A Fast Distributed Screening for High-dimensional Quantile Regression *

Haili Zhang [1], Yanlin Tang[2], Xuejun Jiang[1], Jianchen Jiang[3], Zhaobo Liu [4]
[1] Southern University of Science and Technology, Shenzhen, China
[2] East China Normal University
[3] University of North Carolina at Charlotte, NC 28223, USA
[4] Shenzhen University

April 10, 2023

Abstract

We address the challenge of handling large datasets with limited memory constraints and performing variable selection in the era of big data. A naive divide-and-conquer approach, which splits the whole data into L parts and runs each part separately on L machines, is commonly used to solve this problem. However, this approach tends to select more noisy variables, and the false discovery rate may not be well controlled as a result. To address these issues, a data-driven screening method for high-dimensional quantile regression in distributed data is proposed in this paper. The method effectively reduces dimensionality and improves computational efficiency by selecting true variables with probability tending to 1 when the number of variables is diverging. Furthermore, the paper demonstrates that the obtained parameter estimation errors are asymptotically normal. The effectiveness of the proposed method is evaluated using simulations and a real-world blog feedback example.

Keywords: Cusum monitoring, distributed algorithm, high-dimensional quantile regression, memory constraints, variable Screening.

1 Introduction

In the context of big data, the large volume of high-dimensional or ultra-high-dimensional (often referred to as high-dimensional) data needs to be distributed stored or distributed collected, which presents new opportunities and challenges for statistical analysis. The distributed efficient statistical analysis of high-dimensional data is becoming more and more essential. Quantile regression is a detailed description of the data characteristics of the response variables at each location with the change of covariates. Quantile regression method involves using the quantiles of the explanatory variables to obtain the quantile equations of the conditional distribution of the explanatory variables, thus providing global information about the response variables. Compared to the mean regression, quantile regression can capture the impact of covariates on the tail quantile of the response variable. For outliers, quantile regression is more robust than least square estimation. Quantile regression is more flexible and suitable for heavy tail error data,

*Corresponding authors: X. Jiang (jiangxj@sustech.edu.cn)

skewed distribution error data, and heteroscedastic data automatically. Therefore, distributed high-dimensional quantile regression has crucial theoretical and practical value in the era of big data.

Currently, studies on high-dimensional quantile regression are mainly based on penalized method, for example, L_1 penalty (Belloni and Chernozhukov, 2011 [3]), and seamless L_0 penalty (Gabrela, 2015) [8]; variable screening method (He et al., 2013, [14]); and some computing algorithms, such as semismooth newton algorithm (SNA), semismooth newton coordinate descent algorithm (SNCD, [24]), and alternating direction method of multipliers (ADMM, [12]). The computing algorithms sometimes have good performance in simulations, but convergence results of all these computing algorithms rely on a good initial point, which is usually not guaranteed in practice. For low dimensional problems we can use grid search method to ensure global convergence with an arbitrary initial point, but since grid search methods involve considerable amounts of function and gradient evaluations, they are not well-suited for high-dimensional cases. The penalized methods have several disadvantages: 1) the choice of tuning parameters. The selection of tuning parameters is uncertain, and the results of the estimations and predictions will be unreliable if they are not appropriately chosen. 2) In statistical inference for big data scenarios, corrections are needed for penalized methods, a large number of calculations are usually required in estimating the precision matrix, and the estimation results often have difficulty in achieving the required precision. The variable screening methods only consider the marginal correlation between covariates and response variables, and ignore the correlations in different covariates. However, the correlations between high-dimensional covariates can not be ignored in massive data. In addition, the models constructed by these two methods may not have oracle properties, and will lead to unknown asymptotic distribution of test statistics under the misspecified model, thus making subsequent statistical inferences difficult. Therefore, it is a new idea to develop a quantile model selection for distributed high-dimensional data without penalty. Based on this, a new projection method is proposed, which can help solve the statistical inference problem of quantile regression model for high-dimensional data.

In recent years, driven by national policy support and social development promotion, large data has been widely distributed in the economic, financial, meteorological, transportation and medical fields. Due to computer memory constraints, these huge datasets are too massive to be stored on a single computer, so they are distributed across multiple machines or servers, resulting in distributed data. At the same time, because of their bulky size, a personal computer cannot be utilized for analyzing entire data set. On the other hand, distributed data, such as from bank branches and hospitals in different cities, is generated naturally due to confidentiality, privacy, and technical limitations of the data. Therefore, the traditional methods of estimation and inference need to be revisited. Distributed data and related methods are rapidly becoming important issues in the big data backgrounds. It creates more innovative challenges and opportunities for data scientists [10]. The divide and conquer trick is a practicable and common approach. It divides data into several groups and then aggregate all subgroup estimators to lessen the computational burden and alleviate such bottlenecks, and can be implemented by parallel computing systems [26, 7, 25, 23]. For example, variable selection [7], statistical optimization [26], logistic regression [22], equation estimation [18], kernel ridge regression [25, 23], quantile regression [6], and distributed principal component analysis (PCA) [11, 1]. Some distributed statistical methods based on likelihood framework are also proposed, and the theoretical upper bound of the information loss for the distributed algorithm is obtained [c.f., 2].

Feature selection and parameter estimation are two basic problems in large-scale data problems. Under limited memory constraints, there are many researches on variable selection of general model, but few on variable selection. To tackle the quantile regression of distributed data, Chen et al. (2019) adapt a divide-and-conquer trick for fixed dimensional (or low dimensional) quantile model under memory constraints. A distributed iterative algorithm was proposed to improve the computational speed and estimation precision. For the high-dimensional distributed data, focusing on distributed fast computation, Chen et al. (2020) converted quantile regression into linear regression, and proposed an efficient distributed iterative algorithm for both computing and communication. The new algorithm simply transmited gradient information on each machine or server at each iteration. Chen et al. (2020) also demonstrated that the new algorithm can achieve the near-oracle convergence rate with limited iterating shots.

However, they are not devoted to model selection, resulting in uncertainty in post-model-selection estimators, which will lead to inflated family-wise error rate. In fact, getting an asymptotic distribution of their estimators is a challenge because it is a complex iterative and aggregated estimator. Moreover, the extremely large datasets can not be handled by a single computer due to bottlenecks in processors, memory or disks, so it is important to develop efficient algorithms to minimize computing overhead. A fast parallel algorithm with high computation speed, and low computation cost for high-dimensional quantile regression, is suitable and attractive for distributed data. It improves the efficiency of calculation and is conducive to generalization of statistical inference. Our work aims to resolve the uncertainty of the high-dimensional quantile regression model, improve the prediction accuracy, promote the development of the data-analysis method of the high-dimensional distributed data, further excavate the high-dimensional data information, and provide decision support for the practical application. Therefore, we consider the variable selection problem under memory constraints, and will illustrate that it can be possible to design algorithms with statistical accuracy guaranteed.

We take the divide-and-conquer approach to split the data. Usually the arithmetic average of estimators from every machines can be used as a nice estimator of parameter, but it does not work well in our situation, since the features selected from each machine can be different and averaging them may lead to too many features selected. To solve this problem, we propose a distributed method. Naturally, several important questions must be answered for the proposed approach. For example, does it bring us a computationally efficient algorithm? Is the proposed estimator statistically efficient relative to the quantile estimator if there is no limited memory constraint? Our answers are yes. In the following we briefly highlight our contributions:

(1) For a general high-dimensional quantile regression model, we solve the limited memory constraint problem. Our results extend those of Wang and Leng (2016) under more relaxed assumptions, which we do not require that the random error has mean 0 and q-exponential tails. Our method also acts as a fast solver for quantile loss on a single machine and thus

is more efficient in computation.

(2) Compared with the Chen et al.(2020), our algorithm directly uses the data driven screening estimator and avoids large additional computational burdens on fitting the selected model.

(3) Our method can be easily extend to high-dimensional quantile varying coefficient models for distributed data.

The rest of this paper is organized as follows. In Section 2, we introduce the distributed projection quantile (DPQ) algorithm for high-dimensional quantile regression model under memory constraints. Theoretic properties of the proposed estimator are investigated in Section 3. In Section 4, we examine finite sample performance of our method through simulation studies. In Section 5, we apply the proposed method to the song data. Our findings and conclusions are summarized in Section 6. The detailed proofs are relegated to Appendix.

2 Methodology

2.1 A HOLP type estimator for high-dimensional quantile regression

In this section, we illustrate our method and present our efficient algorithm. Suppose there is an i.i.d. sample $\{(Y_i, \mathbf{X}_i), i = 1, \ldots, N\}$ from the population $(Y, \mathbf{X})$, where Y is a scaler response variable, and $\mathbf{X}$ is a p-dimensional covariate vector, with $p \gg N$. Given a quantile level $0 < \tau < 1$, we assume that the conditional τ-quantile of Y_i given $\mathbf{X}_i$ is given by

$$Q_\tau(Y_i \mid \mathbf{X}_i) = \mathbf{X}_i^\top \boldsymbol{\beta}^*(\tau) \triangleq \mu_i, \quad i = 1, \ldots, N, \tag{1}$$

where $\boldsymbol{\beta}^*(\tau)$ is an unknown true quantile slope vector. Let

$$\varepsilon_i(\tau) = Y_i - Q_\tau(Y_i \mid \mathbf{X}_i) = Y_i - \mathbf{X}_i^\top \boldsymbol{\beta}^*(\tau), \tag{2}$$

so that $Q_\tau\{\varepsilon_i(\tau) \mid \mathbf{X}_i\} = 0$. Let $F_{i,\tau}(\cdot)$ and $f_{i,\tau}(\cdot)$ be the cumulative distribution function and probability density function of $\varepsilon_i(\tau)$ given $\mathbf{X}_i$. Since we focus on i.i.d sample, so $F_{i,\tau}(\cdot) \equiv F_\tau(\cdot)$ and $f_{i,\tau}(\cdot) = f_\tau(\cdot)$ for some $F_\tau(\cdot)$ and $f_\tau(\cdot)$. In this high-dimensional setting, a common estimator for $\boldsymbol{\beta}^*(\tau)$ is the L_1-QR estimator of $\boldsymbol{\beta}^*(\tau)$ [3] :

$$\widehat{\beta}_{\lambda_N}(\tau) = \arg\min_{\beta} \sum_{i=1}^{N} \rho_\tau\left(Y_i - X_i^\top \beta\right) + \lambda_N \|\beta\|_1, \tag{3}$$

where $\lambda_N = 2Q_{0.9}(\Lambda_N|X)$, $\Lambda_N = \|\{\tau - I(u \leq \tau)\}^T \mathbb{X}\|_\infty$, $u = (u_1, \ldots, u_N)^T$, and u_i i.i.d. from the uniform distribution $U(0, 1)$.

However, when the N and dimension number p are massive, we could not use (3) with the whole sample on a single machine due to the memory constraint. Noting the linearity and easy parallel computing properties of least squares estimation and the idea of high-dimensional ordinary least squares projection (HOLP) of Wang and Leng (2016)[20], we proposed a HOLP type estimator for high-dimensional quantile regression. The procedure is shown as follow.

For model (1) with fixed dimension, the standard QR estimator is

$$\widehat{\boldsymbol{\beta}}_{fix}(\tau) = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau \left(Y_i - \mathbf{X}_i^\top \boldsymbol{\beta} \right),$$

and we have

$$\widehat{\boldsymbol{\beta}}(\tau) = (\mathbb{X}^\top \mathbb{X})^+ \mathbb{X}^\top \hat{\mathbf{Y}}(\tau),$$

where $\hat{\mathbf{Y}}(\tau) = \mathbb{X}\widehat{\boldsymbol{\beta}}_{fix}(\tau)$. From the form of this equation, we propose an approximating estimator of $\boldsymbol{\beta}^*(\tau)$ for high-dimensional setting, that is

$$\widetilde{\boldsymbol{\beta}}(\tau) = (\mathbb{X}'^\top \mathbb{X}')^+ \mathbb{X}'^\top \hat{\mathbf{Y}}_0(\tau),$$

where each row of $\mathbb{X}'$ and $\mathbb{X}$ is generated by the same distribution, but $\mathbb{X}'$ and $\mathbb{X}$ are independent, $\hat{\mathbf{Y}}_0(\tau) = \mathbb{X}'\widehat{\boldsymbol{\beta}}_0(\tau)$, and $\widehat{\boldsymbol{\beta}}_0(\tau)$ may be a sparse and consistent estimator of high-dimensional $\boldsymbol{\beta}^*(\tau)$. The consistency can be expressed as $\left\| \widehat{\boldsymbol{\beta}}_0(\tau) - \boldsymbol{\beta}^*(\tau) \right\|^2 = o_p(1)$. See for example, the L_1-QR estimator [3] and Adaptive Robust Lasso estimator [9] can satisfy the sparsity and consistency property under some conditions. Usually, the $\widehat{\boldsymbol{\beta}}_0(\tau)$ can be estimated from a simple random sub-sample $\{(Y_i, \mathbf{X}_i), i \in \mathcal{H}\}$ with a small sample size $m = \#(\mathcal{H})$ of sample $\{(Y_i, \mathbf{X}_i), i = 1, \ldots, N\}$ and $\mathbb{X}'$ is independent of $\{(Y_i, \mathbf{X}_i), i \in \mathcal{H}\}$ and with the the number of rows being m'. Since $\widehat{\beta}_0(\tau)$ is sparse and consistent, if the number of non-zero variables s satisfies $\sup_{\|\boldsymbol{\theta}\|_2 \le 1, \|\boldsymbol{\theta}\|_0 \le s} \boldsymbol{\theta}^T \left(\frac{1}{m'} \mathbb{X}'^\top \mathbb{X}' \right) \boldsymbol{\theta} = O_p(1)$, we have

$$\left| \frac{1}{m'} \mathbb{X}'^\top \hat{\mathbf{Y}}_0 - \frac{1}{m'} \mathbb{X}'^\top \left\{ \mathbb{X}' \boldsymbol{\beta}^*(\tau) \right\} \right| = o_p(1),$$

then $\widetilde{\boldsymbol{\beta}}(\tau)$ is also a consist estimator of $\boldsymbol{\beta}^*(\tau)$ and may contain information about true coefficients $\boldsymbol{\beta}^*(\tau)$, such as maintaining the same rank order as that of $\boldsymbol{\beta}^*(\tau)$. Since $\mathbb{X}'^T\mathbb{X}'$ is irreversible, with the help of HOLP idea, the HOLP estimator for high-dimensional quantile regression is given by

$$\widehat{\boldsymbol{\beta}}_{Hp}(\tau) = (\mathbb{X}'^\top \mathbb{X}')^+ \mathbb{X}'^\top \hat{\mathbf{Y}}_0(\tau) = \mathbb{X}'^\top (\mathbb{X}' \mathbb{X}'^\top)^{-1} \hat{\mathbf{Y}}_0(\tau).$$

