LS-L	69.422	0.634	0.130	0.464	0.248	0.105
		LAD-L	63.743	1.011	0.216	0.535	0.259	0.126
		PR-aL	74.000	5.554	0.000	0.266	0.016	0.011
		LS-aL	69.365	0.645	0.133	0.439	0.248	0.105
		LAD-aL	71.179	4.347	0.202	0.414	0.053	0.035

Table 5. Estimation results under n=400, p=80, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	N0	N1	Mbias0	Mbias1	MSE	MREE
LN	0	PR-L	72.931	2.798	0.126	0.313	0.080	0.076
		LS-L	73.437	1.752	0.119	0.328	0.115	0.107
		LAD-L	61.476	4.146	0.214	0.282	0.088	0.086
		PR-aL	73.988	5.698	0.139	0.208	0.014	0.014
		LS-aL	73.452	1.725	0.103	0.346	0.115	0.107
		LAD-aL	72.832	5.276	0.213	0.268	0.027	0.027
	0.5	PR-L	73.607	4.463	0.133	0.292	0.041	0.049
		LS-L	73.877	3.699	0.092	0.337	0.063	0.092
		LAD-L	63.492	4.552	0.219	0.366	0.072	0.067
		PR-aL	73.989	5.268	0.185	0.232	0.021	0.014
		LS-aL	73.873	3.678	0.101	0.352	0.063	0.091
		LAD-aL	72.861	4.810	0.201	0.307	0.037	0.026
	-0.5	PR-L	70.906	0.430	0.156	0.579	0.334	0.146
		LS-L	70.925	0.298	0.143	0.527	0.372	0.160
		LAD-L	63.152	1.027	0.258	0.526	0.270	0.139
		PR-aL	73.991	5.171	0.135	0.261	0.023	0.015
		LS-aL	70.950	0.282	0.138	0.547	0.369	0.162
		LAD-aL	72.003	4.271	0.254	0.373	0.057	0.035
f_0	0	PR-L	73.833	4.699	0.085	0.197	0.037	0.035
		LS-L	73.650	4.363	0.066	0.197	0.044	0.041
		LAD-L	64.871	5.417	0.139	0.193	0.038	0.036
		PR-aL	74.000	5.992	0.000	0.112	0.005	0.005
		LS-aL	73.652	4.381	0.068	0.209	0.044	0.041
		LAD-aL	73.224	5.897	0.138	0.160	0.011	0.011
	0.5	PR-L	73.967	5.653	0.085	0.186	0.016	0.019
		LS-L	73.941	5.298	0.064	0.224	0.024	0.035
		LAD-L	66.192	5.484	0.155	0.230	0.031	0.027
		PR-aL	74.000	5.927	0.000	0.162	0.008	0.005
		LS-aL	73.936	5.307	0.062	0.224	0.024	0.035
		LAD-aL	73.394	5.717	0.131	0.216	0.015	0.011
	-0.5	PR-L	73.037	0.909	0.096	0.390	0.167	0.074
		LS-L	72.017	1.274	0.102	0.330	0.134	0.060
		LAD-L	66.002	2.290	0.166	0.337	0.115	0.060
		PR-aL	74.000	5.936	0.000	0.140	0.008	0.005
		LS-aL	71.946	1.254	0.104	0.330	0.133	0.060
		LAD-aL	72.904	5.478	0.153	0.243	0.020	0.013

Simulation results for $\beta_0 = (4, 3, 2, 0.5, 1, 0.5, 0_{p-6}^\top)^\top$ with $n = 400, p = 20, 80$, and $n = 100, p = 200, 400$ are similar to Tables S5–S8, and S13–S16 in the supplementary materials, therefore omitted here, but can be available from the authors upon request.

6. Real data analysis

In this section, we apply a real data set of body fat data to illustrate our proposed method. This data set can be found via `https://lib.stat.cmu.edu/datasets/bodyfat` and collected to investigate the relationship of the Percent of body fat ($\tilde{Y}$) and other 13 measurements, denoted by $X_1, ..., X_{13}$ and described as follows: Age (years); Weight (lbs); Height (inches); Neck circumference (cm); Chest circumference (cm); Abdomen 2 circumference (cm); Hip circumference (cm); Thigh circumference (cm); Knee circumference (cm); Ankle circumference (cm); Biceps (extended) circumference (cm); Forearm circumference (cm) and Wrist circumference (cm). In the original data set, there are 252 observations. After deleting one observation with 0 percent of body fat and three outliers, the remaining 248 observations are analyzed. Using the result in [14], we set the intercept term to be 2.848, and all continuous covariates are normalized.

Under the same parameter setting as in the simulation studies, the analysis results of this data are summarized in Table 6. PR-aL yields the sparsest model, while LAD-aL has the largest number of covariates. What's more, X_6 plays the most important role in all three methods and our result is consistent with the finding in [14], where the regression coefficient of X_6 is estimated to be 0.560.

Table 6. Resuts of analyzing the body fat data.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
PR-aL						0.361							
LS-aL			-0.022			0.357							
lAD-aL				-0.027		0.381							-0.025
	MPE	MAPE	MSPE										
PR-L	0.414	0.857	0.171										
LS-aL	0.406	0.852	0.165										
lAD-aL	0.403	0.844	0.162										

Besides, we also calculate the performance of model fitting, measured by median of absolute prediction errors $\{|\hat{Y}_i - \tilde{Y}_i|\}$, median of additive relative prediction errors $\{|\tilde{Y}_i - \exp(X_i^\top \hat{\beta})|/\tilde{Y}_i + |\tilde{Y}_i - \exp(X_i^\top \hat{\beta})|/\exp(X_i^\top \hat{\beta})\}$, and median of squared prediction errors $\{|\hat{Y}_i - \tilde{Y}_i|^2\}$, denoted by MPE, MAPE, and MSPE, respectively. Table 6 shows that the model fits with the proposed PR-aL estimator is comparable with that of LS-aL and LAD-aL.

7. Conclusions

In this paper, we establish nonasymptotic oracle inequalities for the ℓ_1 estimation error and prediction error measured by the SB divergence of the lasso-penalized LPRE estimator. In addition, we developed a computationally efficient and generic ADMM algorithm for solving the lasso-type penalized problem. Unlike the existing numeric methods, the proposed approach doesn't need to prepare a well-behaved initial value at which the quadratic approximation of the LPRE loss function had to be undertaken. What's more, the convergence of the new algorithm is theoretically guaranteed and is independent of the choice of the initial value, which is more desirable for practical applications.

Extensive simulation studies were conducted and demonstrate that the proposed algorithm works well and efficiently, even when the dimension of covariate is larger than the sample size.

According to the suggestion in [1], the computation cost can be further eased by varying the superfluous parameters η_1 and η_2. How to develop this and design a faster program code will be pursued in the future. Although this paper only focuses on the lasso-type penalty, extension to other penalty functions, such as the nonconvex SCAD or MCP penalty is straightforward if the local linear or quadratic approximation of these penalty functions behaves well. In addition, our research interests include developing more efficient algorithms to solve the general relative error criterion-based regularization problem with high-dimensional data, and similar problems that occur in other semiparametric multiplicative regression models, and addressing multicollinearity and outliers or a geographic weighted method in multiplicative regression like [8, 18].

Author contributions

Mingzhen Wan: Computation, writing, and editing; Wei Chen: Original draft, review, and supervision. Both authors have read and agreed to the submitted version of the manuscript.

Use of Generative-AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors thank the anonymous referees for their helpful comments and suggestions. Wan's research is partly supported by Project 2025SJSZ0663, titled "New Quality Productivity Drives University Employment Guidance Transformation: Multi-Subject Collaborative Cultivation Path Exploration under the Three-Dimensional Coupling Mechanism", which was established as a general project in the 2025 Jiangsu Province University Philosophy and Social Science Research. The project is hosted by the Suzhou Institute of Technology at Jiangsu University of Science and Technology.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Found. Trends Inf. Ret.*, **3** (2011), 1–122. https://doi.org/10.1561/2200000016
2. C. Chen, B. He, Y. Ye, X. Yuan, The direct extension of admm for multi-block convex minimization problems is not necessarily convergent, *Math. Program.*, **155** (2016), 57–79. https://doi.org/10.1561/2200000016

3. K. Chen, S. Guo, Y. Lin, Z. Ying, Least absolute relative error estimation, *J. Am. Stat. Assoc.*, **105** (2010), 1104–1112. https://doi.org/10.1198/jasa.2010.tm09307

4. K. Chen, Y. Lin, Z. Wang, Z. Ying, Least product relative error estimation, *J. Multivariate Anal.*, **144** (2016), 91–98. https://doi.org/10.1016/j.jmva.2015.10.017

5. Y. Chen, H. Liu, J. Ma, Local least product relative error estimation for single-index varying-coefficient multiplicative model with positive responses, *J. Comput. Appl. Math.*, **415** (2022), 114478. https://doi.org/10.1016/j.cam.2022.114478

6. Y. Chen, H. Ming, H. Yang, Efficient variable selection for high-dimensional multiplicative models: a novel LPRE-based approach, *Stat. Papers*, **65** (2024), 3713–3737. https://doi.org/10.1007/s00362-024-01545-1

7. B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, *Ann. Statist.*, **32** (2004), 407–451. https://doi.org/10.1214/009053604000000067

8. A. T. Hammad, I. Elbatal I, E. M. Almetwally, M. M. Abd El-Raouf, M. A. El-Qurashi, Ahmed M. Gemeay, A novel robust estimator for addressing multicollinearity and outliers in Beta regression: simulation and application, *AIMS Math.*, **10** (2025), 21549–21580. https://doi.org/10.3934/math.2025958

9. M. Hao, Y. Lin, X. Zhao, A relative error-based approach for variable selection, *Comput. Stat. Data. An.*, **103** (2016), 250–262. https://doi.org/10.1016/j.csda.2016.05.013

10. D. Hu, Local least product relative error estimation for varying coefficient multiplicative regression model, *Acta. Math. Appl. Sin. Engl. Ser.*, **35** (2019), 274–286. https://doi.org/10.1007/s10255-018-0794-2

11. J. Huang, T. Sun, Z. Ying, Y. Yu, C. H. Zhang, Oracle inequalities for the lasso in the cox model, *Ann. Statist.*, **41** (2013), 1142–1165. https://doi.org/10.1214/13-AOS1098

12. T. M. Khoshgoftaar, B. B. Bhattacharyya, G. D. Richardson, Predicting software errors, during development, using nonlinear regression models: a comparative study, *IEEE T. Reliab.*, **41** (1992), 390–395. https://doi.org/10.1109/24.159804

13. X. Li, L. Mo, X. Yuan, J. Zhang, Linearized alternating direction method of multipliers for sparse group and fused lasso models, *Comput. Stat. Data. An.*, **79** (2014), 203–221. https://doi.org/10.1016/j.csda.2014.05.017

14. X. Liu, Y. Lin, Z. Wang, Group variable selection for relative error regression, *J. Stat. Plan. Infer.*, **175** (2016), 40–50. https://doi.org/10.1016/j.jspi.2016.02.006

15. P. Liu, L. Chen, M. Bai, An accelerated semi-proximal ADMM with applications to multi-block sparse optimization problems, *J. Sci. Comput.*, **104** (2025), 30. https://doi.org/10.1007/s10915-025-02951-9

16. S. C. Narula, J. F. Wellington, Prediction, linear regression and the minimum sum of relative errors, *Technometrics*, **19** (1977), 185–190. https://doi.org/10.1080/00401706.1977.10489526

17. H. Park, L. A. Stefanski, Relative-error prediction, *Stat. Probabil. Lett.*, **40** (1998), 227–236. https://doi.org/10.1016/S0167-7152(98)00088-1

18. O. F. Salih Al-Rawi, Z. Y. Algamal, Geospatial modeling of under-five mortality in Iraq based on geographic weighted regression model, *Netw. Model. Anal. Health Inform. Bioinform.*, **15** (2026), 5. https://doi.org/10.1007/s13721-025-00694-z

19. P. J. Bickel, Y. Ritov, A. B. Tsybakov, Simultaneous analysis of Lasso and Dantzig selector, *Ann. Statist.*, **37** (2009), 1705–1732. https://doi.org/10.1214/08-AOS620

20. S. A. van de Geer, High-dimensional generalized linear models and the lasso, *Ann. Statist.*, **36** (2008), 614–645. https://doi.org/10.1214/009053607000000929

21. S. A. van de Geer, P. Bühlmann, On the conditions used to prove oracle results for the lasso, *Electron. J. Statist.*, **3** (2009), 1360–1392. https://doi.org/10.1214/09-EJS506

22. H. Wang, C. Leng., Unified lasso estimation by least squares approximation, *J. Am. Stat. Assoc.*, **102** (2007), 1039–1048. https://doi.org/10.1198/016214507000000509

23. X. Xia, Z. Liu, H. Yang, Regularized estimation for the least absolute relative error models with a diverging number of covariates, *Comput. Stat. Data. An.*, **96** (2016), 104–119. https://doi.org/10.1016/j.csda.2015.10.012

24. F. Ye, C. H. Zhang, Rate minimaxity of the lasso and dantzig selector for the lq loss in lr balls, *J. Mach. Learn. Res.*, **11** (2010), 3519–3540.

25. H. Zhang, L. Sun, Y. Zhou, J. Huang, Oracle inequalities and selection consistency for weighted lasso in high-dimensional additive hazards model, *Stat. Sinica*, **27** (2017), 1903–1920. https://doi.org/10.5705/ss.202015.0075

26. J. Zhang, Z. Feng, H. Peng, Estimation and hypothesis test for partial linear multiplicative models, *Comput. Stat. Data. An.*, **128** (2018), 87–103. https://doi.org/10.1016/j.csda.2018.06.017

27. J. Zhang, J. Zhu, Z. Feng, Estimation and hypothesis test for single-index multiplicative models, *Test*, **28** (2019), 242–268. https://doi.org/10.1007/s11749-018-0586-2

28. Q. Zhang, Q. Wang, Local least absolute relative error estimating approach for partially linear multiplicative model, *Stat. Sinica*, **23** (2013), 1091–1116. https://doi.org/10.5705/ss.2012.133

Supplementary

The whole results of the simulation studies are presented in the supplementary materials as follows. In addition, R codes for implementing the simulation study will be provided upon request.

Table S1. Variable selection results under n=200, p=20, and β_0 = $(4, 3, 2, 0.5, 1, 0.5, 0, ..., 0)^\top$.

ϵ	ρ	method	OF	CF	FPR	FNR	K0-M	K0-ME	K0-SD	K1-M	K1-ME	K1-SD
LN	0	PR-L	0.565	0.435	0.565	0.000	13.140	13.000	0.951	6.000	6.000	0.000
		LS-L	0.406	0.594	0.406	0.000	13.381	14.000	0.960	6.000	6.000	0.000
		LAD-L	0.997	0.003	0.997	0.000	9.279	9.000	1.835	6.000	6.000	0.000
		PR-aL	0.029	0.971	0.029	0.000	13.971	14.000	0.168	6.000	6.000	0.000
		LS-aL	0.408	0.592	0.408	0.000	13.359	14.000	0.997	6.000	6.000	0.000
		LAD-aL	0.525	0.475	0.525	0.000	12.706	13.000	1.679	6.000	6.000	0.000
	0.5	PR-L	0.301	0.699	0.301	0.000	13.631	14.000	0.624	6.000	6.000	0.000
		LS-L	0.136	0.864	0.136	0.000	13.831	14.000	0.476	6.000	6.000	0.000
		LAD-L	0.998	0.002	0.998	0.000	9.799	10.000	1.620	6.000	6.000	0.000
		PR-aL	0.016	0.966	0.016	0.018	13.984	14.000	0.126	5.979	6.000	0.163
		LS-aL	0.137	0.863	0.137	0.000	13.827	14.000	0.487	6.000	6.000	0.000
		LAD-aL	0.471	0.524	0.471	0.005	12.923	14.000	1.512	5.995	6.000	0.071
	-0.5	PR-L	0.816	0.120	0.857	0.064	11.995	12.000	1.432	5.928	6.000	0.288
		LS-L	0.813	0.157	0.837	0.030	12.043	12.000	1.523	5.968	6.000	0.187
		LAD-L	0.981	0.004	0.996	0.015	9.739	10.000	1.637	5.985	6.000	0.122
		PR-aL	0.060	0.837	0.065	0.103	13.924	14.000	0.307	5.886	6.000	0.351
		LS-aL	0.804	0.160	0.832	0.036	12.074	12.000	1.521	5.963	6.000	0.194
		LAD-aL	0.697	0.270	0.709	0.033	12.163	13.000	1.809	5.966	6.000	0.187
LU	0	PR-L	0.522	0.477	0.522	0.001	13.183	13.000	0.975	5.999	6.000	0.032
		LS-L	0.515	0.485	0.515	0.000	13.072	13.000	1.230	6.000	6.000	0.000
		LAD-L	0.994	0.001	0.999	0.005	8.903	9.000	1.761	5.995	6.000	0.071
		PR-aL	0.017	0.982	0.017	0.001	13.983	14.000	0.129	5.999	6.000	0.032
		LS-aL	0.529	0.471	0.529	0.000	13.083	13.000	1.178	6.000	6.000	0.000
		LAD-aL	0.593	0.352	0.603	0.055	12.257	13.000	2.073	5.941	6.000	0.252
	0.5	PR-L	0.290	0.709	0.290	0.001	13.664	14.000	0.565	5.999	6.000	0.032
		LS-L	0.231	0.767	0.231	0.002	13.680	14.000	0.707	5.998	6.000	0.045
		LAD-L	0.988	0.001	0.999	0.011	9.230	9.000	1.619	5.988	6.000	0.118
		PR-aL	0.005	0.979	0.005	0.016	13.995	14.000	0.071	5.984	6.000	0.126
		LS-aL	0.239	0.759	0.239	0.002	13.680	14.000	0.675	5.998	6.000	0.045
		LAD-aL	0.512	0.354	0.549	0.134	12.410	13.000	2.031	5.850	6.000	0.400
	-0.5	PR-L	0.799	0.145	0.836	0.056	12.081	12.000	1.368	5.940	6.000	0.254
		LS-L	0.839	0.100	0.887	0.061	11.602	12.000	1.651	5.939	6.000	0.239
		LAD-L	0.796	0.002	0.997	0.202	9.378	9.000	1.695	5.789	6.000	0.430
		PR-aL	0.058	0.847	0.060	0.095	13.934	14.000	0.275	5.898	6.000	0.325
		LS-aL	0.852	0.103	0.890	0.045	11.528	12.000	1.638	5.954	6.000	0.214
		LAD-aL	0.623	0.107	0.744	0.270	11.627	12.000	2.197	5.668	6.000	0.592
f_0	0	PR-L	0.162	0.838	0.162	0.000	13.827	14.000	0.411	6.000	6.000	0.000
		LS-L	0.309	0.691	0.309	0.000	13.565	14.000	0.786	6.000	6.000	0.000
		LAD-L	0.988	0.012	0.988	0.000	10.202	10.000	1.660	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	14.000	14.000	0.000	6.000	6.000	0.000
		LS-aL	0.299	0.701	0.299	0.000	13.582	14.000	0.764	6.000	6.000	0.000
		LAD-aL	0.356	0.644	0.356	0.000	13.308	14.000	1.154	6.000	6.000	0.000
	0.5	PR-L	0.056	0.944	0.056	0.000	13.944	14.000	0.230	6.000	6.000	0.000
		LS-L	0.121	0.879	0.121	0.000	13.857	14.000	0.430	6.000	6.000	0.000
		LAD-L	0.985	0.015	0.985	0.000	10.402	11.000	1.633	6.000	6.000	0.000
		PR-aL	0.000	0.999	0.000	0.001	14.000	14.000	0.000	5.999	6.000	0.032
		LS-aL	0.123	0.877	0.123	0.000	13.844	14.000	0.467	6.000	6.000	0.000
		LAD-aL	0.361	0.639	0.361	0.000	13.353	14.000	1.040	6.000	6.000	0.000
	-0.5	PR-L	0.569	0.430	0.569	0.001	13.172	13.000	0.894	5.999	6.000	0.032
		LS-L	0.779	0.221	0.779	0.000	12.361	13.000	1.369	6.000	6.000	0.000
		LAD-L	0.977	0.023	0.977	0.000	10.358	10.000	1.617	6.000	6.000	0.000
		PR-aL	0.000	0.996	0.000	0.004	14.000	14.000	0.000	5.996	6.000	0.063
		LS-aL	0.778	0.222	0.778	0.000	12.378	13.000	1.349	6.000	6.000	0.000
		LAD-aL	0.551	0.449	0.551	0.000	12.917	13.000	1.300	6.000	6.000	0.000

Table S2. Variable selection results under n=200, p=20, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	OF	CF	FPR	FNR	K0-M	K0-ME	K0-SD	K1-M	K1-ME	K1-SD
LN	0	PR-L	0.550	0.450	0.550	0.000	13.158	13.000	0.964	6.000	6.000	0.000
		LS-L	0.405	0.595	0.405	0.000	13.383	14.000	0.949	6.000	6.000	0.000
		LAD-L	0.997	0.003	0.997	0.000	9.279	9.000	1.835	6.000	6.000	0.000
		PR-aL	0.018	0.982	0.018	0.000	13.982	14.000	0.133	6.000	6.000	0.000
		LS-aL	0.408	0.592	0.408	0.000	13.359	14.000	0.996	6.000	6.000	0.000
		LAD-aL	0.387	0.613	0.387	0.000	13.015	14.000	1.619	6.000	6.000	0.000
	0.5	PR-L	0.306	0.694	0.306	0.000	13.622	14.000	0.638	6.000	6.000	0.000
		LS-L	0.130	0.870	0.130	0.000	13.837	14.000	0.472	6.000	6.000	0.000
		LAD-L	0.998	0.002	0.998	0.000	9.799	10.000	1.620	6.000	6.000	0.000
		PR-aL	0.011	0.989	0.011	0.000	13.989	14.000	0.104	6.000	6.000	0.000
		LS-aL	0.142	0.858	0.142	0.000	13.823	14.000	0.486	6.000	6.000	0.000
		LAD-aL	0.335	0.665	0.335	0.000	13.214	14.000	1.413	6.000	6.000	0.000
	-0.5	PR-L	0.889	0.111	0.889	0.000	11.875	12.000	1.427	6.000	6.000	0.000
		LS-L	0.838	0.162	0.838	0.000	12.041	12.000	1.516	6.000	6.000	0.000
		LAD-L	0.996	0.004	0.996	0.000	9.734	10.000	1.636	6.000	6.000	0.000
		PR-aL	0.017	0.983	0.017	0.000	13.981	14.000	0.151	6.000	6.000	0.000
		LS-aL	0.837	0.163	0.837	0.000	12.050	12.000	1.512	6.000	6.000	0.000
		LAD-aL	0.512	0.488	0.512	0.000	12.785	13.000	1.669	6.000	6.000	0.000
LU	0	PR-L	0.537	0.463	0.537	0.000	13.151	13.000	0.983	6.000	6.000	0.000
		LS-L	0.510	0.490	0.510	0.000	13.077	13.000	1.227	6.000	6.000	0.000
		LAD-L	0.999	0.001	0.999	0.000	8.903	9.000	1.763	6.000	6.000	0.000
		PR-aL	0.007	0.993	0.007	0.000	13.993	14.000	0.083	6.000	6.000	0.000
		LS-aL	0.536	0.464	0.536	0.000	13.061	13.000	1.199	6.000	6.000	0.000
		LAD-aL	0.420	0.580	0.420	0.000	12.870	14.000	1.865	6.000	6.000	0.000
	0.5	PR-L	0.289	0.711	0.289	0.000	13.654	14.000	0.590	6.000	6.000	0.000
		LS-L	0.228	0.772	0.228	0.000	13.682	14.000	0.700	6.000	6.000	0.000
		LAD-L	0.999	0.001	0.999	0.000	9.235	9.000	1.614	6.000	6.000	0.000
		PR-aL	0.003	0.997	0.003	0.000	13.997	14.000	0.055	6.000	6.000	0.000
		LS-aL	0.241	0.759	0.241	0.000	13.676	14.000	0.681	6.000	6.000	0.000
		LAD-aL	0.381	0.619	0.381	0.000	12.911	14.000	1.839	6.000	6.000	0.000
	-0.5	PR-L	0.862	0.138	0.862	0.000	11.978	12.000	1.364	6.000	6.000	0.000
		LS-L	0.891	0.109	0.891	0.000	11.573	12.000	1.632	6.000	6.000	0.000
		LAD-L	0.997	0.003	0.997	0.000	9.400	9.000	1.703	6.000	6.000	0.000
		PR-aL	0.008	0.992	0.008	0.000	13.992	14.000	0.089	6.000	6.000	0.000
		LS-aL	0.899	0.101	0.899	0.000	11.509	12.000	1.636	6.000	6.000	0.000
		LAD-aL	0.653	0.347	0.653	0.000	12.047	13.000	2.190	6.000	6.000	0.000
f_0	0	PR-L	0.165	0.835	0.165	0.000	13.825	14.000	0.408	6.000	6.000	0.000
		LS-L	0.316	0.684	0.316	0.000	13.559	14.000	0.779	6.000	6.000	0.000
		LAD-L	0.988	0.012	0.988	0.000	10.202	10.000	1.660	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	14.000	14.000	0.000	6.000	6.000	0.000
		LS-aL	0.303	0.697	0.303	0.000	13.584	14.000	0.763	6.000	6.000	0.000
		LAD-aL	0.284	0.716	0.284	0.000	13.415	14.000	1.140	6.000	6.000	0.000
	0.5	PR-L	0.067	0.933	0.067	0.000	13.933	14.000	0.250	6.000	6.000	0.000
		LS-L	0.124	0.876	0.124	0.000	13.855	14.000	0.422	6.000	6.000	0.000
		LAD-L	0.985	0.015	0.985	0.000	10.402	11.000	1.633	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	14.000	14.000	0.000	6.000	6.000	0.000
		LS-aL	0.126	0.874	0.126	0.000	13.836	14.000	0.489	6.000	6.000	0.000
		LAD-aL	0.277	0.723	0.277	0.000	13.468	14.000	1.029	6.000	6.000	0.000
	-0.5	PR-L	0.581	0.419	0.581	0.000	13.142	13.000	0.903	6.000	6.000	0.000
		LS-L	0.786	0.214	0.786	0.000	12.361	13.000	1.349	6.000	6.000	0.000
		LAD-L	0.977	0.023	0.977	0.000	10.358	10.000	1.617	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	14.000	14.000	0.000	6.000	6.000	0.000
		LS-aL	0.782	0.218	0.782	0.000	12.357	13.000	1.362	6.000	6.000	0.000
		LAD-aL	0.380	0.620	0.380	0.000	13.210	14.000	1.276	6.000	6.000	0.000