2.2 Distributed high-dimensional quantile algorithm

To save calculation time and computation running memory, we use a common divide-and-conquer strategy. We firstly randomly split the whole sample into L subsets $\{\mathcal{H}_l\}_{l=1}^L$, each of equal size $n = N/L$. Correspondingly, the entire dataset $\{(Y_i, \mathbf{X}_i), i = 1, \ldots, N\}$ is divided into L subsets $\mathcal{D}_1, \ldots, \mathcal{D}_L$, where $\mathcal{D}_l = \{(Y_i, \boldsymbol{X}_i), i \in \mathcal{H}_l\}$ for $1 \le l \le L$. Every subset is stored in different machine node.

We use the data in the first machine node to get the initial estimator $\hat{\boldsymbol{\beta}}_0$, the data in the rest machine node to sort the variables, and select the model and obtain the estimator simultaneously. Then we translate the selected model and estimators to central machine node, and obtain the averaging final estimator by vote. The algorithm is summarized in Table 1.

The rationalities of the **Steps 2** and **3** of the proposed algorithm are guaranteed in Section 3. Now, we will elaborate on **Step 4, Data driven variables screening process**, which is original proposed in current paper.

Table 1: Fast Distributed Screening Algorithm for High-dimensional Quantile Regression

Algorithm 1: Distributed High-dimensional Quantile Regression (DHQ)

Input: $\{(Y_i, \mathbf{X}_i), i = 1, \ldots, N\}$.

Output: Final estimator $\hat{\boldsymbol{\beta}}(\tau)$.

Step 1: Partition the samples.

1.1:Partition the samples to L subsets, and each subset contains $N/L = n$ samples randomly and is stored in each machine node.

Step 2: Obtain the initial estimator $\hat{\boldsymbol{\beta}}_0$.

We use the data in the first machine to get the initial estimator

$$\hat{\boldsymbol{\beta}}_0 = \arg \min_{\boldsymbol{\beta}=\left(\beta_1,\beta_2,\ldots,\beta_{p+1}\right)^\top \in R^{p+1}} \frac{1}{n} \sum_{i \in \mathcal{H}_1} \rho_\tau(Y_i - X_i^T \boldsymbol{\beta}) + \lambda_n \|\boldsymbol{\beta}\|_1,$$

where λ_n can be chosen by 10-fold cross-validation and methods in [3] and [9]. To save the compute time, we suggest λ_n to be a small positive number, such as $\sqrt{\log p}/n$.

Step 3: Sort variables.

3.1: In the lth machine, we obtain $\hat{\mathbf{Y}}^*_{l,0} = \mathbb{X}_l \hat{\beta}_0$.

3.2: Sort the $\hat{\boldsymbol{\beta}}_l = \mathbb{X}_l^T (\mathbb{X}_l \mathbb{X}_l^T)^{-1} \hat{\mathbf{Y}}^*_{(l,0)}$ in decreasing order, and select the top q index set denoted by

$$\hat{\mathcal{F}}_{l,0} = \{i_1^{(l)}, i_2^{(l)}, \cdots, i_q^{(l)}\}. \ (s < q < p).$$

Step 4: Data driven variables screening process.

4.1: Consider a sequence of nested candidate models by the index $\hat{\mathcal{F}}_{l,0} = \{i_1^{(l)}, i_2^{(l)}, \ldots, i_q^{(l)}\}$. The jth candidate model is set by the first j variables in the index $\hat{\mathcal{F}}_{l,0}$.

4.2: Use the proposed data-driven screening method to obtain the threshold T_l in feature screening index set, the corresponding parameter estimation $\hat{\boldsymbol{\beta}}^{(l)}_{(T_l)}$, the definition of T_l is contained in Section 2.3. That is, the selected model in the lth subsets consists of the following variables $\hat{S}_l = \{i_1^{(l)}, i_2^{(l)}, \ldots, i_{T_l}^{(l)}\}$. The estimator of $\boldsymbol{\beta}^*(\tau)$ is $\widehat{\boldsymbol{\beta}}_l = \Pi_{T_l}^\top \hat{\boldsymbol{\beta}}^{(l)}_{(T_l)}$, where Π_k and $\hat{\boldsymbol{\beta}}^{(l)}_{k)}$ are the selection matrix and QR estimator for the candidate model with variable index set $\{i_1^{(l)}, i_2^{(l)}, \ldots, i_k^{(l)}\}$, respectively, $k = 1, \ldots, q$. For $j = 1 : p$, do $\hat{v}_j^{(l)} = I(\hat{\beta}_{l,j} \neq 0)$.

Step 5: Vote to select the important variables.

5.1: On the central machine, count the number of occurrences of each variable j in $\hat{S}_1 \cup \hat{S}_2 \cup \cdots \cup \hat{S}_L$, that is, $\hat{w}_j = I\left\{L^{-1} \sum_{l=1}^L \hat{v}_j^{(l)} > 1/2\right\}$.

5.2: Obtain the final estimator of $\beta_j(\tau)$, i.e., $\hat{\beta}_j(\tau) = \frac{1}{L} \sum_{l=1}^L \hat{w}_j \hat{\beta}_{l,j}$. Then

$$\widehat{\boldsymbol{\beta}}(\tau) = \left(\hat{\beta}_1(\tau), \hat{\beta}_2(\tau), \cdots .\hat{\beta}_{p+1}(\tau)\right)^T$$

End

2.3 Data driven screening method for high-dimensional quantile regression

We compare the difference of goodness of fit between under-fitted scenarios and corrected scenarios and propose a data driven variable screening for quantile regression to select the model with oracle property. Since that $\hat{\mathcal{F}}_{l,0} = \{j_1, j_2, \cdots, j_q\}$ is the top q index set by sorting $|\widehat{\boldsymbol{\beta}}_0(\tau)|$ in decreasing order. For $k = 1, \ldots, n$, let $\widehat{\mathbf{S}}_k = \{j_1, \cdots, j_k\}$, and

$$\begin{aligned} QL_k &= \min_{\boldsymbol{\beta}_{\widehat{\mathbf{S}}_k}} \sum_{i=1}^{n} \rho_\tau(Y_i - \mathbf{X}_{i,\widehat{\mathbf{S}}_k}^T \boldsymbol{\beta}_{\widehat{\mathbf{S}}_k}) = \sum_{i=1}^{n} \rho_\tau\{Y_i - \mathbf{X}_{i,\widehat{\mathbf{S}}_k}^T \widehat{\boldsymbol{\beta}}_{\widehat{\mathbf{S}}_k}(\tau)\}, \quad (4) \\ QL_{k-1} &= \min_{\boldsymbol{\beta}_{\widehat{\mathbf{S}}_{k-1}}} \sum_{i=1}^{n} \rho_\tau(Y_i - \mathbf{X}_{i,\widehat{\mathbf{S}}_{k-1}}^T \boldsymbol{\beta}_{\widehat{\mathbf{S}}_{k-1}}) = \sum_{i=1}^{n} \rho_\tau\{Y_i - \mathbf{X}_{i,\widehat{\mathbf{S}}_{k-1}}^T \widehat{\boldsymbol{\beta}}_{\widehat{\mathbf{S}}_{k-1}}(\tau)\}. \end{aligned}$$

We do variable screening by giving a threshold $T = \min_{k\in[1,q]}(QL_{k-1} - QL_k)/\sqrt{n} < c_n^*$ for the difference sequence series $\{QL_{k-1} - QL_k\}_{k=1}^{q}$, the selection strategy of c_n^* is contained in Section 3.3, for convenience, we generally set it to be a small constant. From a visualization perspective, we use Shewhart control chart chart method of quality control.

3 Theoretical Results

We establish the important properties for proposed method by presenting the following theoretical results. In the following, we assume the density function f is continuously,strictly positive in a neighborhood of zero and has a bounded first derivative in the neighborhood of 0.

3.1 Screening property for penalized high-dimensional quantile regression

In this subsection, we investigate the screening consistency for penalized high-dimensional quantile regressions. Under some regular conditions, we prove that the important and unimportant variables are separable by simply thresholding the estimated coefficients in $\hat{\beta}_0$. Our theoretical results holds for L_1 penalty (see for example, Lasso, [3]), L_2 penalty (see for example, Ridge regression) and seamless L_0 penalty [8].

Conditions

Define $\mathbf{Z}$ and z respectively as

$$\begin{aligned} &\mathbf{Z} = \mathbb{X}\Sigma^{-1/2}, \\ &z = \Sigma^{-1/2}x, \end{aligned} \quad (5)$$

where x is a repetition of $\mathbf{X}_i$ and $\Sigma = Cov(x)$ is the covariance matrix of the predictors. For simplicity, we assume that the $\mathbf{X}_i$s are standardized to have mean 0 and standard deviation 1, i.e. Σ is the correlation matrix. It is easy to see that the covariance matrix of z is an identity matrix.

Condition 1 The transformed z has a spherically symmetric distribution and there are some $c_1 > 1$ and $C_1 > 0$ such that

$$P\left\{\lambda_{\max}\left(p^{-1}ZZ^T\right) > c_1 \quad \text{or} \quad \lambda_{\min}\left(p^{-1}ZZ^T\right) < c_1^{-1}\right\} \leq \exp\left(-C_1 n\right), \tag{6}$$

where $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ are the largest and smallest eigenvalues of a matrix respectively. Assume that $p > cn$ for some $c > 1$.

Condition 2 Asume that $\boldsymbol{\beta}^{*T}(\tau)\Sigma\boldsymbol{\beta}^*(\tau) \leqslant c'$ and there exists some $\kappa \geq 0$, $\nu \geq 0$, $\zeta \geq 0$ such that

$$\min_{j\in S}\left|\boldsymbol{\beta}_j\right| \geq \frac{c_2}{n^\kappa}, \quad s \leq c_3 n^\nu, \quad \text{and} \quad \operatorname{cond}(\Sigma) \leq c_4 n^\zeta, \tag{7}$$

where $\operatorname{cond}(\Sigma)$ is the conditional number of Σ.

Condition 3 $k_n \triangleq \max_{i,j}|x_{i,j}| = o_p\left(\sqrt{n}s^{-1}\right)$.

Condition 4 There exist universal constants $\tilde{c}_1 > 0$ and $\tilde{c}_2 > 0$ such that for any u satisfying $|u| \leqslant \tilde{c}_1$, $f_i(u)$'s are uniformly bounded away from 0 and ∞ and

$$|F_i(u) - F_i(0) - uf_i(0)| \leqslant \tilde{c}_2 u^2, \tag{8}$$

where $f_i(u)$ and $F_i(u)$ are the density function and distribution function of the error ε_i, respectively.

Condition 5 With γ_n defined in Lemma 2, assume that

$$\lambda_n > 2\gamma_n \left\|\frac{1}{n}Q^T S\right\|_{2,\infty}, \tag{9}$$

where $\|A\|_{2,\infty} = \sup_{x\neq 0}\|AX\|_\infty / \|x\|_2$ for matrix A and vector x.

Theorem 1. *Under Conditions 1-5. Let* $\mathbf{S} = \{j, \beta_j(\tau) \neq 0, 1 \leq j \leq p_n\}$. *If there exists* r_n *with*

$$c_2' \frac{n^{1+\zeta+\nu/2}}{p}\gamma_n < r_n < C\frac{n^{1-\zeta-\kappa}}{p} \tag{10}$$

for constants $c_2', C > 0$, *then*

$$P\left(\mathbf{S} \in \mathcal{M}_{r_n}\right) \to 1,$$

where $\mathcal{M}_{r_n} = \left\{j, \left|\hat{\beta}_j(\tau)\right| > r_n, 1 \leq j \leq p_n\right\}$.

Theorem 2. *If there exists* r_n *with*

$$\frac{\tilde{c}}{\sqrt{\log n}}\frac{n^{1-\zeta-\kappa}}{p} < r_n, \tag{11}$$

such that

$$P\left(\min_{j\in\mathbf{S}}|\widehat{\beta}_j| > r_n > \max_{j\in\mathbf{S}^c}|\widehat{\beta}_j|\right) \to 1.$$

3.2 Model selection criterion for high-dimensional quantile regressions

In this section, we derive rates of convergence for quantile regression coefficients under specifications. Assume that $\mathbf{X}_{i(k-1)}^T = \mathbf{X}_i^T \Pi_k^T$ and $\Pi_k \Pi_k^T = I_{p_k}$ where I_{p_k} is an identity matrix of order p_k . For notational ease, we use $\boldsymbol{\beta}_{(k)}$, $\mathbf{X}_{i(k)}$, $\boldsymbol{\beta}_{(k-1)}$ and $\mathbf{X}_{i(k-1)}$ to replace $\boldsymbol{\beta}_{\widehat{\mathbf{S}}_k}$, $\mathbf{X}_{i,\widehat{\mathbf{S}}_k}$, $\boldsymbol{\beta}_{\widehat{\mathbf{S}}_{k-1}}$ and $\mathbf{X}_{i,\widehat{\mathbf{S}}_{k-1}}$, respectively, and introduce

$$\begin{aligned} \boldsymbol{\beta}_{(k)}^{**}(\tau) &= \arg\min_{\boldsymbol{\beta}_{(k)}} E[\rho_\tau\{Y_i - \mathbf{X}_{i(k)}\boldsymbol{\beta}_{(k)}\}], \\ \boldsymbol{\beta}_{(k-1)}^{**}(\tau) &= \arg\min_{\boldsymbol{\beta}_{(k-1)}} E[\rho_\tau\{Y_i - \mathbf{X}_{i(k-1)}\boldsymbol{\beta}_{(k-1)}\}], \end{aligned}$$

the pseudo true parameter vectors under restricted models, and denote QL_k^{**} and QL_{k-1}^{**} as the corresponding versions of QL_k and QL_{k-1}, respectively; if $\mathbf{S} \subseteq \widehat{\mathbf{S}}_k$, then $\Pi_k^T \boldsymbol{\beta}_{(k)}^{**}(\tau) = \boldsymbol{\beta}^*(\tau)$, the true parameter sub vector, and $\varepsilon_i(\tau) = Y_i - \mathbf{X}_{i(k)}^T \boldsymbol{\beta}_{(k)}^{**}(\tau)$.

For $k = 1, \ldots, p$, denote

$$\begin{aligned} A_{(k)} &\equiv E\left[f\left(-u_{i(k)}\right)\mathbf{X}_{i(k)}\mathbf{X}_{i(k)}^T\right], \\ B_{(k)} &\equiv E\left[\psi_\tau\left(\varepsilon_i + u_{i(k)}\right)^2 \mathbf{X}_{i(k)}\mathbf{X}_{i(k)}^T\right], \\ V_{(k)} &\equiv A_{(k)}^{-1} B_{(k)} A_{(k)}^{-1}, \end{aligned}$$

where $u_{i(k)} \equiv \mu_i - \mathbf{x}_{i(k)}^T \boldsymbol{\beta}_{(k)}^{**}(\tau)$ indicates the approximation bias for the k'th QR model and $\psi_\tau(\varepsilon_i) \equiv \tau - \mathbf{1}\{\varepsilon_i \le 0\}$.

Condition 6 Define $\Sigma_{(k)} = E\{\mathbf{X}_{i(k)}\mathbf{X}_{i(k)}^T\}$, $\sup_{k\in[1,s]}(\boldsymbol{\beta}_{(k)}^{**}(\tau))^T \Sigma_{(k)} \boldsymbol{\beta}_{(k)}^{**}(\tau) \le c'$.

Condition 7 The density of the error, f, is uniformly bounded from above by $f_{\max} < +\infty$ and from below by $f_{\min} > 0$. Furthermore, f is continuously differentiable with the derivative, f', bounded by $f'_{\max}$.

Condition 8 Assume that there is a $\epsilon_n > 0$ subject to $\|\boldsymbol{\beta}_{(k)}^{**}(\tau) - (\boldsymbol{\beta}_{(k-1)}^{**}(\tau)^T, 0)^T\| \ge \epsilon_n$ for $k \in [1, s]$, satisfying $\epsilon_n^2 / \left(n^{-1}s^{9/2} + n^{-1/2}s\right) \ge c > 0$.

Condition 9 $P(\lambda_{max}\left\{\sum_{i=1}^n \mathbf{X}_{i(q)}\mathbf{X}_{i(q)}^T - n\Sigma_{(q)}\right\} > nq^{-1/2}\log^{-1} n) = o((q-s)^{-2}n^{-1})$, and there is $\mu > 0$ satisfies $(q-s)n^{\mu-1/2} = o(1)$ and $P(\max_{i\in[1,n]} \|\mathbf{X}_{i(q)}\| > n^\mu) = o((q-s)^{-2}n^{-1})$.

Condition 10 (i) There exist constants $\underline{c}_{A(k)}$ and $\bar{c}_{A(k)}$ that may depend on k such that $0 < \underline{c}_{A(k)} \le \lambda_{\min}\left(A_{(k)}\right) \le \lambda_{\max}\left(A_{(k)}\right) \le f_{\max}\lambda_{\max}\left(E\left[\mathbf{X}_{i(k)}\mathbf{X}_{i(k)}^T\right]\right) \le \bar{c}_{A(k)} < \infty$.
(ii) There exist constants $\underline{c}_{B(k)}$ and $\bar{c}_{B(k)}$ that may depend on k such that $0 < \underline{c}_{B(k)} \le \lambda_{\min}\left(B_{(k)}\right) \le \lambda_{\max}\left(B_{(k)}\right) \le \bar{c}_{B(k)} < \infty$.
(iii) $\sup_{k\in[1,q]}\left(\bar{c}_{A(k)} + \bar{c}_{B(k)}\right)/k\underline{c}_{A(k)}^2 = O(1)$.

Condition 11 $E\left(\mu_i^4\right) < \infty$ and $\sup_{k\in[1,p]} E\left(\|\mathbf{X}_{i(k)}\|^8\right) \le c_{\mathbf{x}}$ for some $c_{\mathbf{x}} < \infty$.

Condition 12 As $n \to \infty, s^4\bar{c}_{A(s)}/(n\underline{c}_{B(s)}) \to 0$, and $s^4(\log n)^4/(n\underline{c}_{B(s)}^2) \to 0$.

Theorem 3. *Let QL_k and QL_{k-1} be the empirical quantile loss defined in (4). Under conditions 6-10, we have*
(i) If X_{j_k} is an important predictor, then $(QL_{k-1} - QL_k)/\sqrt{n} \geq c_0$ for some positive constant $c_0 > 0$. Specifically, with probability approaching 1,

$$\inf_{k\in[1,s]} (QL_{k-1} - QL_k) \geq c_0(s^4 + s^{1/2}n^{1/2});$$

(ii) If $\mathbf{S} \subseteq \widehat{\mathbf{S}}_{k-1}$, i.e. $\widehat{\mathbf{S}}_{k-1}$ has already included all the important predictors, we have

$$(QL_{k-1} - QL_k)/\sqrt{n} \to 0$$

if $(q-s)^2 \log n + q^2(q-s)n^{\mu-1/2} \log^2 n = o(n^{1/2})$. Specifically,

$$\sup_{k\in[s+1,q]} |QL_{k-1} - QL_k| = o_p((q-s)^2 \log n) + O_p(q^2(q-s)n^{\mu-1/2} \log^2 n); \tag{12}$$

(iii) If $q^2(q-s)n^{\mu-1/2}\log^2 n = o(1)$. For two random variables a and b on $\mathbb{R}$, we define the convex-indicator (CI) distance Δ_{CI} by $\Delta_{CI}(a,b) = \sup_{A\in\mathbb{R}\ convex} |P(a \in A) - P(b \in A)|$, then

$$\sup_{k\in[s+1,q]} \Delta_{CI}(2(QL_{k-1} - QL_k)/\{\lambda(\tau)w(\tau)\}, \chi^2_1) = o(1), \quad \text{as } n \to \infty. \tag{13}$$

3.3 A data-driven method to select s

We set $q = [n/log(n)]^+$, where $[a]^+$ means the smallest integer greater than a.

Denote $W_{m,l} = \left(QL_{n-m,l} - QL_{n-(m-1),l}\right)/\sqrt{n}$, $m = 1, 2, \ldots, n-q+1$, $l = 1, \ldots, L$.

We propose a data-driven method to select s. This method is based on statistical control chart which is a graphical display of the $W_{m,l}$ used for model selection. This method does not depend on the design of information criteria, and is more flexible to identify the correct variables and true model.

The detailed procedures are as following:

1. Calculate the following statistics:

$$\hat{\mu}_{W,l} = \frac{1}{q'}\sum_{m=1}^{q'} W_{m,l},$$

   and

$$\hat{\sigma}_{W,l} = \sqrt{\frac{\sum_{m=1}^{q'}\left(W_{m,l} - \hat{\mu}_{W,l}\right)^2}{q'-1}},$$

   where $q' = 30$.

2. Then select the

$$\hat{s}^*_l = \inf_{q'\leq m\leq q} (W_{m,l} - \hat{\mu}_w) > b\hat{\sigma}_{W,l}.$$

   We set $b = 6$. If there is no available variable can be selected when $b = 6$, we change b to 3 to select signals. The choice of b can be aided by visual diagrams. Usually, we set $b = 6$

firstly to select the strong signals, and then the weak signals are selected by setting $b = 3$ with the models containing selected strong signals. More discussion about choice of b can be our further work.

There are also some other type Shewhart control charts to apply and more detailed instructions are provided in Wang et al.(2013).

3.4 A majority vote weighted method to combine the local estimators

Since the randomness of the estimators $\hat{\mathcal{F}}_l$ and $\hat{s}_l^*$, we apply a vote weighted aggregating method to obtain our final estimator. The aggregating method can be seen in Wang et al. (2014). The detailed process is as follows:

1. For each variable, we calculate the weight $\hat{w}_j = I\left\{L^{-1}\sum_{l=1}^{L} I(\hat{\beta}_{l,j} \neq 0) > 1/2\right\}$
2. The final weighted estimator of $\boldsymbol{\beta}(\tau)$ is given by $\widehat{\boldsymbol{\beta}}(\tau) = \left(\hat{\beta}_1(\tau), \hat{\beta}_2(\tau), \cdots .\hat{\beta}_{p+1}(\tau)\right)^T$ with $\hat{\beta}_j(\tau) = \frac{1}{L}\sum_{l=1}^{L} \hat{w}_j \hat{\beta}_{l,j}$.

The differences between our method and Wang et al. (2014) are mainly in two aspects: (i) The model considered in Wang et al. (2014) is high-dimensional linear model, while we consider high-dimensional quantile regression model; (i) The second step in Wang et al. (2014) is averaging the refitted estimator in each sub-machine, while we conduct variable screening method before the aggregating step and use the weighted estimator as final estimator. Since the method in Wang et al. (2014) is for linear model, so We do not compare this approach in our simulation studies.

3.5 Asymptotic properties for final estimator

Theorem 4. *(Consistency in variable selection) Under conditions 6-10, suppose the parameter c_n^* satisfies $c_n^* = o(s^{1/2})$ and $(q-s)^2 \log n + q^2(q-s)n^{\mu-1/2}\log^2 n = o(c_n^* n^{1/2})$, denote $S_n = \{j : \hat{\beta}_j(\tau) = 1, j = 1, \ldots, p\}$, then $P(S_n = \mathbf{S}) \to 1$.*

Proof. Denote $h_n = P(\widehat{S}_1 \neq \mathbf{S})$, so $E\sum_{l=1}^{L} I_{\{\widehat{S}_l \neq \mathbf{S}\}} = Lh_n$. Note that as long as more than half subsets select the correct model, we have $S_n = \mathbf{S}$, then by Chernoff's inequality,

$$
\begin{aligned}
P(S_n \neq \mathbf{S}) &\leq P(\sum_{l=1}^{L} I_{\{\widehat{S}_1 = \mathbf{S}\}} < L/2) \leq P(\sum_{l=1}^{L} I_{\{\widehat{S}_1 \neq \mathbf{S}\}} > L/2) \\
&\leq \exp\{L/2 - Lh_n\} \cdot (2h_n)^{L/2}. \qquad (14)
\end{aligned}
$$

Note that T_1 is the least integer such that $\left(QL_{T_1-1} - QL_{T_1}\right)/\sqrt{n} < c_n^*$, we have

$$
\begin{aligned}
& P(\widehat{S}_1 \neq \mathbf{S}) = P(\widehat{S}_1 \neq \mathbf{S}, \left(QL_{T_1-1} - QL_{T_1}\right)/\sqrt{n} < c_n^*) \\
= \ & P(T_1 > s+1, \left(QL_{T_1-1} - QL_{T_1}\right)/\sqrt{n} < c_n^*) + P(T_1 < s+1, \left(QL_{T_1-1} - QL_{T_1}\right)/\sqrt{n} < c_n^*) \\
\leq \ & P(\left(QL_s - QL_{s+1}\right)/\sqrt{n} \geq c_n^*) + P(\inf_{k\in[1,s]} \left(QL_{k-1} - QL_k\right)/\sqrt{n} < c_n^*) \qquad (15)
\end{aligned}
$$

since $c_n^* = o(s^{1/2})$ and $(q-s)^2 \log n + q^2(q-s)n^{\mu-1/2}\log^2 n = o(c_n^* n^{1/2})$ and (i), (ii) of Theorem 3, it is easy to see the right of (15) tends to zero as $n \to \infty$. Hence $h_n \to 0$, which together with (14) leads to $P(S_n = \mathbf{S}) \to 1$.