Table S3. Variable selection results under n=200, p=80, and β_0 = $(4, 3, 2, 0.5, 1, 0.5, 0, ..., 0)^\top$.

ϵ	ρ	method	OF	CF	FPR	FNR	K0-M	K0-ME	K0-SD	K1-M	K1-ME	K1-SD
LN	0	PR-L	0.734	0.265	0.734	0.001	72.529	73.000	1.294	5.999	6.000	0.032
		LS-L	0.633	0.367	0.633	0.000	72.394	73.000	2.142	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	59.713	60.000	3.233	6.000	6.000	0.000
		PR-aL	0.044	0.949	0.044	0.007	73.955	74.000	0.212	5.993	6.000	0.083
		LS-aL	0.622	0.378	0.622	0.000	72.432	73.000	2.033	6.000	6.000	0.000
		LAD-aL	0.770	0.218	0.776	0.012	70.480	72.000	4.106	5.988	6.000	0.109
	0.5	PR-L	0.378	0.620	0.379	0.002	73.488	74.000	0.766	5.998	6.000	0.045
		LS-L	0.210	0.789	0.210	0.001	73.625	74.000	0.967	5.999	6.000	0.032
		LAD-L	1.000	0.000	1.000	0.000	61.825	62.000	2.967	6.000	6.000	0.000
		PR-aL	0.044	0.932	0.044	0.024	73.953	74.000	0.225	5.976	6.000	0.153
		LS-aL	0.224	0.776	0.224	0.000	73.618	74.000	0.972	6.000	6.000	0.000
		LAD-aL	0.553	0.439	0.554	0.008	71.795	73.000	3.270	5.992	6.000	0.089
	-0.5	PR-L	0.624	0.012	0.959	0.364	70.222	70.000	2.229	5.587	6.000	0.584
		LS-L	0.855	0.008	0.981	0.137	68.257	69.000	3.619	5.859	6.000	0.360
		LAD-L	0.899	0.000	1.000	0.101	61.492	62.000	3.043	5.896	6.000	0.315
		PR-aL	0.128	0.418	0.186	0.454	73.764	74.000	0.563	5.476	6.000	0.624
		LS-aL	0.849	0.013	0.976	0.138	68.203	69.000	3.811	5.857	6.000	0.364
		LAD-aL	0.766	0.024	0.930	0.210	67.661	69.000	5.209	5.774	6.000	0.455
LU	0	PR-L	0.676	0.321	0.678	0.003	72.719	73.000	1.265	5.997	6.000	0.055
		LS-L	0.737	0.259	0.738	0.004	71.767	73.000	2.549	5.996	6.000	0.063
		LAD-L	0.990	0.000	1.000	0.010	58.141	58.000	3.327	5.990	6.000	0.100
		PR-aL	0.025	0.969	0.026	0.006	73.973	74.000	0.168	5.994	6.000	0.077
		LS-aL	0.742	0.254	0.743	0.004	71.790	73.000	2.434	5.996	6.000	0.063
		LAD-aL	0.733	0.157	0.789	0.110	69.668	72.000	5.104	5.884	6.000	0.339
	0.5	PR-L	0.348	0.652	0.348	0.000	73.532	74.000	0.749	6.000	6.000	0.000
		LS-L	0.369	0.630	0.369	0.001	73.281	74.000	1.366	5.999	6.000	0.032
		LAD-L	0.989	0.000	1.000	0.011	60.265	60.000	3.020	5.989	6.000	0.104
		PR-aL	0.023	0.957	0.023	0.020	73.976	74.000	0.160	5.979	6.000	0.150
		LS-aL	0.361	0.639	0.361	0.000	73.248	74.000	1.645	6.000	6.000	0.000
		LAD-aL	0.598	0.305	0.620	0.097	70.973	73.000	4.386	5.899	6.000	0.314
	-0.5	PR-L	0.596	0.016	0.946	0.388	70.491	71.000	2.154	5.557	6.000	0.598
		LS-L	0.695	0.002	0.982	0.303	67.327	68.000	4.020	5.675	6.000	0.513
		LAD-L	0.496	0.000	1.000	0.504	60.111	60.000	3.045	5.454	5.000	0.576
		PR-aL	0.121	0.410	0.196	0.469	73.769	74.000	0.510	5.460	6.000	0.625
		LS-aL	0.702	0.004	0.979	0.294	67.375	68.000	3.990	5.684	6.000	0.510
		LAD-aL	0.410	0.008	0.860	0.582	67.850	70.000	6.123	5.238	5.000	0.746
f_0	0	PR-L	0.267	0.733	0.267	0.000	73.699	74.000	0.532	6.000	6.000	0.000
		LS-L	0.551	0.449	0.551	0.000	72.886	73.000	1.481	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	62.657	63.000	3.041	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	74.000	74.000	0.000	6.000	6.000	0.000
		LS-aL	0.535	0.465	0.535	0.000	72.909	73.000	1.454	6.000	6.000	0.000
		LAD-aL	0.592	0.408	0.592	0.000	72.094	73.000	2.605	6.000	6.000	0.000
	0.5	PR-L	0.055	0.945	0.055	0.000	73.942	74.000	0.250	6.000	6.000	0.000
		LS-L	0.180	0.820	0.180	0.000	73.690	74.000	0.905	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	63.868	64.000	2.845	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	74.000	74.000	0.000	6.000	6.000	0.000
		LS-aL	0.178	0.822	0.178	0.000	73.695	74.000	0.910	6.000	6.000	0.000
		LAD-aL	0.457	0.543	0.457	0.000	72.626	74.000	2.273	6.000	6.000	0.000
	-0.5	PR-L	0.739	0.204	0.766	0.057	72.507	73.000	1.224	5.943	6.000	0.232
		LS-L	0.969	0.031	0.969	0.000	69.385	70.000	2.945	6.000	6.000	0.000
		LAD-L	0.999	0.000	1.000	0.001	63.743	64.000	2.892	5.999	6.000	0.032
		PR-aL	0.000	0.914	0.000	0.086	74.000	74.000	0.000	5.913	6.000	0.286
		LS-aL	0.970	0.030	0.970	0.000	69.349	70.000	3.085	6.000	6.000	0.000
		LAD-aL	0.872	0.118	0.879	0.010	69.702	71.000	3.840	5.990	6.000	0.100

Table S4. Variable selection results under n=200, p=80, and β_0 = $(4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	OF	CF	FPR	FNR	K0-M	K0-ME	K0-SD	K1-M	K1-ME	K1-SD
LN	0	PR-L	0.728	0.272	0.728	0.000	72.553	73.000	1.293	6.000	6.000	0.000
		LS-L	0.631	0.369	0.631	0.000	72.359	73.000	2.166	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	59.713	60.000	3.233	6.000	6.000	0.000
		PR-aL	0.031	0.969	0.031	0.000	73.968	74.000	0.182	6.000	6.000	0.000
		LS-aL	0.624	0.376	0.624	0.000	72.438	73.000	2.016	6.000	6.000	0.000
		LAD-aL	0.567	0.433	0.567	0.000	71.609	73.000	3.858	6.000	6.000	0.000
	0.5	PR-L	0.374	0.626	0.374	0.000	73.499	74.000	0.751	6.000	6.000	0.000
		LS-L	0.211	0.789	0.211	0.000	73.627	74.000	0.964	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	61.825	62.000	2.967	6.000	6.000	0.000
		PR-aL	0.037	0.963	0.037	0.000	73.961	74.000	0.204	6.000	6.000	0.000
		LS-aL	0.218	0.782	0.218	0.000	73.626	74.000	0.969	6.000	6.000	0.000
		LAD-aL	0.398	0.602	0.398	0.000	72.328	74.000	3.131	6.000	6.000	0.000
	-0.5	PR-L	0.984	0.016	0.984	0.000	69.383	70.000	2.383	6.000	6.000	0.000
		LS-L	0.988	0.012	0.988	0.000	68.049	69.000	3.573	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	61.492	62.000	3.046	6.000	6.000	0.000
		PR-aL	0.033	0.967	0.033	0.000	73.962	74.000	0.225	6.000	6.000	0.000
		LS-aL	0.985	0.015	0.985	0.000	67.951	69.000	3.784	6.000	6.000	0.000
		LAD-aL	0.915	0.079	0.921	0.006	68.593	70.000	5.072	5.994	6.000	0.077
LU	0	PR-L	0.703	0.297	0.703	0.000	72.711	73.000	1.222	6.000	6.000	0.000
		LS-L	0.735	0.265	0.735	0.000	71.780	73.000	2.548	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	58.142	58.000	3.323	6.000	6.000	0.000
		PR-aL	0.015	0.985	0.015	0.000	73.984	74.000	0.133	6.000	6.000	0.000
		LS-aL	0.739	0.261	0.739	0.000	71.801	73.000	2.408	6.000	6.000	0.000
		LAD-aL	0.584	0.416	0.584	0.000	71.237	73.000	4.455	6.000	6.000	0.000
	0.5	PR-L	0.365	0.635	0.365	0.000	73.496	74.000	0.785	6.000	6.000	0.000
		LS-L	0.366	0.634	0.366	0.000	73.286	74.000	1.372	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	60.263	60.000	3.019	6.000	6.000	0.000
		PR-aL	0.016	0.984	0.016	0.000	73.983	74.000	0.137	6.000	6.000	0.000
		LS-aL	0.357	0.643	0.357	0.000	73.259	74.000	1.629	6.000	6.000	0.000
		LAD-aL	0.441	0.559	0.441	0.000	71.776	74.000	4.141	6.000	6.000	0.000
	-0.5	PR-L	0.979	0.021	0.979	0.000	69.632	70.000	2.321	6.000	6.000	0.000
		LS-L	0.993	0.007	0.993	0.000	66.808	67.000	3.911	6.000	6.000	0.000
		LAD-L	0.999	0.000	1.000	0.001	60.132	60.000	3.036	5.999	6.000	0.032
		PR-aL	0.021	0.979	0.021	0.000	73.979	74.000	0.143	6.000	6.000	0.000
		LS-aL	0.993	0.007	0.993	0.000	66.800	67.000	3.888	6.000	6.000	0.000
		LAD-aL	0.920	0.037	0.953	0.043	66.391	68.000	6.294	5.957	6.000	0.203
f_0	0	PR-L	0.273	0.727	0.273	0.000	73.691	74.000	0.536	6.000	6.000	0.000
		LS-L	0.558	0.442	0.558	0.000	72.874	73.000	1.487	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	62.657	63.000	3.041	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	74.000	74.000	0.000	6.000	6.000	0.000
		LS-aL	0.543	0.457	0.543	0.000	72.905	73.000	1.450	6.000	6.000	0.000
		LAD-aL	0.447	0.553	0.447	0.000	72.421	74.000	2.750	6.000	6.000	0.000
	0.5	PR-L	0.056	0.944	0.056	0.000	73.943	74.000	0.236	6.000	6.000	0.000
		LS-L	0.168	0.832	0.168	0.000	73.697	74.000	0.950	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	63.868	64.000	2.845	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	74.000	74.000	0.000	6.000	6.000	0.000
		LS-aL	0.176	0.824	0.176	0.000	73.690	74.000	0.926	6.000	6.000	0.000
		LAD-aL	0.330	0.670	0.330	0.000	72.874	74.000	2.329	6.000	6.000	0.000
	-0.5	PR-L	0.802	0.198	0.802	0.000	72.379	73.000	1.262	6.000	6.000	0.000
		LS-L	0.971	0.029	0.971	0.000	69.422	70.000	2.905	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	63.743	64.000	2.892	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	74.000	74.000	0.000	6.000	6.000	0.000
		LS-aL	0.968	0.032	0.968	0.000	69.365	70.000	3.035	6.000	6.000	0.000
		LAD-aL	0.724	0.276	0.724	0.000	71.179	72.000	3.416	6.000	6.000	0.000