Theorem 5. *(Oracle properties)) Under conditions 6–12, suppose the parameter c_n^* satisfies $c_n^* = o(s_n^{1/2})$ and $(q-s)^2 \log n + q^2(q-s)n^{\mu-1/2}\log^2 n = o(c_n^* n^{1/2})$, we have*

$$\sqrt{n} C_n V_{(s)}^{-1/2}(\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}^*(\tau)) \xrightarrow{d} N(0, G). \tag{16}$$

where C_n is a $m \times s$ matrix such that $C_n C_n^T \to G$, and G is a $m \times m$ nonnegative symmetric matrix

Proof. By Theorem 3.1 of [19], we have

$$\sqrt{n} C_n V_{(s)}^{-1/2}(\widehat{\boldsymbol{\beta}}_{(s)}^{(l)}(\tau) - \boldsymbol{\beta}^*(\tau)) \xrightarrow{d} N(0, G), \tag{17}$$

note that Theorem 4 infers with probability tends to 1,

$$\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}^*(\tau) = \frac{1}{L}\sum_{l=1}^{L}(\widehat{\boldsymbol{\beta}}_{(s)}^{(l)}(\tau) - \boldsymbol{\beta}^*(\tau)). \tag{18}$$

Since the data in L subset are mutually independent, (16) holds.

4 Simulations

In this section, we conduct simulation experiments to compare the finite sample performance of our distributed screening methods and some commonly used distributed screening methods. The effectiveness of our distributed method is demonstrated by comparing with the whole data method. In detail, we compare the following method:

(i) simple averaging Lasso estimator (DLS);

Distributed HOLP method:

(ii) simple averaging distributed estimators (DHXS),

(iii) voting averaging distributed estimators (DHXV),

(iv) simple averaging distributed estimators with only one initial estimator(DlHXS),

(v) voting averaging distributed estimators with only one initial estimator(DlHXV);

Distributed Rank control charts (QC) method without turning parameter selections:

(vi) simple averaging QC selected estimator with Lasso (DLHS),

(vii) voting averaging QC selected estimator with Lasso (DLHV);

and two whole data method :

(viii) Lasso (WL),

(ix) quantile regression (WQ).

The data generating process is

$$Y_i = \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau) + e_i - F^{-1}(\tau), \quad i = 1, 2, \ldots, n, \tag{19}$$

where F is the cdf and F^{-1} is the corresponding quantile function so that the τth conditional quantile function is just $\boldsymbol{X}_i^{\mathrm{T}}\beta^*$, and $\boldsymbol{X}_i^{\mathrm{T}} = (1, X_{i,1}, \ldots, X_{i,p})$ is a $(p+1)$-dimensional covariate vector and $(X_{i,1}, \ldots, X_{i,p})$s are drawn *i.i.d.* from a multivariate normal distribution $N(0, \boldsymbol{S})$ with $\boldsymbol{S} = (S_{i,j})_{1 \le i,j \le p}$. Specifically, two cases for (19) are considered:

- Case 1: $p = 1000$ and $\beta^* = (1, 1, 1/2, 1/3, 1/4, 1/5, 0, \ldots, 0)^T$. The covariance matrix $\boldsymbol{S}$ is constructed by Equi-correlation $S_{ij} = \rho = 0.8$.
- Case 2: $p = 1000$ and $\beta^* = (10/s, 20/s, 30/s, \ldots, 10(s-1)/s, 0, \ldots, 0)^T$ with $s = 20$. The covariance matrix $\boldsymbol{S}$ is constructed by Toeplitz $S_{ij} = \rho^{|i-j|}$ for $1 \le i, j \le p$.

The values of τ are considered to be 0.3 and 0.5. For the noise e_i, we compare the results in the following five kinds of distribution:

(i) Cauchy distribution: Cauchy$(0, 1)$,

(ii) Exponential distribution: $\exp(1)$,

(iii) Normal distribution: $\mathrm{N}(0, 1)$,

(iv) t distribution: $\mathrm{t}(2)$,

(v) χ^2 distribution: ${\chi^2}_{(1)}$.

We mainly focus on the effect of the number of subsets L and effect of sub-sample size and number of distribute machines when the total sample size is fixed, which will be studied thoroughly in Section 4.1 and 4.2 for Case 1, and Section 4.3 and 4.4 for Case 2. The performance is evaluated for each method in terms of the following measurements: (i) L_2-error from the estimations to the true QR coefficients($Loss\_beta$), (ii) the compute time for each method ($Time$), (iii) the number of non-zero coefficients in each method ($Model\_Size$), the number of true positive results (TP), the number of false positive results (FP, i.e. wrong predictors), and F_1-score for model selection consistency perspective, which is given by

$$F_1 = \left(\frac{recall^{-1} + precision^{-1}}{2}\right)^{-1}.$$

We will report simulation results of Case 1-2 in each subsections. We repeat each experiment 200 times and report the mean value for each measurement

4.1 Effects of the Number of Subsets in Case 1

In this subsection, we focus the effect of the number of subsets L in Case 1. We set $n = 500$ and $L = 10, 20$ and 40. See Tables 2-6 for the results of five types errors described in before. It is clearly that:

1) The *Loss_beta* of all methods decreases as K increasing from 10 to 40 except method DHXS,DHXV ,DLHXS and DlHXV when $L = 20$. The reason may be that when the number of subsets increases from $L = 10$ to $L = 20$ on the condition that all the subsets have the same sub-sample size, simple averaging methods are more likely to lead to overfitting models, which will increase the squared loss of quantile parameter estimators. When From the results of $L = 10$ or $L = 20$ to that of $L = 40$, It can be seen that increasing the number of nodes can improve the accuracy of parameter estimation.

2) The *time* of each method increases with L increasing. Our method DlHSV and DLHV always obtain the top two smallest computing time among all the methods. The method WQ has the longest computing time, followed by WL and distributed methods DHXS and DHXV.

3) The *Loss_beta* of each method with $\tau = 0.5$ is always smaller than that with $\tau = 0.3$. The *time* of each method with $\tau = 0.5$ is often longer than that with $\tau = 0.3$.

4) DLS frequently obtains the the largest *Loss_beta*, followed by WL and WQ. This means our method DlHXV and DLHV can improve the prediction accuracy.

5) None of these methods leave out important variables, and the *model_Size*s of DLS, WL and WQ methods are always the top three largest, followed by DHXS, DLHS. Our DlHXV and DLHV methods always get 1 in F_1 -Score. All these phenomenon reveals that our methods have good oracle properties in variable screening.

6) DLHV often gets the smallest *Loss_beta* and $Time$ in all the distribute methods, followed by DlHXV.

For different error distributions across Table 2 to Table 6, it can be find that for each method in each setting, *Loss_beta* of Cauchy distribution is smaller than that of other distributions, and *Loss_beta* of t distribution is larger than that of the rest distributions when $\tau = 0.3$ and $\tau = 0.5$. As for $Time$, there is no obvious or unusual rhythms in the computation time $Time$ for different settings across each type distribution of error.

In summary, our method DlHXV and DHXV performs the best in all the distribution methods and similarly to the best single machine method in each situation.

4.2 Effects of Data Splitting Tricks in Case 1

The effects of splitting sample trick are considered in this subsection for Case 1. We compare three splitting settings, $n = 250$ with $L = 40$, $n = 500$ with $L = 20$, and $n = 1000$ with $L = 10$ for all the methods. The results are summarized in Tables 7-11 for the different types errors. From these Tables, we can conclude that:

1) The *Loss_beta* of each method with $\tau = 0.5$ are smaller than that with $\tau = 0.3$. The $Time$ of each method with $\tau = 0.5$ are mostly longer than that with $\tau = 0.3$ except for χ_2 distribution in Table 11.

2) DLHV always has oracle model selection consistency and leads to the smallest *Loss_beta* with the shortest computing time (See $Time$ in Table 7-11). $Time$ of each method is increasing when n varies from 250 to 1000. The performances of distributed methods DHXV and DlHXV are similar and near to obtain the minimal values in *Loss_beta* and $Time$ measurements.

3) The $Model_Size$s of our methods DHXV, DlHXV and DLHV are always near the number of the true variables except for the cases that $n = 1000$ with $L = 10$. The F_1-Scores of our methods are significantly higher than that of the other methods.

4) With n increasing from 250 to 1000, the *Loss_beta* of all methods decrease except the methods DLHS and DLHV, which are raising when n fluctuates from 500 to 1000.

Table 7 with Cauchy errors shows that when $n = 1000$, DlHXV method behaves better in the *Loss_beta* sense than that of other settings. For $\tau = 0.3$ and 0.5, each of DLHS and DLHV methods attends the smallest *Loss_beta* value of itself when $n = 250$ with $L = 40$, the rest distributed methods get their smallest *Loss_beta* value when $n = 1000$ with $L = 10$. The reason may be that big n can give a better rank estimator than small n, which will improve the performances of the distributed methods except DLHS and DLHV methods. As for $Time$, each method in the setting that $n = 500$ with $L = 20$ is shorter than that in the rest settings when $\tau = 0.3$. However, there is no similar pattern for $\tau = 0.5$. DHXV and DLHV often have smaller $Model\_Size$ with containing all important variables when $\tau = 0.3$ and 0.5. DLHV often gets the smallest *Loss_beta* with TP being 6. This means DLHV method can achieve model selection consistency. To save space, we the detailed analysis for the results in Tables 8-11 are omitted.

In conclusion, our proposed methods DHXV, DlHXV, and DLHV often achieve the smallest *Loss_beta* and $Loss$ with shortest computing time.

4.3 Effects of Number of Subsets in Case 2

For Case 2, simulation results about the effects of number of subsets are given in Tables 12-16. We let $n = 1000$ and $L = 5, 10$ and 20 for each distribution of error e_i.

In this subsection, we focus the effect of the number of subsets L. We set $n = 1000$ and $L = 5, 10$ and 20. We repeat each experiment 200 times and report the mean value for each measurement. we present the results of Case 2 in Tables 12-16. From these Tables, it can be seen that:

1) The *Loss_beta* of all methods decreases when L increases from 5 to 20 .

2) The $time$ of each method increases with L increasing.

3) The *Loss_beta* of each method with $\tau = 0.5$ are smaller than that with $\tau = 0.3$. The $time$ of each method with $\tau = 0.5$ is longer than that with $\tau = 0.3$.

4) DLS always obtain the the largest $Loss\_beta$, followed by WL.

5) The time of WQ is always the largest.

6) DHXV, DlHXV, and DLHV methods are always have smaller $Model\_Size$ than that of the rest methods.

8) Our DHXV, DlHXV, and DLHV methods always have higher F_1-Scores than that of other methods.

For Cauchy error distribution in Table 12, when $\tau = 0.3$ DlHXV gets the smallest $Loss\_beta$ in the case $L = 5$ and DHXV gets the smallest $Loss\_beta$ in the case $L = 10$ and 20. When $\tau = 0.5$, DLHV always obtains the smallest value of $Loss\_beta$.

In summary, our methods DHXV, DlHXV, and DLHV always perform the best in all the distribution methods and similarly to the best single machine method in each situation.

4.4 Effects of Data Splitting in Case 2

We report the effects of data splitting results of Case 2 in Tables 17-21. Three splitting settings, $n = 500$ with $L = 20$, $n = 1000$ with $L = 10$, and $n = 2000$ with $L = 5$ are considered for all the methods. The results are summarized in Tables 17-21 for the five types errors described in before.

By Tables 17-21, it is clear that: the $Loss\_beta$ of DHXS, DHXV, DlHXS, DlHXV have bigger drops in value when $n = 500$ to 1000, than that with n varying from 1000 to 2000. The $Loss\_beta$ of each method with $\tau = 0.5$ are smaller than that with $\tau = 0.3$. The *time* of each method with $\tau = 0.5$ is often longer than that with $\tau = 0.3$.

In conclusion, our proposed methods DHXV, DlHXV, and DLHV often achieve the smallest $Loss\_beta$ and $Loss$ with shortest computing time.

5 Real Data Example

We now illustrate the application of the proposed method by analyzing Million Song Dataset data, which collected records of the songs released from 1922 to 2011, and is freely available. We studied the year prediction of the data, one of important and difficult tasks for this dataset due to the large sample sizes and high-dimensional covariates. For more information and details, see [5]. We compare the considered method of Section 4 in real data. The details are as follows. we randomly repeat the calculations 20 times.

We set released year of the song as response variable, and the rest timbre information as covariates, and consider 463, 715 training examples and 51, 630 testing examples, like [23]. We consider the quantile regression of $\tau = 0.5$ and will report the in-sample quantile loss, out-sample quantile loss of prediction, and the variables selected by our methods. We use the data from the first 20,000 records as the training data set and the next 100 records as the test data set. We set $L = 20, 40$ and 80 We repeat the computation 20 times. The prediction results are reported in Table 2.

Table 2: Results for Million Song Dataset

	Method	DLS	DHXS	DHXV	DlHXS	DlHXV	DLHS	DLHV	WL	WQ
L=80	MFSE	0.296	0.280	0.293	0.276	0.281	0.280	0.319	0.262	0.265
	Time	55.769	57.470	57.469	51.387	51.385	52.576	52.574	792.792	1151.471
L=40	MFSE	0.290	0.268	0.264	0.268	0.267	0.272	0.263	0.261	0.264
	Time	87.385	173.204	173.204	165.342	165.341	165.532	165.530	810.164	1282.787
L=20	MFSE	0.284	0.265	0.262	0.265	0.264	0.267	0.265	0.262	0.265
	Time	203.043	229.845	229.845	208.745	208.744	210.802	210.800	879.697	1109.097

6 Conclusion

We propose a distributed data driven screening method for high-dimensional quantile regressions. The theoretical properties are studied and proved. Numerical experimental results show that our proposed methods have certain advantages in calculating time and variable selection consistency. The next step is to extend it to the high-dimensional varying-coefficient quantile regressions.

For the high-dimensional quantile regression, there are various choices of the penalty methods. Examples include but not limited to LASSO, Ridge, and L_0. In this paper we employ the LASSO penalty for illustration although other penalties can also be used. The corresponding theoretical properties are the content of our future research.

Acknowledgment

Jiang's work is supported by NSFC (11871263), and NSF grant of Guangdong Province of China (No. 2017A030313012). Tang's work is partially supported by NCFC(11871376) and NSF grant of Shanghai (21ZR1420700).

A Technical Proofs

Let $e_j = (0, \dots, 0, 1, 0, \dots, 0)^T$ denote the jth natural base in the p dimension space and $\tilde{e}_1$ denote the n-dimensional column vector $(1, 0, \dots, 0)^T$. We cite the following result for the proof of our theorems.

A.1 Lemmas and Proofs

Lemma 1. *Under Conditions 1 and 2, for any $C > 0$, there exists some $c, \tilde{c} > 0$ such that for any $i \in S$,*

$$P\left(\left|e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X}\boldsymbol{\beta}(\tau)\right| < c\frac{n^{1-\zeta-\kappa}}{p}\right) = O\left\{\exp\left(\frac{-Cn^{1-5\zeta-2\kappa-v}}{2\log n}\right)\right\}, \tag{20}$$

and for any $i \notin S$,

$$P\left(\left|e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X}\boldsymbol{\beta}(\tau)\right| > \frac{\tilde{c}}{\sqrt{\log n}}\frac{n^{1-\zeta-\kappa}}{p}\right) = O\left\{\exp\left(\frac{-Cn^{1-5\zeta-2\kappa-v}}{2\log n}\right)\right\}, \tag{21}$$

where ζ, κ, ν are the parameters defined in Condition 2.

Lemma 2. *For any $C>0$, let $\gamma_n = C\left(\sqrt{\frac{s(\log n)}{n}}+\lambda_n\sqrt{s}\right)$. Suppose Conditions 3 and 4 hold and $\lambda_n s k_n = o_p(1)$, then there exists some constant $c>0$ such that*

$$P\left(\left\|\widehat{\boldsymbol{\beta}}_1^{(0)}(\tau)-\boldsymbol{\beta}_1(\tau)\right\|\leqslant\gamma_n\right)\geqslant 1-n^{-cs}. \tag{22}$$

Lemma 3. *Suppose Conditions 3-5 hold. If we choose λ_n such that*

$$\sqrt{1+\gamma_n s^{3/2}k_n^2}\log_2 n = o\left(\sqrt{n}\lambda_n\right), \tag{23}$$

$$n^{1/2}\lambda_n\left(\log p\right)^{-1/2}\to\infty, \tag{24}$$

and

$$k_n\gamma_n^2 = o(\lambda_n) \tag{25}$$

then there exists some constant $c>0$ such that

$$P\left(\left\|\widehat{\boldsymbol{\beta}}_1^{(0)}(\tau)-\boldsymbol{\beta}_1(\tau)\right\|\leqslant\gamma_n, \left\|\widehat{\boldsymbol{\beta}}_2^{(0)}(\tau)\right\|=0\right)\geqslant 1-n^{-cs}. \tag{26}$$

A.1.1 Property of $\mathbb{X}^T\left(\mathbb{X}\mathbb{X}^T\right)^{-1}\mathbb{X}\left\{\widehat{\boldsymbol{\beta}}^{(0)}(\tau)-\boldsymbol{\beta}(\tau)\right\}$

In this part, we aim to bound the magnitude of $\mathbb{X}^T\left(\mathbb{X}\mathbb{X}^T\right)^{-1}\mathbb{X}\left\{\widehat{\boldsymbol{\beta}}^{(0)}(\tau)-\boldsymbol{\beta}(\tau)\right\}$, we have the following lemma.

Lemma 4. *Assume Conditions 1-5 and (23)-(25) hold, we have for any $i\notin S$,*

$$P\left(\left|e_i^T\mathbb{X}^T\left(\mathbb{X}\mathbb{X}^T\right)^{-1}\mathbb{X}\left\{\widehat{\boldsymbol{\beta}}^{(0)}(\tau)-\boldsymbol{\beta}(\tau)\right\}\right|>\frac{\sqrt{Cc_4}}{c_3^{'}}\frac{n^{1+\zeta+\nu/2-\alpha}}{(\log n)p}\gamma_n\right)=O\left\{n^{\nu}\exp\left(\frac{-Cn^{1-2\alpha}}{2\log n}\right)\right\}+O\left\{n^{-cs}\right\}, \tag{27}$$

and for any $i\in S$,

$$P\left(\left|e_i^T\mathbb{X}^T\left(\mathbb{X}\mathbb{X}^T\right)^{-1}\mathbb{X}\left\{\widehat{\boldsymbol{\beta}}^{(0)}(\tau)-\boldsymbol{\beta}(\tau)\right\}\right|>c_2^{'}\frac{n^{1+\zeta+\nu/2}}{p}\gamma_n\right)=O\left\{n^{\nu}\exp\left(-Cn\right)\right\}+O\left\{n^{-cs}\right\}, \tag{28}$$

where $\alpha>0$, ζ,κ,ν are the parameters defined in Condition 2, and γ_n is defined in Lemma 2 and Lemma 3 .

Proof. The lemma 4 of Wang and Leng(2016) [21] gives us that the diagonal terms of $\mathbb{X}^T\left(\mathbb{X}\mathbb{X}^T\right)^{-1}\mathbb{X}$ have the following results, that is, for any $C>0$, there exist constants $c_1^{'}, c_2^{'}$ with $0<c_1^{'}<1<c_2^{'}$, such that

$$P\left(e_i^T\mathbb{X}^T\left(\mathbb{X}\mathbb{X}^T\right)^{-1}\mathbb{X}e_i<c_1^{'}\frac{n^{1-\varsigma}}{p} or e_i^T\mathbb{X}^T\left(\mathbb{X}\mathbb{X}^T\right)^{-1}\mathbb{X}e_i>c_2^{'}\frac{n^{1+\varsigma}}{p}\right)<4\exp^{-Cn}, \tag{29}$$

and the off diagonal terms of $\mathbb{X}^T\left(\mathbb{X}\mathbb{X}^T\right)^{-1}\mathbb{X}$ have an upper bound, that is, for any $C>0$, there exists some $c_3^{'}>0$ such that

$$P\left(\left|e_i^T\mathbb{X}^T\left(\mathbb{X}\mathbb{X}^T\right)^{-1}\mathbb{X}e_j\right|>\frac{M}{\log n}\frac{n^{1+\varsigma-\alpha}}{p}\right)=O\left\{\exp\left(\frac{-Cn^{1-2\alpha}}{2\log n}\right)\right\}, \tag{30}$$

where $i, j = 1, 2, \ldots, p$ and $i \neq j$, and $M = \frac{\sqrt{Cc_4}}{c_3'}$. Using these results, we have for any $i \notin S$

$$
\begin{aligned}
& \left| e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} \left\{ \widehat{\boldsymbol{\beta}}^{(0)}(\tau) - \boldsymbol{\beta}(\tau) \right\} \right| \\
= & \left| \sum_{j=1}^{p} e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} e_j e_j^T \left\{ \widehat{\boldsymbol{\beta}}^{(0)}(\tau) - \boldsymbol{\beta}(\tau) \right\} \right| \\
= & \left| \sum_{j \in S} e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} e_j e_j^T \left\{ \widehat{\boldsymbol{\beta}}^{(0)}(\tau) - \boldsymbol{\beta}(\tau) \right\} + \sum_{j \notin S} e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} e_j e_j^T \left\{ \widehat{\boldsymbol{\beta}}^{(0)}(\tau) - \boldsymbol{\beta}(\tau) \right\} \right| \\
\leqslant & \left\{ \left| \sum_{j \in S} e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} e_j e_j^T \left\{ \widehat{\boldsymbol{\beta}}^{(0)}(\tau) - \boldsymbol{\beta}(\tau) \right\} \right| + \left| \sum_{j \notin S} e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} e_j e_j^T \left\{ \widehat{\boldsymbol{\beta}}^{(0)}(\tau) - \boldsymbol{\beta}(\tau) \right\} \right| \right\} \\
\leqslant & \left\{ \sum_{j \in S} \left| e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} e_j e_j^T \left\{ \widehat{\boldsymbol{\beta}}^{(0)}(\tau) - \boldsymbol{\beta}(\tau) \right\} \right| + \sum_{j \notin S} \left| e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} e_j e_j^T \left\{ \widehat{\boldsymbol{\beta}}^{(0)}(\tau) - \boldsymbol{\beta}(\tau) \right\} \right| \right\} \\
\leqslant & \left\{ \sqrt{\sum_{j \in S} \left| e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} e_j \right|^2} \left\| \widehat{\boldsymbol{\beta}}_1^{(0)}(\tau) - \boldsymbol{\beta}_1(\tau) \right\|_2 + \sqrt{\sum_{j \notin S} \left| e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} e_j \right|^2} \left\| \widehat{\boldsymbol{\beta}}_2^{(0)}(\tau) \right\|_2 \right\} \\
\leqslant & \frac{M}{\log n} \frac{n^{1+\zeta+\nu/2-\alpha}}{p} \gamma_n, \qquad (31)
\end{aligned}
$$

with probability at least

$$
\begin{aligned}
& 1 - O\left\{ n^{\nu} \exp\left(\frac{-Cn^{1-2\alpha}}{2\log n} \right) \right\} - P\left(\left\| \widehat{\boldsymbol{\beta}}^{(0)}(\tau) - \boldsymbol{\beta}(\tau) \right\|_2 > \gamma_n \right) - P\left(\left\| \widehat{\boldsymbol{\beta}}_2^{(0)}(\tau) \right\|_2 > 0 \right) \\
& = 1 - O\left\{ n^{\nu} \exp\left(\frac{-Cn^{1-2\alpha}}{2\log n} \right) \right\} - O\left\{ n^{-cs} \right\}, \qquad (32)
\end{aligned}
$$

where the last equation is by Lemma 2 and 3.

Choose the appropriate α, then it follows that

$$
P\left(|\eta_i| > \frac{M}{\log n} \frac{n^{1+\tau-\alpha}}{p^{1/2}} \gamma_n \right) = \exp\left(\frac{-Cn^{1-2\alpha}}{2\log n} \right), \qquad (33)
$$

where and is any constant that is less than C (C is also an arbitrary constant).

Next, for $i \in S$, it can be seen that

$$
\begin{aligned}
& \left| e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} \left\{ \widehat{\boldsymbol{\beta}}^{(0)}(\tau) - \boldsymbol{\beta}(\tau) \right\} \right| \\
\leqslant & \left\{ \sqrt{\sum_{j\in S} \left| e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} e_j \right|^2} \left\| \widehat{\boldsymbol{\beta}}_1^{(0)}(\tau) - \boldsymbol{\beta}_1(\tau) \right\|_2 + \sqrt{\sum_{j\notin S} \left| e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} e_j \right|^2} \left\| \widehat{\boldsymbol{\beta}}_2^{(0)}(\tau) \right\|_2 \right\} \\
\leqslant & \left\{ \sqrt{\frac{1}{2}\sum_{j\in S} \left| e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} e_i \right|^2 + \left| e_j^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} e_j \right|^2} \left\| \widehat{\boldsymbol{\beta}}_1^{(0)}(\tau) - \boldsymbol{\beta}_1(\tau) \right\|_2 \right. \\
& \left. + \sqrt{\sum_{j\notin S} \left| e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} e_j \right|^2} \left\| \widehat{\boldsymbol{\beta}}_2^{(0)}(\tau) \right\|_2 \right\} \\
\leqslant & c_2^{'} \frac{n^{1+\zeta+\nu/2}}{p} \gamma_n, \qquad (34)
\end{aligned}
$$

with probability at least $1 - O\left\{n^\nu \exp(-Cn)\right\} - O\left\{n^{-cs}\right\}$.