Table S5. Variable selection results under n=400, p=20, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	OF	CF	FPR	FNR	K0-M	K0-ME	K0-SD	K1-M	K1-ME	K1-SD
LN	0	PR-L	0.492	0.508	0.492	0.000	13.326	14.000	0.801	6.000	6.000	0.000
		LS-L	0.162	0.838	0.162	0.000	13.782	14.000	0.573	6.000	6.000	0.000
		LAD-L	0.994	0.006	0.994	0.000	10.042	10.000	1.656	6.000	6.000	0.000
		PR-aL	0.006	0.994	0.006	0.000	13.994	14.000	0.077	6.000	6.000	0.000
		LS-aL	0.147	0.853	0.147	0.000	13.806	14.000	0.528	6.000	6.000	0.000
		LAD-aL	0.284	0.716	0.284	0.000	13.485	14.000	0.990	6.000	6.000	0.000
	0.5	PR-L	0.278	0.722	0.278	0.000	13.660	14.000	0.602	6.000	6.000	0.000
		LS-L	0.061	0.939	0.061	0.000	13.936	14.000	0.257	6.000	6.000	0.000
		LAD-L	0.991	0.009	0.991	0.000	10.252	10.000	1.636	6.000	6.000	0.000
		PR-aL	0.006	0.994	0.006	0.000	13.994	14.000	0.077	6.000	6.000	0.000
		LS-aL	0.052	0.948	0.052	0.000	13.941	14.000	0.275	6.000	6.000	0.000
		LAD-aL	0.256	0.744	0.256	0.000	13.558	14.000	0.886	6.000	6.000	0.000
	-0.5	PR-L	0.826	0.174	0.826	0.000	12.327	12.000	1.242	6.000	6.000	0.000
		LS-L	0.680	0.320	0.680	0.000	12.879	13.000	1.082	6.000	6.000	0.000
		LAD-L	0.989	0.011	0.989	0.000	10.225	10.000	1.548	6.000	6.000	0.000
		PR-aL	0.009	0.991	0.009	0.000	13.989	14.000	0.122	6.000	6.000	0.000
		LS-aL	0.667	0.333	0.667	0.000	12.874	13.000	1.119	6.000	6.000	0.000
		LAD-aL	0.383	0.617	0.383	0.000	13.309	14.000	1.099	6.000	6.000	0.000
LU	0	PR-L	0.499	0.501	0.499	0.000	13.283	14.000	0.864	6.000	6.000	0.000
		LS-L	0.354	0.646	0.354	0.000	13.465	14.000	0.869	6.000	6.000	0.000
		LAD-L	0.998	0.002	0.998	0.000	9.174	9.000	1.714	6.000	6.000	0.000
		PR-aL	0.003	0.997	0.003	0.000	13.997	14.000	0.055	6.000	6.000	0.000
		LS-aL	0.364	0.636	0.364	0.000	13.440	14.000	0.912	6.000	6.000	0.000
		LAD-aL	0.359	0.641	0.359	0.000	13.094	14.000	1.539	6.000	6.000	0.000
	0.5	PR-L	0.286	0.714	0.286	0.000	13.665	14.000	0.579	6.000	6.000	0.000
		LS-L	0.119	0.881	0.119	0.000	13.859	14.000	0.411	6.000	6.000	0.000
		LAD-L	0.997	0.003	0.997	0.000	9.691	10.000	1.704	6.000	6.000	0.000
		PR-aL	0.002	0.998	0.002	0.000	13.998	14.000	0.045	6.000	6.000	0.000
		LS-aL	0.114	0.886	0.114	0.000	13.859	14.000	0.435	6.000	6.000	0.000
		LAD-aL	0.341	0.659	0.341	0.000	13.210	14.000	1.367	6.000	6.000	0.000
	-0.5	PR-L	0.798	0.202	0.798	0.000	12.399	13.000	1.250	6.000	6.000	0.000
		LS-L	0.789	0.211	0.789	0.000	12.311	12.000	1.370	6.000	6.000	0.000
		LAD-L	0.999	0.001	0.999	0.000	9.557	10.000	1.681	6.000	6.000	0.000
		PR-aL	0.002	0.998	0.002	0.000	13.998	14.000	0.045	6.000	6.000	0.000
		LS-aL	0.790	0.210	0.790	0.000	12.331	13.000	1.402	6.000	6.000	0.000
		LAD-aL	0.501	0.499	0.501	0.000	12.829	13.000	1.624	6.000	6.000	0.000
f_0	0	PR-L	0.145	0.855	0.145	0.000	13.850	14.000	0.371	6.000	6.000	0.000
		LS-L	0.146	0.854	0.146	0.000	13.826	14.000	0.458	6.000	6.000	0.000
		LAD-L	0.972	0.028	0.972	0.000	10.833	11.000	1.557	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	14.000	14.000	0.000	6.000	6.000	0.000
		LS-aL	0.140	0.860	0.140	0.000	13.838	14.000	0.431	6.000	6.000	0.000
		LAD-aL	0.212	0.787	0.212	0.001	13.700	14.000	0.656	5.999	6.000	0.032
	0.5	PR-L	0.052	0.948	0.052	0.000	13.946	14.000	0.235	6.000	6.000	0.000
		LS-L	0.042	0.958	0.042	0.000	13.954	14.000	0.228	6.000	6.000	0.000
		LAD-L	0.962	0.038	0.962	0.000	11.017	11.000	1.537	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	14.000	14.000	0.000	6.000	6.000	0.000
		LS-aL	0.039	0.961	0.039	0.000	13.955	14.000	0.247	6.000	6.000	0.000
		LAD-aL	0.179	0.821	0.179	0.000	13.758	14.000	0.576	6.000	6.000	0.000
	-0.5	PR-L	0.459	0.541	0.459	0.000	13.447	14.000	0.674	6.000	6.000	0.000
		LS-L	0.518	0.482	0.518	0.000	13.247	13.000	0.918	6.000	6.000	0.000
		LAD-L	0.957	0.043	0.957	0.000	11.093	11.000	1.477	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	14.000	14.000	0.000	6.000	6.000	0.000
		LS-aL	0.520	0.480	0.520	0.000	13.234	13.000	0.928	6.000	6.000	0.000
		LAD-aL	0.252	0.748	0.252	0.000	13.615	14.000	0.759	6.000	6.000	0.000

Table S6. Variable selection results under n=400, p=80, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	OF	CF	FPR	FNR	K0-M	K0-ME	K0-SD	K1-M	K1-ME	K1-SD
LN	0	PR-L	0.635	0.365	0.635	0.000	72.931	73.000	1.086	6.000	6.000	0.000
		LS-L	0.336	0.664	0.336	0.000	73.437	74.000	1.054	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	61.476	62.000	3.260	6.000	6.000	0.000
		PR-aL	0.011	0.989	0.011	0.000	73.988	74.000	0.118	6.000	6.000	0.000
		LS-aL	0.345	0.655	0.345	0.000	73.452	74.000	0.986	6.000	6.000	0.000
		LAD-aL	0.359	0.641	0.359	0.000	72.832	74.000	2.270	6.000	6.000	0.000
	0.5	PR-L	0.302	0.698	0.302	0.000	73.607	74.000	0.669	6.000	6.000	0.000
		LS-L	0.083	0.917	0.083	0.000	73.877	74.000	0.490	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	63.492	64.000	2.949	6.000	6.000	0.000
		PR-aL	0.010	0.990	0.010	0.000	73.989	74.000	0.114	6.000	6.000	0.000
		LS-aL	0.086	0.914	0.086	0.000	73.873	74.000	0.536	6.000	6.000	0.000
		LAD-aL	0.335	0.665	0.335	0.000	72.861	74.000	2.327	6.000	6.000	0.000
	-0.5	PR-L	0.938	0.062	0.938	0.000	70.906	71.000	1.873	6.000	6.000	0.000
		LS-L	0.914	0.086	0.914	0.000	70.925	71.000	2.381	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	63.152	63.000	2.993	6.000	6.000	0.000
		PR-aL	0.009	0.991	0.009	0.000	73.991	74.000	0.094	6.000	6.000	0.000
		LS-aL	0.903	0.097	0.903	0.000	70.950	71.000	2.414	6.000	6.000	0.000
		LAD-aL	0.584	0.416	0.584	0.000	72.003	73.000	2.918	6.000	6.000	0.000
LU	0	PR-L	0.603	0.397	0.603	0.000	73.027	73.000	1.038	6.000	6.000	0.000
		LS-L	0.535	0.465	0.535	0.000	72.799	73.000	1.717	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	59.043	59.000	3.354	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	74.000	74.000	0.000	6.000	6.000	0.000
		LS-aL	0.542	0.458	0.542	0.000	72.824	73.000	1.630	6.000	6.000	0.000
		LAD-aL	0.415	0.585	0.415	0.000	72.282	74.000	3.296	6.000	6.000	0.000
	0.5	PR-L	0.272	0.728	0.272	0.000	73.679	74.000	0.569	6.000	6.000	0.000
		LS-L	0.169	0.831	0.169	0.000	73.733	74.000	0.746	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	61.036	61.000	3.131	6.000	6.000	0.000
		PR-aL	0.003	0.997	0.003	0.000	73.997	74.000	0.055	6.000	6.000	0.000
		LS-aL	0.176	0.824	0.176	0.000	73.733	74.000	0.716	6.000	6.000	0.000
		LAD-aL	0.364	0.636	0.364	0.000	72.548	74.000	2.812	6.000	6.000	0.000
	-0.5	PR-L	0.909	0.091	0.909	0.000	71.194	71.000	1.800	6.000	6.000	0.000
		LS-L	0.961	0.039	0.961	0.000	69.535	70.000	2.963	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	61.149	61.000	3.074	6.000	6.000	0.000
		PR-aL	0.009	0.991	0.009	0.000	73.991	74.000	0.094	6.000	6.000	0.000
		LS-aL	0.960	0.040	0.960	0.000	69.540	70.000	2.924	6.000	6.000	0.000
		LAD-aL	0.738	0.262	0.738	0.000	70.907	72.000	3.689	6.000	6.000	0.000
f_0	0	PR-L	0.151	0.849	0.151	0.000	73.833	74.000	0.414	6.000	6.000	0.000
		LS-L	0.245	0.755	0.245	0.000	73.650	74.000	0.728	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	64.871	65.000	2.742	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	74.000	74.000	0.000	6.000	6.000	0.000
		LS-aL	0.250	0.750	0.250	0.000	73.652	74.000	0.710	6.000	6.000	0.000
		LAD-aL	0.319	0.681	0.319	0.000	73.224	74.000	1.485	6.000	6.000	0.000
	0.5	PR-L	0.033	0.967	0.033	0.000	73.967	74.000	0.179	6.000	6.000	0.000
		LS-L	0.050	0.950	0.050	0.000	73.941	74.000	0.279	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	66.192	66.000	2.595	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	74.000	74.000	0.000	6.000	6.000	0.000
		LS-aL	0.048	0.952	0.048	0.000	73.936	74.000	0.349	6.000	6.000	0.000
		LAD-aL	0.262	0.738	0.262	0.000	73.394	74.000	1.255	6.000	6.000	0.000
	-0.5	PR-L	0.624	0.376	0.624	0.000	73.037	73.000	0.958	6.000	6.000	0.000
		LS-L	0.815	0.185	0.815	0.000	72.017	72.000	1.686	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	66.002	66.000	2.509	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	74.000	74.000	0.000	6.000	6.000	0.000
		LS-aL	0.818	0.182	0.818	0.000	71.946	72.000	1.784	6.000	6.000	0.000
		LAD-aL	0.428	0.572	0.428	0.000	72.904	74.000	1.782	6.000	6.000	0.000