By now we have all the technical results needed to prove the main theorems. The proof of Theorem 1 follows the basic scheme of Wang and Leng (2016) but with a modification of their Theorem 1 by using our Lemma 4.

A.2 Proofs of Theorems

Proof of Theorem 1

In this section, we give rigorous proofs of our theorems. The framework of the proof follows Fan and Lv (2008), but with many modifications in details. Recall the proposed HOLP-Quantile screening estimator

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_i(\tau) &= e_j^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} Y^* \left\{ \widehat{\boldsymbol{\beta}}^0(\tau) \right\} \\
&= e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} \widehat{\boldsymbol{\beta}}^{(0)}(\tau) \\
&= e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} \boldsymbol{\beta}(\tau) + e_i^T \mathbb{X}^T \left(\mathbb{X}\mathbb{X}^T\right)^{-1} \mathbb{X} \left\{ \widehat{\boldsymbol{\beta}}^{(0)}(\tau) - \boldsymbol{\beta}(\tau) \right\} \\
&\triangleq \xi_i + \eta_i
\end{aligned}
$$

where ξ_i can be seen as the signal part and η_i the noise part. For ξ_i, the techniques are similar to Wang and Leng(2016) [21]. We only to check the proof and give the corresponding condition requirements. In this part, we aim to evaluate the magnitude of η_i.

Applying of (20) of Lemma 1 and (28) of Lemma 2 all $i \in S$, we have

$$
P\left(\min_{i\in S} |\xi_i| < C \frac{n^{1-\zeta-\kappa}}{p} \right) = O\left\{ s \cdot \exp\left(\frac{-Cn^{1-5\zeta-2\kappa-v}}{2\log n} \right) \right\} \qquad (35)
$$

and

$$
P\left(\max_{i\in S} |\eta_i| > c_2^{'} \frac{n^{1+\zeta+\nu/2}}{p} \gamma_n \right) = O\left\{ sn^\nu \exp(-Cn) \right\} + O\left\{ sn^{-cs} \right\} \qquad (36)
$$

Therefore, if we choose r_n such that

$$c_2^{'} \frac{n^{1+\zeta+\nu/2}}{p}\gamma_n < r_n < C\frac{n^{1-\zeta-\kappa}}{p}, \tag{37}$$

or in asymptotic form satisfying

$$\frac{pr_n}{n^{1-\zeta-\kappa}} \to 0 \quad \text{and} \quad \frac{pr_n}{n^{1+\zeta+\nu/2}\gamma_n} \to \infty, \tag{38}$$

then we have

$$\begin{aligned}
& p\left(\min_{i\in S}\left|\widehat{\boldsymbol{\beta}}_i(\tau)\right| < r_n\right) \\
&= p\left(\min_{i\in S}|\xi_i+\eta_i| < r_n\right) \\
&\leqslant P\left(\min_{i\in S}|\xi_i| < C\frac{n^{1-\zeta-\kappa}}{p}\right) + P\left(\max_{i\in S}|\eta_i| > c_2^{'}\frac{n^{1+\zeta+\nu/2}}{p}\gamma_n\right) \\
&= O\left\{\exp\left(\frac{-C_1 n^{1-5\zeta-2\kappa-v}}{2\log n}\right) + sn^{\nu}\exp\left(-Cn\right) + sn^{-cs}\right\} \to 0.
\end{aligned} \tag{39}$$

This completes the proof of Theorem 1.

Proof of Theorem 2

Let

$$\frac{\tilde{c}}{\sqrt{\log n}}\frac{n^{1-\zeta-\kappa}}{p} < r_n, \tag{40}$$

then we have that

$$\begin{aligned}
& p\left(\max_{i\notin S}\left|\widehat{\boldsymbol{\beta}}_i(\tau)\right| > r_n\right) \\
&= p\left(\max_{i\notin S}|\xi_i+\eta_i| > r_n\right) \\
&\leqslant P\left(\max_{i\notin S}|\xi_i| > \frac{\tilde{c}}{\sqrt{\log n}}\frac{n^{1-\zeta-\kappa}}{p}\right) + P\left(\max_{i\notin S}|\eta_i| > \frac{\sqrt{Cc_4}}{c_3^{'}}\frac{n^{1+\zeta+\nu/2-\alpha}}{(\log n)p}\gamma_n\right) \\
&= O\left\{p\exp\left(\frac{-Cn^{1-5\zeta-2\kappa-v}}{2\log n}\right) + pn^{\nu}\exp\left(\frac{-Cn^{1-2\alpha}}{2\log n}\right)\right\} + O\left\{pn^{-cs}\right\} \to 0.
\end{aligned} \tag{41}$$

Therefore, combing the above result with Theorem 1, we have

$$\begin{aligned}
& P\left(\max_{i\notin S}\left|\widehat{\boldsymbol{\beta}}_i(\tau)\right| < r_n < \min_{i\in S}\left|\widehat{\boldsymbol{\beta}}_i(\tau)\right|\right) \\
&= 1 - O\left\{\exp\left(\frac{-C_1 n^{1-5\zeta-2\kappa-v}}{2\log n}\right) + sn^{\nu}\exp\left(-Cn\right) + p\exp\left(\frac{-Cn^{1-5\zeta-2\kappa-v}}{2\log n}\right)\right. \\
&\qquad \left. + pn^{\nu}\exp\left(\frac{-Cn^{1-2\alpha}}{2\log n}\right) + pn^{-cs}\right\}
\end{aligned} \tag{42}$$

Obviously, if we choose a submodel with size d that $d \geqslant s$, we will have

$$P\left(\mathcal{M}_S \subset \mathcal{M}_d\right) \to 1, \tag{43}$$

which completes the proof of Theorem 2.

Proof of Theorem 3

(i) To estimate $\inf_{k\in[1,s]}(QL_{k-1}-QL_k)$, we decompose $QL_{k-1}-QL_k$ as

$$QL_{k-1}-QL_k=(QL_{k-1}-QL_{k-1}^{**})-(QL_k-QL_k^{**})+(QL_{k-1}^{**}-QL_k^{**}). \tag{44}$$

(I) Estimation of $QL_k-QL_k^{**}$. We first prove

$$\sup_{k\in[1,s]} n^{1/2}k^{-1/2}s^{-1/2}(\widehat{\beta}_{(k)}(\tau)-\beta_{(k)}^{**}(\tau))=O_p(1). \tag{45}$$

Imitating Theorem 3.1(i) of [19], by condition 7 and 10, for any $\varepsilon>0$, we can find a large constant C independent of k such that

$$P\left[\inf_{\|v_k\|=C} QL_k\left\{\beta_{(k)}^{**}(\tau)+n^{-1/2}k^{1/2}s^{1/2}v_k\right\}>QL_k\left\{\beta_{(k)}^{**}(\tau)\right\},k=1,\ldots,s\right]>1-\varepsilon.$$

So, with probability approaching to 1, for each $k\in[1,s]$, there is a local minimum in the ball

$$\left\{\beta_{(k)}^{**}(\tau)+n^{-1/2}k^{1/2}s^{1/2}v_k:\|v_k\|\le C\right\}.$$

By the convexity of QL_k, $\widehat{\beta}_{(k)}(\tau)$ is also the global minimum , and hence

$$n^{1/2}k^{-1/2}s^{-1/2}\|\widehat{\beta}_{(k)}(\tau)-\beta_{(k)}^{**}(\tau)\|\le C,$$

which confirms (45).

Denote $u_i=\mathbf{X}_i^T\boldsymbol{\beta}^*(\tau)$ and $u_{i(k)}=\mathbf{X}_{i(k)}^T\boldsymbol{\beta}_{(k)}^{**}(\tau)$. By Knight's identity [15], we have

$$\begin{aligned}
&QL_k-QL_k^{**}\\
=\;&\sum_{i=1}^n\rho_\tau\{Y_i-\mathbf{X}_{i(k)}^T\widehat{\boldsymbol{\beta}}_{(k)}(\tau)\}-\sum_{i=1}^n\rho_\tau\{Y_i-\mathbf{X}_{i(k)}^T\boldsymbol{\beta}_{(k)}^{**}(\tau)\}\\
=\;&\sum_{i=1}^n(\rho_\tau[\varepsilon_i(\tau)+u_i-u_{i(k)}+\mathbf{X}_{i(k)}^T\{\boldsymbol{\beta}_{(k)}^{**}(\tau)-\widehat{\boldsymbol{\beta}}_{(k)}(\tau)\}]-\rho_\tau\{\varepsilon_i(\tau)+u_i-u_{i(k)}\})\\
=\;&\sum_{i=1}^n\mathbf{X}_{i(k)}^T\{\boldsymbol{\beta}_{(k)}^{**}(\tau)-\widehat{\boldsymbol{\beta}}_{(k)}(\tau)\}\psi_\tau\{\varepsilon_i(\tau)+u_i-u_{i(k)}\}+\sum_{i=1}^n\int_0^{-\mathbf{X}_{i(k)}^T\{\boldsymbol{\beta}_{(k)}^{**}(\tau)-\widehat{\boldsymbol{\beta}}_{(k)}(\tau)\}}\alpha_{i(k)}(x)dx\\
=\;&\{\boldsymbol{\beta}_{(k)}^{**}(\tau)-\widehat{\boldsymbol{\beta}}_{(k)}(\tau)\}^T\sum_{i=1}^n\psi_\tau\{\varepsilon_i(\tau)+u_i-u_{i(k)}\}\mathbf{X}_{i(k)}+\sum_{i=1}^n\int_0^{-\mathbf{X}_{i(k)}^T\{\boldsymbol{\beta}_{(k)}^{**}(\tau)-\widehat{\boldsymbol{\beta}}_{(k)}(\tau)\}}\alpha_{i(k)}(x)dx\\
\doteq\;&Z_{n(k),1}+Z_{n(k),2},
\end{aligned}$$

where $\alpha_{i(k)}(x)=I\{\varepsilon_i(\tau)+u_i-u_{i(k)}\le x\}-I\{\varepsilon_i(\tau)+u_i-u_{i(k)}\le 0\}$.

We first consider $Z_{n(k),1}$. The first order condition for the population minimization problem implies that

$$E\{\psi_\tau(\varepsilon_i(\tau)+u_i-u_{i(k)})\mathbf{X}_{i(k)}\}=0.$$

For the variance of $Z_{n(k),1}$,

$$\mathrm{var}\Big[\sum_{i=1}^n\psi_\tau\{\varepsilon_i(\tau)+u_i-u_{i(k)}\}\mathbf{X}_{i(k)}\Big]\le\sum_{i=1}^n E[\psi_\tau^2\{\varepsilon_i(\tau)+u_i-u_{i(k)}\}\mathbf{X}_{i(k)}\mathbf{X}_{i(k)}^T]\le\lambda_{\max}(\Sigma_{(k)})\cdot n.$$

Hence, by (45),

$$
\begin{aligned}
\sup_{k\in[1,s]}|Z_{n(k),1}| &= \sup_{k\in[1,s]}\|\widehat{\beta}_{(k)}(\tau)-\beta^{**}_{(k)}(\tau)\| \sup_{k\in[1,s]}\left\|\sum_{i=1}^{n}\psi_\tau\{\varepsilon_i(\tau)+u_i-u_{i(k)}\}\mathbf{X}_{i(k)}\right\| \\
&\le O_p\left(sn^{-1/2}\right)\cdot O\left(\lambda^{1/2}_{\max}(\Sigma_{(s)})(ns)^{1/2}\right)=O_p\left(\lambda^{1/2}_{\max}(\Sigma_{(s)})s^{3/2}\right) \qquad (46)
\end{aligned}
$$