Table S7. Variable selection results under n=100, p=200, and β_0 = $(4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	OF	CF	FPR	FNR	K0-M	K0-ME	K0-SD	K1-M	K1-ME	K1-SD
LN	0	PR-L	0.849	0.151	0.849	0.000	191.762	192.000	1.804	6.000	6.000	0.000
		LS-L	0.907	0.093	0.907	0.000	189.603	190.500	4.117	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	170.228	170.000	3.718	6.000	6.000	0.000
		PR-aL	0.008	0.992	0.008	0.000	193.991	194.000	0.105	6.000	6.000	0.000
		LS-aL	0.908	0.092	0.908	0.000	189.463	191.000	4.582	6.000	6.000	0.000
		LAD-aL	0.932	0.032	0.968	0.036	186.783	188.000	4.853	5.964	6.000	0.186
	0.5	PR-L	0.360	0.640	0.360	0.000	193.518	194.000	0.756	6.000	6.000	0.000
		LS-L	0.442	0.558	0.442	0.000	192.770	194.000	2.636	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	172.930	173.000	3.612	6.000	6.000	0.000
		PR-aL	0.008	0.992	0.008	0.000	193.991	194.000	0.105	6.000	6.000	0.000
		LS-aL	0.427	0.573	0.427	0.000	192.871	194.000	2.296	6.000	6.000	0.000
		LAD-aL	0.582	0.418	0.582	0.000	191.810	193.000	2.898	6.000	6.000	0.000
	-0.5	PR-L	0.516	0.000	1.000	0.484	185.581	186.000	3.427	5.470	6.000	0.591
		LS-L	0.877	0.000	1.000	0.123	178.192	179.000	6.972	5.876	6.000	0.333
		LAD-L	0.763	0.000	1.000	0.237	171.459	172.000	3.613	5.756	6.000	0.446
		PR-aL	0.094	0.391	0.199	0.515	193.704	194.000	0.696	5.426	5.000	0.614
		LS-aL	0.878	0.000	0.999	0.122	178.340	179.000	7.139	5.875	6.000	0.340
		LAD-aL	0.395	0.000	0.991	0.605	184.281	185.000	4.022	5.279	5.000	0.687
LU	0	PR-L	0.844	0.156	0.844	0.000	191.944	192.000	1.551	6.000	6.000	0.000
		LS-L	0.925	0.075	0.925	0.000	188.250	190.000	6.081	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	168.552	169.000	3.752	6.000	6.000	0.000
		PR-aL	0.008	0.992	0.008	0.000	193.992	194.000	0.089	6.000	6.000	0.000
		LS-aL	0.919	0.081	0.919	0.000	188.527	190.000	5.639	6.000	6.000	0.000
		LAD-aL	0.903	0.041	0.956	0.056	186.165	187.000	5.491	5.943	6.000	0.236
	0.5	PR-L	0.318	0.682	0.318	0.000	193.601	194.000	0.650	6.000	6.000	0.000
		LS-L	0.543	0.457	0.543	0.000	192.023	193.000	3.464	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	171.218	171.000	3.547	6.000	6.000	0.000
		PR-aL	0.000	0.998	0.000	0.002	194.000	194.000	0.000	5.998	6.000	0.045
		LS-aL	0.541	0.459	0.541	0.000	192.042	193.000	3.608	6.000	6.000	0.000
		LAD-aL	0.624	0.376	0.624	0.000	191.335	193.000	3.385	6.000	6.000	0.000
	-0.5	PR-L	0.436	0.000	1.000	0.564	185.289	185.000	3.413	5.350	5.000	0.651
		LS-L	0.750	0.000	0.999	0.250	178.268	179.000	7.520	5.738	6.000	0.466
		LAD-L	0.549	0.000	1.000	0.451	170.496	171.000	3.537	5.528	6.000	0.540
		PR-aL	0.080	0.331	0.221	0.589	193.688	194.000	0.682	5.308	5.000	0.671
		LS-aL	0.750	0.000	0.999	0.250	178.298	179.000	7.710	5.735	6.000	0.476
		LAD-aL	0.256	0.000	0.984	0.744	183.818	184.000	4.799	5.073	5.000	0.720
f_0	0	PR-L	0.446	0.554	0.446	0.000	193.416	194.000	0.773	6.000	6.000	0.000
		LS-L	0.880	0.120	0.880	0.000	190.107	191.000	3.691	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	173.736	174.000	3.653	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	194.000	194.000	0.000	6.000	6.000	0.000
		LS-aL	0.877	0.123	0.877	0.000	190.178	191.000	3.766	6.000	6.000	0.000
		LAD-aL	0.915	0.064	0.936	0.021	187.980	189.000	4.261	5.979	6.000	0.143
	0.5	PR-L	0.056	0.944	0.056	0.000	193.941	194.000	0.248	6.000	6.000	0.000
		LS-L	0.403	0.597	0.403	0.000	193.112	194.000	1.640	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	176.122	176.000	3.385	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	194.000	194.000	0.000	6.000	6.000	0.000
		LS-aL	0.419	0.581	0.419	0.000	193.025	194.000	1.823	6.000	6.000	0.000
		LAD-aL	0.515	0.485	0.515	0.000	192.316	193.000	2.366	6.000	6.000	0.000
	-0.5	PR-L	0.844	0.004	0.994	0.152	188.684	189.000	2.749	5.837	6.000	0.406
		LS-L	0.999	0.000	1.000	0.001	179.996	181.000	6.144	5.999	6.000	0.032
		LAD-L	0.970	0.000	1.000	0.030	174.034	174.000	3.813	5.969	6.000	0.179
		PR-aL	0.010	0.832	0.020	0.158	193.979	194.000	0.150	5.830	6.000	0.414
		LS-aL	0.998	0.000	1.000	0.002	180.022	181.000	6.164	5.998	6.000	0.045
		LAD-aL	0.534	0.001	0.997	0.465	185.470	186.000	3.518	5.478	6.000	0.628

Table S8. Variable selection results under n=100, p=400, and β_0 = $(4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	OF	CF	FPR	FNR	K0-M	K0-ME	K0-SD	K1-M	K1-ME	K1-SD
LN	0	PR-L	0.902	0.096	0.904	0.002	391.019	391.000	2.174	5.998	6.000	0.045
		LS-L	0.943	0.057	0.943	0.000	388.348	390.000	5.207	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	362.897	363.000	3.957	6.000	6.000	0.000
		PR-aL	0.009	0.989	0.009	0.002	393.990	394.000	0.109	5.998	6.000	0.045
		LS-aL	0.938	0.062	0.938	0.000	388.150	390.000	5.555	6.000	6.000	0.000
		LAD-aL	0.874	0.014	0.983	0.112	386.293	387.000	4.756	5.885	6.000	0.328
	0.5	PR-L	0.403	0.597	0.403	0.000	393.422	394.000	0.842	6.000	6.000	0.000
		LS-L	0.504	0.496	0.504	0.000	392.396	393.000	2.934	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	366.082	366.000	3.893	6.000	6.000	0.000
		PR-aL	0.010	0.990	0.010	0.000	393.990	394.000	0.100	6.000	6.000	0.000
		LS-aL	0.513	0.487	0.513	0.000	392.337	393.000	3.163	6.000	6.000	0.000
		LAD-aL	0.614	0.386	0.614	0.000	391.833	393.000	2.599	6.000	6.000	0.000
	-0.5	PR-L	0.205	0.000	1.000	0.795	383.312	384.000	3.935	4.967	5.000	0.728
		LS-L	0.577	0.000	0.999	0.423	375.810	377.000	8.957	5.548	6.000	0.562
		LAD-L	0.485	0.000	1.000	0.515	362.964	363.000	4.013	5.440	5.000	0.585
		PR-aL	0.063	0.119	0.340	0.818	393.416	394.000	1.067	4.889	5.000	0.766
		LS-aL	0.559	0.000	0.999	0.441	375.518	377.000	9.164	5.537	6.000	0.545
		LAD-aL	0.158	0.000	0.969	0.842	383.627	384.000	5.577	4.431	5.000	1.103
LU	0	PR-L	0.911	0.087	0.912	0.002	391.054	391.000	1.998	5.998	6.000	0.045
		LS-L	0.963	0.037	0.963	0.000	387.341	389.000	6.027	6.000	6.000	0.000
		LAD-L	0.999	0.000	1.000	0.001	360.917	361.000	3.944	5.999	6.000	0.032
		PR-aL	0.012	0.983	0.013	0.005	393.985	394.000	0.144	5.995	6.000	0.071
		LS-aL	0.956	0.044	0.956	0.000	387.454	389.000	5.602	6.000	6.000	0.000
		LAD-aL	0.816	0.009	0.984	0.175	385.561	386.000	5.188	5.818	6.000	0.404
	0.5	PR-L	0.350	0.650	0.350	0.000	393.529	394.000	0.735	6.000	6.000	0.000
		LS-L	0.568	0.432	0.568	0.000	391.832	393.000	3.954	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	363.796	364.000	3.885	6.000	6.000	0.000
		PR-aL	0.005	0.993	0.005	0.002	393.995	394.000	0.071	5.998	6.000	0.045
		LS-aL	0.565	0.435	0.565	0.000	391.669	393.000	4.320	6.000	6.000	0.000
		LAD-aL	0.670	0.330	0.670	0.000	391.068	392.000	3.388	6.000	6.000	0.000
	-0.5	PR-L	0.165	0.000	1.000	0.835	383.203	383.000	3.945	4.828	5.000	0.791
		LS-L	0.408	0.000	0.998	0.592	376.511	378.000	9.084	5.304	5.000	0.669
		LAD-L	0.301	0.000	1.000	0.699	361.625	362.000	4.005	5.178	5.000	0.636
		PR-aL	0.043	0.096	0.397	0.861	393.310	394.000	1.129	4.749	5.000	0.810
		LS-aL	0.418	0.000	0.996	0.582	376.352	378.000	9.614	5.304	5.000	0.691
		LAD-aL	0.069	0.000	0.956	0.931	383.545	384.000	6.317	4.151	4.000	1.086
f_0	0	PR-L	0.548	0.452	0.548	0.000	393.141	393.000	0.994	6.000	6.000	0.000
		LS-L	0.921	0.079	0.921	0.000	389.329	390.000	4.302	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	367.741	368.000	3.893	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	394.000	394.000	0.000	6.000	6.000	0.000
		LS-aL	0.917	0.083	0.917	0.000	389.417	390.000	4.371	6.000	6.000	0.000
		LAD-aL	0.893	0.026	0.973	0.081	387.869	388.000	3.778	5.917	6.000	0.283
	0.5	PR-L	0.054	0.946	0.054	0.000	393.943	394.000	0.245	6.000	6.000	0.000
		LS-L	0.418	0.582	0.418	0.000	392.928	394.000	2.014	6.000	6.000	0.000
		LAD-L	1.000	0.000	1.000	0.000	370.431	370.000	3.708	6.000	6.000	0.000
		PR-aL	0.000	1.000	0.000	0.000	394.000	394.000	0.000	6.000	6.000	0.000
		LS-aL	0.427	0.573	0.427	0.000	392.919	394.000	1.987	6.000	6.000	0.000
		LAD-aL	0.581	0.419	0.581	0.000	392.238	393.000	2.157	6.000	6.000	0.000
	-0.5	PR-L	0.597	0.002	0.995	0.401	386.031	386.000	3.450	5.535	6.000	0.637
		LS-L	0.964	0.000	1.000	0.036	374.789	376.000	7.265	5.963	6.000	0.194
		LAD-L	0.826	0.000	1.000	0.174	365.830	366.000	4.395	5.820	6.000	0.400
		PR-aL	0.016	0.572	0.052	0.412	393.929	394.000	0.338	5.507	6.000	0.673
		LS-aL	0.964	0.000	1.000	0.036	374.730	375.500	7.414	5.964	6.000	0.186
		LAD-aL	0.242	0.000	0.980	0.758	384.581	385.000	4.658	4.648	5.000	1.100

Table S9. Estimation results under n=200, p=20, and $\beta_0 = (4, 3, 2, 0.5, 1, 0.5, 0, ..., 0)^\top$.

ϵ	ρ	method	N0	N1	Mbias0	Mbias1	MSE	MREE
LN	0	PR-L	13.140	2.774	0.211	0.404	0.102	0.104
		LS-L	13.381	1.756	0.178	0.430	0.150	0.155
		LAD-L	9.279	3.645	0.284	0.416	0.103	0.101
		PR-aL	13.971	4.761	0.267	0.341	0.033	0.033
		LS-aL	13.359	1.760	0.188	0.481	0.152	0.155
		LAD-aL	12.706	4.205	0.316	0.391	0.065	0.065
	0.5	PR-L	13.631	3.833	0.226	0.489	0.065	0.068
		LS-L	13.831	3.143	0.200	0.409	0.097	0.141
		LAD-L	9.799	3.623	0.308	0.440	0.103	0.079
		PR-aL	13.984	4.172	0.260	0.500	0.049	0.033
		LS-aL	13.827	3.131	0.200	0.424	0.098	0.142
		LAD-aL	12.923	3.537	0.305	0.500	0.091	0.062
	-0.5	PR-L	11.995	1.180	0.283	0.747	0.299	0.141
		LS-L	12.043	0.658	0.220	0.598	0.403	0.185
		LAD-L	9.739	1.553	0.318	0.699	0.279	0.149
		PR-aL	13.924	3.617	0.337	0.591	0.071	0.043
		LS-aL	12.074	0.640	0.235	0.629	0.404	0.186
		LAD-aL	12.163	3.193	0.370	0.562	0.114	0.071
LU	0	PR-L	13.183	2.726	0.231	0.500	0.104	0.098
		LS-L	13.072	1.738	0.177	0.459	0.171	0.158
		LAD-L	8.903	2.421	0.436	0.547	0.233	0.230
		PR-aL	13.983	4.815	0.234	0.500	0.032	0.033
		LS-aL	13.083	1.712	0.196	0.440	0.170	0.158
		LAD-aL	12.257	2.995	0.471	0.537	0.168	0.172
	0.5	PR-L	13.664	3.711	0.270	0.500	0.068	0.073
		LS-L	13.680	2.923	0.257	0.531	0.111	0.145
		LAD-L	9.230	2.520	0.511	0.750	0.240	0.198
		PR-aL	13.995	4.171	0.314	0.500	0.047	0.035
		LS-aL	13.680	2.962	0.267	0.500	0.113	0.145
		LAD-aL	12.410	2.436	0.507	0.660	0.239	0.182
	-0.5	PR-L	12.081	1.152	0.259	0.613	0.287	0.114
		LS-L	11.602	0.658	0.258	0.668	0.441	0.164
		LAD-L	9.378	0.962	0.449	0.845	0.619	0.276
		PR-aL	13.934	3.924	0.283	0.590	0.061	0.036
		LS-aL	11.528	0.681	0.258	0.682	0.434	0.162
		LAD-aL	11.627	2.077	0.593	1.000	0.344	0.201
f_0	0	PR-L	13.827	3.989	0.120	0.255	0.049	0.050
		LS-L	13.565	3.647	0.102	0.260	0.057	0.058
		LAD-L	10.202	4.843	0.234	0.275	0.043	0.043
		PR-aL	14.000	5.729	0.000	0.225	0.013	0.012
		LS-aL	13.582	3.593	0.111	0.259	0.057	0.060
		LAD-aL	13.308	5.307	0.198	0.348	0.025	0.024
	0.5	PR-L	13.944	5.045	0.139	0.283	0.027	0.026
		LS-L	13.857	4.603	0.107	0.270	0.036	0.044
		LAD-L	10.402	4.799	0.277	0.374	0.044	0.036
		PR-aL	14.000	5.340	0.000	0.500	0.019	0.013
		LS-aL	13.844	4.593	0.117	0.285	0.037	0.043
		LAD-aL	13.353	4.828	0.270	0.381	0.033	0.023
	-0.5	PR-L	13.172	1.440	0.145	0.599	0.166	0.064
		LS-L	12.361	1.592	0.131	0.397	0.143	0.055
		LAD-L	10.358	2.568	0.224	0.386	0.119	0.057
		PR-aL	14.000	5.053	0.000	0.500	0.025	0.015
		LS-aL	12.378	1.577	0.126	0.372	0.146	0.056
		LAD-aL	12.917	4.630	0.234	0.373	0.042	0.026

Table S10. Estimation results under n=200, p=20, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	N0	N1	Mbias0	Mbias1	MSE	MREE
LN	0	PR-L	13.158	2.689	0.213	0.437	0.106	0.107
		LS-L	13.383	1.752	0.178	0.432	0.151	0.155
		LAD-L	9.279	3.645	0.284	0.416	0.103	0.101
		PR-aL	13.982	4.908	0.273	0.269	0.031	0.031
		LS-aL	13.359	1.740	0.188	0.485	0.152	0.156
		LAD-aL	13.015	4.322	0.309	0.360	0.056	0.055
	0.5	PR-L	13.622	3.816	0.230	0.476	0.067	0.069
		LS-L	13.837	3.141	0.203	0.419	0.097	0.141
		LAD-L	9.799	3.623	0.308	0.440	0.103	0.079
		PR-aL	13.989	4.298	0.257	0.370	0.044	0.029
		LS-aL	13.823	3.132	0.203	0.419	0.098	0.141
		LAD-aL	13.214	3.780	0.294	0.465	0.075	0.054
	-0.5	PR-L	11.875	1.194	0.293	0.670	0.299	0.141
		LS-L	12.041	0.655	0.227	0.630	0.403	0.187
		LAD-L	9.734	1.554	0.318	0.786	0.279	0.149
		PR-aL	13.981	4.171	0.295	0.365	0.048	0.030
		LS-aL	12.050	0.657	0.233	0.624	0.408	0.187
		LAD-aL	12.785	3.352	0.377	0.468	0.097	0.061
LU	0	PR-L	13.151	2.692	0.234	0.409	0.107	0.102
		LS-L	13.077	1.724	0.181	0.456	0.172	0.158
		LAD-L	8.903	2.420	0.436	0.547	0.233	0.230
		PR-aL	13.993	4.853	0.236	0.340	0.030	0.031
		LS-aL	13.061	1.702	0.200	0.445	0.170	0.157
		LAD-aL	12.870	3.070	0.479	0.561	0.141	0.143
	0.5	PR-L	13.654	3.664	0.274	0.518	0.070	0.075
		LS-L	13.682	2.932	0.253	0.592	0.111	0.146
		LAD-L	9.235	2.518	0.524	0.750	0.240	0.198
		PR-aL	13.997	4.182	0.265	0.397	0.047	0.036
		LS-aL	13.676	2.961	0.264	0.632	0.112	0.144
		LAD-aL	12.911	2.637	0.515	0.685	0.189	0.142
	-0.5	PR-L	11.978	1.154	0.255	0.612	0.293	0.116
		LS-L	11.573	0.666	0.259	0.674	0.440	0.162
		LAD-L	9.400	0.962	0.445	0.847	0.625	0.280
		PR-aL	13.992	4.402	0.296	0.377	0.042	0.028
		LS-aL	11.509	0.684	0.263	0.672	0.437	0.163
		LAD-aL	12.047	2.361	0.603	0.900	0.254	0.158
f_0	0	PR-L	13.825	3.906	0.124	0.270	0.051	0.052
		LS-L	13.559	3.655	0.103	0.263	0.057	0.058
		LAD-L	10.202	4.843	0.234	0.275	0.043	0.043
		PR-aL	14.000	5.835	0.000	0.167	0.011	0.010
		LS-aL	13.584	3.591	0.108	0.259	0.057	0.059
		LAD-aL	13.415	5.382	0.195	0.236	0.023	0.022
	0.5	PR-L	13.933	4.961	0.145	0.278	0.028	0.027
		LS-L	13.855	4.596	0.108	0.271	0.036	0.044
		LAD-L	10.402	4.799	0.277	0.374	0.044	0.036
		PR-aL	14.000	5.535	0.000	0.253	0.016	0.011
		LS-aL	13.836	4.591	0.118	0.285	0.037	0.043
		LAD-aL	13.468	4.996	0.272	0.310	0.030	0.021
	-0.5	PR-L	13.142	1.372	0.136	0.430	0.173	0.067
		LS-L	12.361	1.592	0.131	0.405	0.143	0.055
		LAD-L	10.358	2.568	0.224	0.386	0.119	0.057
		PR-aL	14.000	5.556	0.000	0.225	0.016	0.011
		LS-aL	12.357	1.600	0.125	0.364	0.146	0.055
		LAD-aL	13.210	4.768	0.236	0.317	0.037	0.023

Table S11. Estimation results under n=200, p=80, and $\beta_0 = (4, 3, 2, 0.5, 1, 0.5, 0, ..., 0)^\top$.