Next we consider $Z_{n(k),2}$. Define $E_n(C)=\{\sup_{k\in[1,s]}\|\boldsymbol{\beta}^{**}_{(k)}(\tau)-\widehat{\boldsymbol{\beta}}_{(k)}(\tau)\|\le Csn^{-1/2}\}$, we have

$$
|Z_{n(k),2}|\le\Big|\sum_{i=1}^{n}\Big[\int_0^{-\|\mathbf{X}^T_{i(k)}\|Csn^{-1/2}}+\int_0^{\|\mathbf{X}^T_{i(k)}\|Csn^{-1/2}}\alpha_{i(k)}(x)dx\Big]\Big|\doteq|T_{n(k),2}|,\quad\text{on } E_n(C),
$$

$$
\begin{aligned}
|E[T_{n(k),2}]| &= \Big|E\sum_{i=1}^{n}\Big[\int_0^{-\|\mathbf{X}^T_{i(k)}\|Csn^{-1/2}}F(-u_i+u_{i(k)}+x\big|\mathbf{X}_i)-F(-u_i+u_{i(k)}\big|\mathbf{X}_i)dx\Big]\Big| \\
&\quad+\Big|E\sum_{i=1}^{n}\Big[\int_0^{\|\mathbf{X}^T_{i(k)}\|Csn^{-1/2}}F(-u_i+u_{i(k)}+x\big|\mathbf{X}_i)-F(-u_i+u_{i(k)}\big|\mathbf{X}_i)dx\Big]\Big| \\
&\le C^2f_{\max}s^2n^{-1}\Big|E\sum_{i=1}^{n}\|\mathbf{X}_{i(k)}\|^2\Big|=C^2f_{\max}s^2\cdot\mathrm{tr}(\Sigma_{(k)})),
\end{aligned}
$$

which implies $\sup_{k\in[1,s]}|T_{n(k),2}|=O_p(s^3\mathrm{tr}(\Sigma_{(s)}))$. Note that (45) infers for any $\varepsilon>0$, there is C such that

$$
\liminf_{n\to\infty}P(E_n(C))>1-\varepsilon,
$$

thus $\sup_{k\in[1,s]}|Z_{n(k),2}|=O_p(s^3\mathrm{tr}(\Sigma_{(s)}))$ and then

$$
\sup_{k\in[1,s]}|QL_k-QL^{**}_k|=O_p(\lambda^{1/2}_{\max}(\Sigma_{(s)})s^{3/2}+s^3\mathrm{tr}(\Sigma_{(s)})).
$$

(II) Estimation of $QL^{**}_{k-1}-QL^{**}_k$. Calculate

$$
\begin{aligned}
&\rho_\tau\{Y_i-\mathbf{X}^T_{i(k-1)}\boldsymbol{\beta}^{**}_{(k-1)}(\tau)\}-\rho_\tau\{Y_i-\mathbf{X}^T_{i(k)}\boldsymbol{\beta}^{**}_{(k)}(\tau)\} \\
=\;&\{\mathbf{X}^T_{i(k)}\boldsymbol{\beta}^{**}_{(k)}(\tau)-\mathbf{X}^T_{i(k-1)}\boldsymbol{\beta}^{**}_{(k-1)}(\tau)\}\psi_\tau\{Y_i-\mathbf{X}^T_{i(k)}\boldsymbol{\beta}^{**}_{(k)}(\tau)\} \\
&+\int_0^{-\{\mathbf{X}^T_{i(k)}\boldsymbol{\beta}^{**}_{(k)}(\tau)-\mathbf{X}^T_{i(k-1)}\boldsymbol{\beta}^{**}_{(k-1)}(\tau)\}}\Big[I\{Y_i-\mathbf{X}^T_{i(k)}\boldsymbol{\beta}^{**}_{(k)}(\tau)\le x\}-I\{Y_i-\mathbf{X}^T_{i(k)}\boldsymbol{\beta}^{**}_{(k)}(\tau)\le 0\}\Big]dx \\
=\;&\Big(\boldsymbol{\beta}^{**}_{(k)}(\tau)^T\mathbf{X}_{i(k)}\psi_\tau\{Y_i-\mathbf{X}^T_{i(k)}\boldsymbol{\beta}^{**}_{(k)}(\tau)\}-\boldsymbol{\beta}^{**}_{(k-1)}(\tau)^T\mathbf{X}_{i(k-1)}\psi_\tau\{Y_i-\mathbf{X}^T_{i(k)}\boldsymbol{\beta}^{**}_{(k)}(\tau)\}\Big) \\
&+\int_0^{-\{\mathbf{X}^T_{i(k)}\boldsymbol{\beta}^{**}_{(k)}(\tau)-\mathbf{X}^T_{i(k-1)}\boldsymbol{\beta}^{**}_{(k-1)}(\tau)\}}\Big[I\{Y_i-\mathbf{X}^T_{i(k)}\boldsymbol{\beta}^{**}_{(k)}(\tau)\le x\}-I\{Y_i-\mathbf{X}^T_{i(k)}\boldsymbol{\beta}^{**}_{(k)}(\tau)\le 0\}\Big]dx \\
\doteq\;&L_{1,i}(k)+L_{2,i}(k) \qquad (47)
\end{aligned}
$$

Note that $EL_{1,i}(k)=0$, then by chebyshev's theorem and condition 6, for any $\varepsilon>0$,

$$
\begin{aligned}
P\left(\sup_{k\in[1,s]}\left|\sum_{i=1}^{n}L_{1,i}(k)\right|>\varepsilon^{-1}n^{1/2}s^{1/2}\right) &\le \sum_{k=1}^{s}\cdot P\left(\left|\sum_{i=1}^{n}L_{1,i}(k)\right|>\varepsilon^{-1}n^{1/2}s^{1/2}\right) \\
&\le \frac{\varepsilon^2\sum_{k=1}^{s}\sum_{i=1}^{n}\mathrm{Var}(L_{1,i}(k))}{ns}\le 2c'\varepsilon^2, \qquad (48)
\end{aligned}
$$

$$
\begin{aligned}
& P\left(\sup_{k\in[1,s]}\left|\sum_{i=1}^{n} L_{2,i}(k)-\sum_{i=1}^{n} EL_{2,i}(k)\right|>\varepsilon^{-1}n^{1/2}s^{1/2}\right)\\
\le\ & \frac{\varepsilon^2\sum_{k=1}^{s}\sum_{i=1}^{n}\mathrm{Var}(L_{2,i}(k))}{ns}\le\frac{\varepsilon^2\sum_{k=1}^{s}\sum_{i=1}^{n}E(L_{2,i}^2(k))}{ns}\\
\le\ & \frac{\varepsilon^2}{ns}\sum_{k=1}^{s}\sum_{i=1}^{n}E\int_0^{-\{\mathbf{X}_{i(k)}^T\boldsymbol{\beta}_{(k)}^{**}(\tau)-\mathbf{X}_{i(k-1)}^T\boldsymbol{\beta}_{(k-1)}^{**}(\tau)\}}[I\{Y_i-\mathbf{X}_{i(k)}^T\boldsymbol{\beta}_{(k)}^{**}(\tau)\le x\}\\
& -I\{Y_i-\mathbf{X}_{i(k)}^T\boldsymbol{\beta}_{(k)}^{**}(\tau)\le 0\}]^2dx\cdot E|\mathbf{X}_{i(k)}^T\boldsymbol{\beta}_{(k)}^{**}(\tau)-\mathbf{X}_{i(k-1)}^T\boldsymbol{\beta}_{(k-1)}^{**}(\tau)|\\
\le\ & \frac{\varepsilon^2}{ns}\sum_{k=1}^{s}\sum_{i=1}^{n}E|\mathbf{X}_{i(k)}^T\boldsymbol{\beta}_{(k)}^{**}(\tau)-\mathbf{X}_{i(k-1)}^T\boldsymbol{\beta}_{(k-1)}^{**}(\tau)|^2\le 2c'\varepsilon^2, \qquad (49)
\end{aligned}
$$

so

$$
\sup_{k\in[1,s]}\left(\left|\sum_{i=1}^{n} L_{2,i}(k)-\sum_{i=1}^{n} EL_{2,i}(k)\right|+\left|\sum_{i=1}^{n} L_{1,i}(k)\right|\right)=O_p(n^{1/2}s^{1/2}). \qquad (50)
$$

Moreover,

$$
\begin{aligned}
& \sum_{i=1}^{n} EL_{2,i}(k)\\
=\ & \sum_{i=1}^{n} E\int_0^{-\{\mathbf{X}_{i(k)}^T\boldsymbol{\beta}_{(k)}^{**}(\tau)-\mathbf{X}_{i(k-1)}^T\boldsymbol{\beta}_{(k-1)}^{**}(\tau)\}}\Big[F\left(x-u_i+\mathbf{X}_{i(k)}^T\boldsymbol{\beta}_{(k)}^{**}(\tau)|\mathbf{X}_{i(k)}\right)\\
& -F\left(-u_i+\mathbf{X}_{i(k)}^T\boldsymbol{\beta}_{(k)}^{**}(\tau)|\mathbf{X}_{i(k)}\right)\Big]\,dx\\
\ge\ & \frac{1}{2}\sum_{i=1}^{n}\left\{\boldsymbol{\beta}_{(k)}^{**}(\tau)-\left(\boldsymbol{\beta}_{(k-1)}^{**}(\tau)^T,0\right)^T\right\}^T E\left\{f_{\min}\mathbf{X}_{i(k)}\mathbf{X}_{i(k)}^T\right\}\left\{\boldsymbol{\beta}_{(k)}^{**}(\tau)-\left(\boldsymbol{\beta}_{(k-1)}^{**}(\tau)^T,0\right)^T\right\}\\
\ge\ & \frac{f_{\min}\lambda_{\min}(\Sigma_{(k)})n\epsilon_n^2}{2}.
\end{aligned}
$$

So, we have

$$
\inf_{k\in[1,s]}(QL_{k-1}^{**}-QL_k^{**})\ge\frac{f_{\min}\lambda_{\min}(\Sigma_{(s)})n\epsilon_n^2}{2}+O_p(n^{1/2}s^{1/2}), \qquad (51)
$$

by combining the bound for $QL_k-QL_k^{**}$, $QL_{k-1}-QL_{k-1}^{**}$ and $QL_{k-1}^{**}-QL_k^{**}$, we have,

$$
\inf_{k\in[1,s]}(QL_{k-1}-QL_k)\ge\frac{f_{\min}\lambda_{\min}(\Sigma_{(s)})n\epsilon_n^2}{2}+O_p(n^{1/2}s^{1/2})+O_p(\lambda_{\max}^{1/2}(\Sigma_{(s)})s^{3/2}+s^3\mathrm{tr}(\Sigma_{(s)})). \qquad (52)
$$

Since $\mathrm{tr}(\Sigma_{(s)})=s$, $\lambda_{\max}(\Sigma_{(s)}\ge 1$, by condition 8 and 10, $O(\lambda_{\min}^{1/2}(\Sigma_{(s)}))\ge s^{-1}\lambda_{\max}(\Sigma_{(s)})$, $o(\epsilon_n^2)=n^{-1}s^{9/2}+n^{-1/2}s$, there is constant $c_0>0$ such that

$$
\inf_{k\in[1,s]}(QL_{k-1}-QL_k)\ge c_0(s^4+s^{1/2}n^{1/2}).
$$

(ii) if $\mathbf{S}\subseteq\widehat{\mathbf{S}}_{k-1}$, $k\in(s,q]$, with the definition of $\boldsymbol{\beta}_{(k)}^{**}(\tau)$ and $\boldsymbol{\beta}_{(k-1)}^{**}(\tau)$, we have $\boldsymbol{\beta}_{(k)}^{**}(\tau)=(\boldsymbol{\beta}_{(k-1)}^{**}(\tau)^T,0)^T$. In order to find the threshold of $QL_{k-1}-QL_k$ in this case, we adopt the idea of likelihood ratio test and sketch an alternative method of proof using frameworks of [16] and [17]. $\widehat{\boldsymbol{\beta}}_{(k-1)}$ can be seen as the restricted estimator to the problem

$$\min_{\boldsymbol{\beta}_{(k-1)} \in R^{k-1}} \sum_{i=1}^{n} \rho_\tau \{Y_i - \mathbf{X}_{i(k)}^T (\boldsymbol{\beta}_{(k-1)}^T, 0)^T\}.$$

Let

$$\mathbf{V}_{n(k)}(\delta_{(k)}, \tau) = \sum_{i=1}^{n} \rho_\tau \{\varepsilon_i(\tau) - w(\tau)\mathbf{X}_{i(k)}^T \delta_{(k)}/\sqrt{n}\},$$

where $w(\tau) = \lambda(\tau)s(\tau)$, $\lambda^2(\tau) = \tau(1-\tau)$ and $s(\tau) = 1/f(F^{-1}(\tau)) = 1/f(0)$. Define the quadratic objective function

$$\mathbf{Q}_{n(k)}(\delta_{(k)}, \tau) = \frac{1}{2} w^{-1}(\tau) \delta_{(k)}^T D_{(k)} \delta_{(k)} - \delta_{(k)}^T g_{n(k)}(\tau),$$

where $D_{(k)} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_{i(k)}\mathbf{X}_{i(k)}^T$ and $g_{n(k)}(\tau) = n^{-1/2}\sum_{i=1}^{n} \mathbf{X}_{i(k)}\psi_\tau\{\varepsilon_i(\tau)\}$. Set

$$\mathbf{W}_{n(k)}(\delta_{(k)}, \tau) = \{\lambda^2(\tau)s(\tau)\}^{-1}\{\mathbf{V}_{n(k)}(\delta_{(k)}, \tau) - \mathbf{V}_{n(k)}(0, \tau)\} - \mathbf{Q}_{n(k)}(\delta_{(k)}, \tau).$$

Define

$$\hat{\delta}_{(k)}(\tau) = \sqrt{n}\{\widehat{\boldsymbol{\beta}}_{(k)}(\tau) - \boldsymbol{\beta}_{(k)}^{**}(\tau)/w(\tau)\}.$$

We assert that for any $C > 0$,

$$\sup_{k \in [s+1, q]} \sup_{\|t\| \le Cq \log n} \left|\mathbf{W}_{n(k)}(t, \tau)\right| = O_p(q(q-s)n^{\mu - 1/2}\log^2 n). \tag{53}$$