ϵ	ρ	method	N0	N1	Mbias0	Mbias1	MSE	MREE
LN	0	PR-L	72.529	1.473	0.200	0.524	0.185	0.181
		LS-L	72.394	1.122	0.174	0.495	0.210	0.203
		LAD-L	59.713	2.846	0.343	0.438	0.189	0.182
		PR-aL	73.955	4.676	0.293	0.500	0.037	0.036
		LS-aL	72.432	1.118	0.172	0.508	0.209	0.207
		LAD-aL	70.480	3.956	0.285	0.500	0.089	0.084
	0.5	PR-L	73.488	3.348	0.180	0.500	0.091	0.098
		LS-L	73.625	2.864	0.141	0.500	0.115	0.143
		LAD-L	61.825	3.502	0.326	0.439	0.148	0.133
		PR-aL	73.953	4.138	0.279	0.500	0.052	0.032
		LS-aL	73.618	2.891	0.147	0.449	0.113	0.140
		LAD-aL	71.795	3.569	0.359	0.500	0.095	0.066
	-0.5	PR-L	70.222	0.134	0.284	0.949	0.801	0.327
		LS-L	68.257	0.198	0.258	0.786	0.661	0.273
		LAD-L	61.492	0.523	0.312	0.742	0.609	0.290
		PR-aL	73.764	2.834	0.338	0.780	0.212	0.116
		LS-aL	68.203	0.194	0.258	0.786	0.671	0.278
		LAD-aL	67.661	2.499	0.372	0.744	0.247	0.155
LU	0	PR-L	72.719	1.400	0.191	0.561	0.191	0.199
		LS-L	71.767	1.068	0.226	0.510	0.251	0.260
		LAD-L	58.141	1.967	0.454	0.577	0.398	0.396
		PR-aL	73.973	4.861	0.248	0.500	0.032	0.031
		LS-aL	71.790	1.093	0.213	0.534	0.246	0.254
		LAD-aL	69.668	2.699	0.446	0.516	0.229	0.226
	0.5	PR-L	73.532	3.385	0.209	0.439	0.089	0.107
		LS-L	73.281	2.718	0.253	0.500	0.134	0.175
		LAD-L	60.265	2.433	0.466	0.581	0.336	0.303
		PR-aL	73.976	4.238	0.242	0.500	0.048	0.032
		LS-aL	73.248	2.738	0.274	0.472	0.132	0.174
		LAD-aL	70.973	2.493	0.462	0.584	0.244	0.181
	-0.5	PR-L	70.491	0.127	0.251	0.960	0.868	0.331
		LS-L	67.327	0.184	0.287	1.053	0.829	0.329
		LAD-L	60.111	0.328	0.428	1.163	1.117	0.536
		PR-aL	73.769	2.975	0.305	0.708	0.224	0.113
		LS-aL	67.375	0.191	0.282	0.942	0.817	0.327
		LAD-aL	67.850	1.546	0.560	1.094	0.641	0.326
f_0	0	PR-L	73.699	2.701	0.126	0.313	0.081	0.071
		LS-L	72.886	2.883	0.141	0.276	0.076	0.066
		LAD-L	62.657	4.230	0.227	0.303	0.078	0.073
		PR-aL	74.000	5.675	0.000	0.215	0.014	0.013
		LS-aL	72.909	2.856	0.135	0.287	0.077	0.067
		LAD-aL	72.094	5.219	0.200	0.256	0.031	0.029
	0.5	PR-L	73.942	4.717	0.108	0.302	0.034	0.042
		LS-L	73.690	4.411	0.105	0.280	0.041	0.055
		LAD-L	63.868	4.655	0.200	0.314	0.065	0.058
		PR-aL	74.000	5.316	0.000	0.273	0.021	0.013
		LS-aL	73.695	4.405	0.099	0.298	0.042	0.055
		LAD-aL	72.626	4.792	0.215	0.304	0.036	0.026
	-0.5	PR-L	72.507	0.194	0.142	0.684	0.410	0.163
		LS-L	69.385	0.652	0.133	0.474	0.246	0.104
		LAD-L	63.743	1.011	0.216	0.535	0.259	0.126
		PR-aL	74.000	4.530	0.000	0.627	0.036	0.020
		LS-aL	69.349	0.658	0.134	0.440	0.246	0.105
		LAD-aL	69.702	4.100	0.235	0.500	0.070	0.046

Table S12. Estimation results under n=200, p=80, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	N0	N1	Mbias0	Mbias1	MSE	MREE
LN	0	PR-L	72.553	1.439	0.200	0.523	0.188	0.184
		LS-L	72.359	1.122	0.174	0.502	0.211	0.204
		LAD-L	59.713	2.846	0.343	0.438	0.189	0.182
		PR-aL	73.968	4.859	0.300	0.323	0.032	0.030
		LS-aL	72.438	1.110	0.174	0.517	0.208	0.204
		LAD-aL	71.609	4.161	0.281	0.398	0.067	0.064
	0.5	PR-L	73.499	3.357	0.180	0.467	0.092	0.099
		LS-L	73.627	2.875	0.142	0.482	0.116	0.145
		LAD-L	61.825	3.502	0.326	0.439	0.148	0.133
		PR-aL	73.961	4.302	0.286	0.387	0.045	0.029
		LS-aL	73.626	2.894	0.151	0.455	0.113	0.141
		LAD-aL	72.328	3.796	0.383	0.450	0.077	0.054
	-0.5	PR-L	69.383	0.159	0.284	0.857	0.746	0.302
		LS-L	68.049	0.206	0.259	0.752	0.660	0.272
		LAD-L	61.492	0.524	0.312	0.789	0.611	0.293
		PR-aL	73.962	4.194	0.318	0.444	0.049	0.032
		LS-aL	67.951	0.201	0.259	0.785	0.662	0.272
		LAD-aL	68.593	2.907	0.370	1.000	0.162	0.100
LU	0	PR-L	72.711	1.384	0.191	0.559	0.195	0.203
		LS-L	71.780	1.068	0.225	0.518	0.252	0.258
		LAD-L	58.142	1.968	0.454	0.664	0.398	0.396
		PR-aL	73.984	4.931	0.257	0.316	0.030	0.029
		LS-aL	71.801	1.096	0.212	0.539	0.245	0.255
		LAD-aL	71.237	2.943	0.447	0.604	0.165	0.160
	0.5	PR-L	73.496	3.341	0.212	0.463	0.089	0.108
		LS-L	73.286	2.741	0.255	0.491	0.133	0.175
		LAD-L	60.263	2.433	0.466	0.601	0.336	0.303
		PR-aL	73.983	4.246	0.236	0.345	0.047	0.032
		LS-aL	73.259	2.730	0.266	0.467	0.132	0.174
		LAD-aL	71.776	2.631	0.481	0.587	0.188	0.137
	-0.5	PR-L	69.632	0.141	0.257	0.839	0.793	0.314
		LS-L	66.808	0.203	0.296	0.918	0.809	0.323
		LAD-L	60.132	0.318	0.427	1.155	1.222	0.573
		PR-aL	73.979	4.401	0.244	0.380	0.042	0.027
		LS-aL	66.800	0.210	0.285	0.918	0.800	0.322
		LAD-aL	66.391	1.927	0.611	1.000	0.446	0.268
f_0	0	PR-L	73.691	2.640	0.134	0.341	0.083	0.073
		LS-L	72.874	2.900	0.137	0.271	0.075	0.066
		LAD-L	62.657	4.230	0.227	0.303	0.078	0.073
		PR-aL	74.000	5.838	0.000	0.163	0.011	0.010
		LS-aL	72.905	2.872	0.131	0.291	0.076	0.066
		LAD-aL	72.421	5.370	0.198	0.265	0.026	0.025
	0.5	PR-L	73.943	4.669	0.112	0.321	0.036	0.043
		LS-L	73.697	4.404	0.106	0.273	0.042	0.055
		LAD-L	63.868	4.655	0.200	0.314	0.065	0.058
		PR-aL	74.000	5.547	0.000	0.217	0.016	0.011
		LS-aL	73.690	4.399	0.101	0.290	0.042	0.056
		LAD-aL	72.874	4.971	0.200	0.292	0.032	0.023
	-0.5	PR-L	72.379	0.183	0.147	0.639	0.412	0.163
		LS-L	69.422	0.634	0.130	0.464	0.248	0.105
		LAD-L	63.743	1.011	0.216	0.535	0.259	0.126
		PR-aL	74.000	5.554	0.000	0.266	0.016	0.011
		LS-aL	69.365	0.645	0.133	0.439	0.248	0.105
		LAD-aL	71.179	4.347	0.202	0.414	0.053	0.035

Table S13. Estimation results under n=400, p=20, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	N0	N1	Mbias0	Mbias1	MSE	MREE
LN	0	PR-L	13.326	4.085	0.149	0.313	0.048	0.050
		LS-L	13.782	2.410	0.116	0.342	0.093	0.097
		LAD-L	10.042	4.719	0.216	0.297	0.048	0.049
		PR-aL	13.994	5.683	0.171	0.190	0.015	0.015
		LS-aL	13.806	2.331	0.101	0.321	0.094	0.098
		LAD-aL	13.485	5.273	0.212	0.265	0.024	0.025
	0.5	PR-L	13.660	4.819	0.174	0.274	0.033	0.034
		LS-L	13.936	3.859	0.093	0.304	0.059	0.083
		LAD-L	10.252	4.639	0.339	0.328	0.049	0.039
		PR-aL	13.994	5.261	0.146	0.252	0.021	0.015
		LS-aL	13.941	3.854	0.119	0.310	0.059	0.083
		LAD-aL	13.558	4.821	0.249	0.310	0.034	0.024
	-0.5	PR-L	12.327	1.917	0.226	0.466	0.146	0.060
		LS-L	12.879	0.669	0.170	0.459	0.258	0.093
		LAD-L	10.225	2.312	0.255	0.429	0.142	0.067
		PR-aL	13.989	5.203	0.165	0.253	0.022	0.015
		LS-aL	12.874	0.666	0.177	0.471	0.256	0.093
		LAD-aL	13.309	4.507	0.228	0.377	0.042	0.028
LU	0	PR-L	13.283	3.914	0.156	0.311	0.055	0.055
		LS-L	13.465	2.519	0.144	0.331	0.097	0.097
		LAD-L	9.174	3.483	0.303	0.427	0.116	0.115
		PR-aL	13.997	5.668	0.151	0.206	0.015	0.015
		LS-aL	13.440	2.531	0.130	0.342	0.096	0.098
		LAD-aL	13.094	4.099	0.341	0.379	0.067	0.066
	0.5	PR-L	13.665	4.649	0.168	0.280	0.037	0.038
		LS-L	13.859	3.732	0.121	0.379	0.064	0.087
		LAD-L	9.691	3.401	0.331	0.468	0.124	0.097
		PR-aL	13.998	5.073	0.177	0.268	0.025	0.016
		LS-aL	13.859	3.738	0.131	0.327	0.063	0.085
		LAD-aL	13.210	3.498	0.362	0.508	0.093	0.062
	-0.5	PR-L	12.399	1.696	0.198	0.453	0.158	0.071
		LS-L	12.311	0.855	0.183	0.497	0.251	0.109
		LAD-L	9.557	1.418	0.335	0.635	0.313	0.161
		PR-aL	13.998	5.274	0.167	0.282	0.021	0.014
		LS-aL	12.331	0.822	0.186	0.520	0.253	0.108
		LAD-aL	12.829	3.231	0.339	0.634	0.110	0.070
f_0	0	PR-L	13.850	5.466	0.121	0.191	0.022	0.022
		LS-L	13.826	4.983	0.094	0.185	0.033	0.033
		LAD-L	10.833	5.719	0.148	0.190	0.021	0.021
		PR-aL	14.000	5.992	0.000	0.123	0.005	0.005
		LS-aL	13.838	5.040	0.105	0.185	0.033	0.033
		LAD-aL	13.700	5.902	0.129	1.000	0.010	0.010
	0.5	PR-L	13.946	5.770	0.113	0.200	0.013	0.013
		LS-L	13.954	5.424	0.081	0.209	0.021	0.028
		LAD-L	11.017	5.598	0.163	0.219	0.022	0.018
		PR-aL	14.000	5.943	0.000	0.159	0.008	0.005
		LS-aL	13.955	5.426	0.086	0.204	0.021	0.028
		LAD-aL	13.758	5.705	0.146	0.193	0.014	0.010
	-0.5	PR-L	13.447	2.637	0.115	0.336	0.080	0.034
		LS-L	13.247	2.331	0.105	0.294	0.088	0.037
		LAD-L	11.093	3.827	0.133	0.313	0.059	0.029
		PR-aL	14.000	5.923	0.000	0.150	0.008	0.005
		LS-aL	13.234	2.279	0.099	0.282	0.089	0.037
		LAD-aL	13.615	5.557	0.121	0.232	0.016	0.010

Table S14. Estimation results under n=400, p=80, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	N0	N1	Mbias0	Mbias1	MSE	MREE
LN	0	PR-L	72.931	2.798	0.126	0.313	0.080	0.076
		LS-L	73.437	1.752	0.119	0.328	0.115	0.107
		LAD-L	61.476	4.146	0.214	0.282	0.088	0.086
		PR-aL	73.988	5.698	0.139	0.208	0.014	0.014
		LS-aL	73.452	1.725	0.103	0.346	0.115	0.107
		LAD-aL	72.832	5.276	0.213	0.268	0.027	0.027
	0.5	PR-L	73.607	4.463	0.133	0.292	0.041	0.049
		LS-L	73.877	3.699	0.092	0.337	0.063	0.092
		LAD-L	63.492	4.552	0.219	0.366	0.072	0.067
		PR-aL	73.989	5.268	0.185	0.232	0.021	0.014
		LS-aL	73.873	3.678	0.101	0.352	0.063	0.091
		LAD-aL	72.861	4.810	0.201	0.307	0.037	0.026
	-0.5	PR-L	70.906	0.430	0.156	0.579	0.334	0.146
		LS-L	70.925	0.298	0.143	0.527	0.372	0.160
		LAD-L	63.152	1.027	0.258	0.526	0.270	0.139
		PR-aL	73.991	5.171	0.135	0.261	0.023	0.015
		LS-aL	70.950	0.282	0.138	0.547	0.369	0.162
		LAD-aL	72.003	4.271	0.254	0.373	0.057	0.035
LU	0	PR-L	73.027	2.519	0.136	0.305	0.093	0.091
		LS-L	72.799	1.671	0.136	0.371	0.132	0.130
		LAD-L	59.043	2.728	0.299	0.434	0.207	0.201
		PR-aL	74.000	5.636	0.000	0.244	0.015	0.015
		LS-aL	72.824	1.661	0.130	0.372	0.133	0.129
		LAD-aL	72.282	4.015	0.315	0.412	0.072	0.071
	0.5	PR-L	73.679	4.333	0.163	0.318	0.045	0.056
		LS-L	73.733	3.473	0.138	0.369	0.074	0.107
		LAD-L	61.036	3.277	0.317	0.470	0.172	0.159
		PR-aL	73.997	5.130	0.139	0.270	0.024	0.017
		LS-aL	73.733	3.489	0.126	0.401	0.072	0.108
		LAD-aL	72.548	3.554	0.319	0.508	0.091	0.068
	-0.5	PR-L	71.194	0.352	0.182	0.621	0.374	0.146
		LS-L	69.535	0.306	0.191	0.656	0.419	0.165
		LAD-L	61.149	0.538	0.390	0.783	0.606	0.289
		PR-aL	73.991	5.324	0.167	0.255	0.021	0.013
		LS-aL	69.540	0.335	0.191	0.689	0.415	0.163
		LAD-aL	70.907	2.905	0.380	0.576	0.155	0.094
f_0	0	PR-L	73.833	4.699	0.085	0.197	0.037	0.035
		LS-L	73.650	4.363	0.066	0.197	0.044	0.041
		LAD-L	64.871	5.417	0.139	0.193	0.038	0.036
		PR-aL	74.000	5.992	0.000	0.112	0.005	0.005
		LS-aL	73.652	4.381	0.068	0.209	0.044	0.041
		LAD-aL	73.224	5.897	0.138	0.160	0.011	0.011
	0.5	PR-L	73.967	5.653	0.085	0.186	0.016	0.019
		LS-L	73.941	5.298	0.064	0.224	0.024	0.035
		LAD-L	66.192	5.484	0.155	0.230	0.031	0.027
		PR-aL	74.000	5.927	0.000	0.162	0.008	0.005
		LS-aL	73.936	5.307	0.062	0.224	0.024	0.035
		LAD-aL	73.394	5.717	0.131	0.216	0.015	0.011
	-0.5	PR-L	73.037	0.909	0.096	0.390	0.167	0.074
		LS-L	72.017	1.274	0.102	0.330	0.134	0.060
		LAD-L	66.002	2.290	0.166	0.337	0.115	0.060
		PR-aL	74.000	5.936	0.000	0.140	0.008	0.005
		LS-aL	71.946	1.254	0.104	0.330	0.133	0.060
		LAD-aL	72.904	5.478	0.153	0.243	0.020	0.013

Table S15. Estimation results under n=100, p=200, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	N0	N1	Mbias0	Mbias1	MSE	MREE
LN	0	PR-L	191.762	0.369	0.301	1.092	0.737	0.753
		LS-L	189.603	0.513	0.332	0.938	0.548	0.566
		LAD-L	170.228	1.348	0.515	0.770	0.576	0.581
		PR-aL	193.991	3.751	0.428	0.592	0.068	0.065
		LS-aL	189.463	0.533	0.326	0.866	0.549	0.571
		LAD-aL	186.783	2.433	0.637	1.500	0.268	0.269
	0.5	PR-L	193.518	1.959	0.311	0.787	0.264	0.347
		LS-L	192.770	1.933	0.297	0.732	0.246	0.339
		LAD-L	172.930	2.330	0.453	0.723	0.385	0.370
		PR-aL	193.991	3.254	0.477	0.546	0.094	0.058
		LS-aL	192.871	1.935	0.297	0.796	0.251	0.345
		LAD-aL	191.810	2.758	0.543	0.819	0.173	0.128
	-0.5	PR-L	185.581	0.005	0.602	2.396	4.185	1.733
		LS-L	178.192	0.044	0.485	1.577	2.373	1.048
		LAD-L	171.459	0.075	0.649	2.074	3.160	1.485
		PR-aL	193.704	1.829	0.649	2.000	1.254	0.727
		LS-aL	178.340	0.045	0.450	2.000	2.363	1.050
		LAD-aL	184.281	0.333	0.863	3.000	2.505	1.187
LU	0	PR-L	191.944	0.343	0.358	0.892	0.720	0.631
		LS-L	188.250	0.467	0.377	0.837	0.695	0.599
		LAD-L	168.552	0.968	0.625	0.966	0.989	0.918
		PR-aL	193.992	3.957	0.473	0.606	0.058	0.053
		LS-aL	188.527	0.472	0.424	0.893	0.694	0.608
		LAD-aL	186.165	1.777	0.569	1.500	0.533	0.504
	0.5	PR-L	193.601	1.937	0.250	0.919	0.265	0.331
		LS-L	192.023	1.770	0.368	0.802	0.311	0.396
		LAD-L	171.218	1.763	0.635	0.926	0.716	0.689
		PR-aL	194.000	3.227	0.000	1.000	0.095	0.065
		LS-aL	192.042	1.744	0.361	0.840	0.315	0.397
		LAD-aL	191.335	1.882	0.591	0.950	0.386	0.292
	-0.5	PR-L	185.289	0.010	0.504	2.402	4.584	1.780
		LS-L	178.268	0.049	0.544	2.145	3.180	1.329
		LAD-L	170.496	0.058	0.923	2.594	4.521	2.063
		PR-aL	193.688	1.774	0.619	2.000	1.396	0.558
		LS-aL	178.298	0.047	0.576	2.000	3.196	1.335
		LAD-aL	183.818	0.264	0.865	3.000	3.340	1.502
f_0	0	PR-L	193.416	0.534	0.194	0.736	0.316	0.315
		LS-L	190.107	1.158	0.187	0.518	0.205	0.199
		LAD-L	173.736	2.246	0.339	0.465	0.241	0.237
		PR-aL	194.000	5.257	0.000	0.248	0.022	0.020
		LS-aL	190.178	1.172	0.188	0.518	0.205	0.200
		LAD-aL	187.980	3.589	0.376	1.500	0.103	0.099
	0.5	PR-L	193.941	2.991	0.167	0.530	0.102	0.133
		LS-L	193.112	3.097	0.183	0.413	0.096	0.121
		LAD-L	176.122	3.496	0.288	0.442	0.160	0.151
		PR-aL	194.000	4.710	0.000	0.324	0.034	0.022
		LS-aL	193.025	3.109	0.166	0.407	0.095	0.124
		LAD-aL	192.316	3.918	0.257	0.384	0.070	0.050
	-0.5	PR-L	188.684	0.012	0.385	2.000	2.080	0.863
		LS-L	179.996	0.126	0.302	1.000	0.879	0.400
		LAD-L	174.034	0.150	0.487	2.000	1.250	0.618
		PR-aL	193.979	3.855	0.330	2.000	0.050	0.030
		LS-aL	180.022	0.132	0.306	1.000	0.876	0.405
		LAD-aL	185.470	0.648	0.680	3.000	1.668	0.791

Table S16. Estimation results under n=100, p=400, and $\beta_0 = (4, 3, 2, 1.5, 1, 2.5, 0, ..., 0)^\top$.

ϵ	ρ	method	N0	N1	Mbias0	Mbias1	MSE	MREE
LN	0	PR-L	391.019	0.208	0.454	1.070	0.915	1.029
		LS-L	388.348	0.329	0.430	0.993	0.666	0.743
		LAD-L	362.897	1.033	0.467	0.976	0.712	0.756
		PR-aL	393.990	3.800	0.498	1.000	0.066	0.064
		LS-aL	388.150	0.337	0.424	0.904	0.656	0.737
		LAD-aL	386.293	1.786	0.705	1.500	0.382	0.418
	0.5	PR-L	393.422	1.784	0.232	0.855	0.298	0.495
		LS-L	392.396	1.861	0.298	0.715	0.274	0.451
		LAD-L	366.082	2.263	0.490	0.752	0.441	0.481
		PR-aL	393.990	3.251	0.517	0.494	0.092	0.066
		LS-aL	392.337	1.852	0.276	0.738	0.270	0.450
		LAD-aL	391.833	2.647	0.409	0.766	0.179	0.159
	-0.5	PR-L	383.312	0.001	0.586	3.000	6.283	2.724
		LS-L	375.810	0.013	0.598	2.497	4.026	1.844
		LAD-L	362.964	0.040	0.734	2.634	4.867	2.397
		PR-aL	393.416	1.270	0.879	3.000	1.681	0.812
		LS-aL	375.518	0.017	0.506	2.646	4.095	1.837
		LAD-aL	383.627	0.084	0.903	3.000	6.496	3.136
LU	0	PR-L	391.054	0.219	0.401	1.026	0.950	1.002
		LS-L	387.341	0.313	0.439	1.187	0.829	0.877
		LAD-L	360.917	0.760	0.612	1.000	1.179	1.192
		PR-aL	393.985	3.956	0.458	1.000	0.060	0.062
		LS-aL	387.454	0.313	0.429	1.082	0.843	0.886
		LAD-aL	385.561	1.285	0.682	2.000	0.799	0.842
	0.5	PR-L	393.529	1.814	0.243	1.163	0.306	0.464
		LS-L	391.832	1.705	0.339	0.836	0.359	0.531
		LAD-L	363.796	1.754	0.630	0.845	0.772	0.807
		PR-aL	393.995	3.234	0.445	1.000	0.094	0.065
		LS-aL	391.669	1.689	0.339	0.766	0.364	0.541
		LAD-aL	391.068	1.815	0.554	0.898	0.391	0.326
	-0.5	PR-L	383.203	0.001	0.707	3.000	7.140	3.757
		LS-L	376.511	0.004	0.536	3.000	5.343	2.846
		LAD-L	361.625	0.021	0.745	3.000	6.460	3.704
		PR-aL	393.310	1.174	0.715	3.000	1.831	0.800
		LS-aL	376.352	0.003	0.537	3.005	5.399	2.887
		LAD-aL	383.545	0.081	1.227	3.000	8.222	4.357
f_0	0	PR-L	393.141	0.320	0.212	0.852	0.411	0.393
		LS-L	389.329	0.789	0.220	0.537	0.249	0.237
		LAD-L	367.741	1.754	0.339	0.556	0.289	0.280
		PR-aL	394.000	5.255	0.000	0.220	0.022	0.022
		LS-aL	389.417	0.813	0.223	0.521	0.248	0.241
		LAD-aL	387.869	2.731	0.447	1.533	0.152	0.148
	0.5	PR-L	393.943	2.770	0.195	0.553	0.117	0.130
		LS-L	392.928	2.945	0.166	0.454	0.104	0.115
		LAD-L	370.431	3.376	0.279	0.433	0.181	0.170
		PR-aL	394.000	4.688	0.000	0.392	0.033	0.020
		LS-aL	392.919	2.940	0.159	0.474	0.104	0.116
		LAD-aL	392.238	3.828	0.257	0.451	0.070	0.049
	-0.5	PR-L	386.031	0.003	0.454	2.304	3.499	1.444
		LS-L	374.789	0.037	0.346	1.500	1.422	0.621
		LAD-L	365.830	0.070	0.619	2.248	2.242	1.055
		PR-aL	393.929	2.984	0.589	2.000	0.146	0.080
		LS-aL	374.730	0.047	0.361	1.485	1.416	0.619
		LAD-aL	384.581	0.123	0.896	3.000	4.637	1.940