To prove (53), we wish to show that for any $\kappa > 0$ and fixed t,

$$P\{|\mathbf{W}_{n(k)}(t, \tau)| \ge (\kappa + 1)B_n | F_n\} \le K n^{-\kappa},$$

with a fixed $K > 0$ and $B_n = qn^{\mu - 1/2}\log^2 n$, where $F_n = \{\max_{i \in [1,n]} \|\mathbf{X}_{i(q)}\| \le n^\mu\}$. To do this, we will use the Markov inequality

$$P\{|\mathbf{W}_{n(k)}(t, \tau)| \ge s_n' | F_n\} \le \exp(-v s_n')\{M(v) + M(-v)\},$$

where $M(v) = E[\exp(v\mathbf{W}_{n(k)}(t, \tau))|F_n]$. Similar to [13], denote

$$\begin{aligned} R_i(t, \tau) \quad &= \quad \{\lambda^2(\tau)s(\tau)\}^{-1}\{\rho_\tau\{\varepsilon_i(\tau) - w(\tau)\mathbf{X}_{i(k)}^T t/\sqrt{n}\} - \rho_\tau\{\varepsilon_i(\tau)\}\} \\ &\quad - \left(\frac{1}{2} n^{-1}(\mathbf{X}_{i(k)}^T t)^2 - \lambda^{-1}(\tau) t^T n^{-1/2} \mathbf{X}_{i(k)} \psi_\tau\{\varepsilon_i(\tau)\}\right). \end{aligned}$$

For $\|t\| \le Cq \log n$, we obtain

$$\left| R_i(t, \tau) + \frac{1}{2} n^{-1}(\mathbf{X}_{i(k)}^T t)^2 \right| \le \lambda^{-1}(\tau)\|\mathbf{X}_{i(k)}\| n^{-1/2}\|t\| \le \lambda^{-1}(\tau)\|\mathbf{X}_{i(k)}\| n^{-1/2} q \log n,$$

and for some constant c,

$$\begin{aligned} E\{R_i(t,\tau)|F_n\} &\le c(n^{-\frac{3}{2}}E[\|\mathbf{X}_{i(k)}\|^3|F_n]\|t\|^3+n^{-2}E[\|\mathbf{X}_{i(k)}\|^4|F_n]\|t\|^4) \\ &< c(1+C)^4(n^{3\mu-3/2}q^2+n^{4\mu-2}q^4)\log^4 n, \end{aligned}$$

$$\mathrm{Var}\{R_i(t,\tau)|F_n\} \le cn^{-\frac{3}{2}}E[\|\mathbf{X}_{i(k)}\|^3|F_n]\|t\|^3 \le c(1+C)^3n^{3\mu-3/2}q^3\log^3 n.$$

Hence

$$\sum_{i=1}^n E\{R_i(t,\tau)|F_n\} \le c(1+C)^4(n^{3\mu-1/2}q^2+n^{4\mu-1}q^4)\log^4 n,$$

and

$$\sum_{i=1}^n \mathrm{Var}\{R_i(t,\tau)|F_n\} \le c(1+C)^3n^{3\mu-1/2}q^3\log^3 n.$$

Let $v=\log n/B_n$, then $vR_i=o(1)$ on set F_n(note that $|R_i|<\lambda^{-1}(\tau)n^{\mu-\frac{1}{2}}q$ on F_n), and Taylor series expansion yields

$$\exp\{vR_i(t,\tau)\} \le 1+vR_i(t,\tau)+\frac{3}{2}v^2R_i^2(t,\tau), \quad \text{on } F_n,$$

and then for some constant K',

$$\begin{aligned} M(v) &= E\left[\prod_{i=1}^n \exp\{vR_i(t,\tau)\}|F_n\right] \le 1+2\sum_{i=1}^n E\left[vR_i(t,\tau)+\frac{3}{2}v^2R_i^2(t,\tau)|F_n\right] \\ &\le 1+2v\sum_{i=1}^n E[R_i(t,\tau)|F_n]+3v^2\sum_{i=1}^n \mathrm{Var}\{R_i(t,\tau)|F_n\}+3v^2\sum_{i=1}^n E^2[R_i(t,\tau)|F_n] \\ &\le K'(n^{\mu+1/2}q+n^{2\mu}q+n^{3\mu-1/2}q^3+n^{4\mu-1}q^2+n^{6\mu-2}q^6)\log^6 n \le 5K'n^{6\mu+8}. \end{aligned} \tag{54}$$

So we have

$$\begin{aligned} P\left\{\left|\mathbf{W}_{n(k)}(t,\tau)\right| \ge (\kappa+9+6\mu)B_n|F_n\right\} &\le \exp\{-(\kappa+9+6\mu)\log n\}\cdot 5K'n^{6\mu+8} \\ &\le 5K'n^{-\kappa}. \end{aligned}$$

Next, choosing balls of $n^{-2-\mu}$ covering $\{t:\|t\|\le Cq\log n\}$, consider $\{t_1,t_2\}$ lie in one of the balls, which is denoted by S_v, then on F_n we deduce

$$\begin{aligned} &\sup_{S_v}\left|\mathbf{W}_{n(k)}(t_1,\tau)-\mathbf{W}_{n(k)}(t_2,\tau)\right| \\ \le\ &\sup_{S_v}\left\{\frac{1}{2n}\sum_{i=1}^n|(\mathbf{X}_{i(k)}^Tt_1)^2-(\mathbf{X}_{i(k)}^Tt_2)^2|+\sum_{i=1}^n\lambda^{-1}(\tau)\|t_1-t_2\|n^{-\frac{1}{2}}\|\mathbf{X}_{i(k)}\|\right\} \\ \le\ &Cn^{\mu-1}q\log n+\lambda^{-1}(\tau)n^{-3/2} \le (C+\lambda^{-1}(\tau))n^{-1/2}B_n, \end{aligned} \tag{55}$$

Since the number of sets needed to cover the set $\{t:\|t\|\le Cq\log n\}$ is bounded by $n^{(3+\mu)q}$, we calculate (let $\kappa=(4+\mu)q$)

$$P\left(\sup_{\|t\|\le Cq}\left|\mathbf{W}_{n(k)}(t,\tau)\right| \ge (\kappa+9+6\mu)B_n+(C+\lambda^{-1}(\tau))n^{-1/2}B_n|F_n\right) \le 5K'n^{(3+\mu)q}n^{-\kappa},$$

so

$$
\begin{aligned}
& P\left(\sup_{k\in[s+1,q]}\sup_{\|t\|\le Cq\log n}\left|\mathbf{W}_{n(k)}(t,\tau)\right| \ge (\kappa+1)(q-s)B_n + (C+\lambda^{-1}(\tau))(q-s)n^{-1/2}B_n|F_n\right) \\
\le\ & 5K'qn^{(3+\mu)q}n^{-\kappa} \qquad (56)
\end{aligned}
$$

note that $P(F_n^c) = o(n^{-1}(p-s)^{-2})$, we can obtain

$$
\sup_{k\in[s+1,q]}\sup_{\|t\|\le Cq\log n}\left|\mathbf{W}_{n(k)}(t,\tau)\right| = O_p(\kappa(q-s)B_n) = O_p(q^2(q-s)n^{\mu-1/2}\log^2 n).
$$

Similar to (45), we have

$$
\sup_{k\in[1,q]} n^{1/2}k^{-1/2}q^{-1/2}(\widehat{\beta}_{(k)}(\tau) - \beta^{**}_{(k)}(\tau)) = O_p(1), \qquad (57)
$$

which together with (53) leads to

$$
\sup_{k\in[s+1,q]} \mathbf{W}_{n(k)}(\hat{\delta}_{(k)}(\tau),\tau) = O_p(q^2(q-s)n^{\mu-1/2}\log^2 n). \qquad (58)
$$

Next, for the $\widehat{\boldsymbol{\beta}}_{(k-1)}(\tau)$, we have

$$
\hat{\delta}_{(k-1)}(\tau) = D^{-1}_{(k-1)}g_{n(k-1)}(\tau)/\lambda(\tau).
$$

Now substituting back into the quadratic approximation, following [16], we obtain that

$$
\begin{aligned}
& 2(QL_{k-1} - QL_k)/\{\lambda(\tau)w(\tau)\} \\
=\ & 2\{\mathbf{V}_{n(k)}\big((\hat{\delta}_{(k-1)}(\tau)^T,0)^T,\tau\big) - \mathbf{V}_{n(k)}(\hat{\delta}_{(k)}(\tau),\tau)\}/\{\lambda(\tau)w(\tau)\} \\
=\ & \{\hat{\delta}^T_{(k-1)}(\tau)D_{(k-1)}\hat{\delta}_{(k-1)}(\tau) - 2\hat{\delta}^T_{(k-1)}(\tau)g_{n(k-1)}(\tau)/\lambda(\tau)\} \\
& -\{\hat{\delta}^T_{(k)}(\tau)D_{(k)}\hat{\delta}_{(k)}(\tau) - 2\hat{\delta}^T_{(k)}(\tau)g_{n(k)}(\tau)/\lambda(\tau)\} + (\mathbf{W}_{n(k-1)}(\hat{\delta}_{(k-1)}(\tau),\tau) - \mathbf{W}_{n(k)}(\hat{\delta}_{(k)}(\tau),\tau)) \\
=\ & h_{(k)}(\tau)^T D^{(k)}h_{(k)}(\tau) + (\mathbf{W}_{n(k-1)}(\hat{\delta}_{(k-1)}(\tau),\tau) - \mathbf{W}_{n(k)}(\hat{\delta}_{(k)}(\tau),\tau)),
\end{aligned}
$$

where $h_{(k)}(\tau) = \{\big(g_{n(k)}(\tau)\big)_{k,1} - D_{(k,k-1)}D^{-1}_{(k-1)}g_{n(k-1)}(\tau)\}/\lambda(\tau)$, $(A)_{i,j}$ means the i,jth element of matrix A, $D_{(k,k-1)} = \Big(\big(D_{(k)}\big)_{k,1},\big(D_{(k)}\big)_{k,2},\cdots,\big(D_{(k)}\big)_{k,k-1}\Big)$, and $D^{(k)} = \big(D^{-1}_{(k)}\big)_{k,k}$.

By (58),

$$
\sup_{k\in[s+1,q]} |2(QL_{k-1} - QL_k)/\{\lambda(\tau)w(\tau)\} - h_{(k)}(\tau)^T D^{(k)}h_{(k)}(\tau)| = O_p(q^2(q-s)n^{\mu-1/2}\log^2 n) \qquad (59)
$$

Denote

$$
Z_{i,k} = \lambda^{-1}(\tau)(D^{(k)})^{1/2}(-D_{(k,k-1)}D^{-1}_{(k-1)},1)\mathbf{X}_{i(k)}\psi_\tau\{\varepsilon_i(\tau)\}, \qquad (60)
$$

rewrite

$$
(D^{(k)})^{1/2}h_{(k)}(\tau) = \lambda^{-1}(\tau)(D^{(k)})^{1/2}(-D_{(k,k-1)}D^{-1}_{(k-1)},1)n^{-1/2}\sum_{i=1}^n \mathbf{X}_{i(k)}\psi_\tau\{\varepsilon_i(\tau)\} = \sum_{i=1}^n Z_{i,k}.
$$

Note that $E[Z_i]=0$ and $\sum_{i=1}^n E[Z_i^2]=1$, Theorem 1.1 in [4] gives us

$$\sup_t \left| P\left(\sum_{i=1}^n Z_i \le t\right) - \Phi(t)\right| \le c' \sum_{i=1}^n E|Z_{i,k}|^3. \tag{61}$$

for an absolute positive constant c'.

Define $Q_k = \{\| \sum_{i=1}^n \mathbf{X}_{i(k)}\mathbf{X}_{i(k)}^T - n\Sigma_{(k)}\| > nq^{-1/2}\log^{-1} n\}$ and $w = (-D_{(k,k-1)}D_{(k-1)}^{-1}, 1)^T$, then

$$\sum_{i=1}^n E|Z_{i,k}|^3 \le \sum_{i=1}^n E|Z_{i,k}|^3 (I_{F_n\cap Q_k^c} + I_{F_n^c} + I_{Q_k}), \tag{62}$$

and on set Q_k^c, by condition 6 we know

$$n = \sum_{i=1}^n w^T \mathbf{X}_{i(k)}\mathbf{X}_{i(k)}^T w \ge nw^T\Sigma_{(k)}w - nq^{-1/2}\log^{-1} n\|w\|^2 > \frac{nq^{-1/2}}{2}\|w\|^2 \tag{63}$$

combine (63) with $\sum_{i=1}^n Z_i^2 \le \lambda^{-1}(\tau)$,

$$\begin{aligned}\sum_{i=1}^n E|Z_{i,k}|^3 I_{F_n\cap Q_k^c} \le n^{-1/2}E\|w\| \max_{i\in[1,n]} \|\mathbf{X}_{i(k)}\| \sum_{i=1}^n |Z_{i,k}|^2 I_{F_n^c\cap Q_k^c} &\le 2q^{1/4}n^{-1/2}En^{\mu}\lambda^{-1}(\tau)I_{F_n^c\cap Q_k^c} \\ &\le 2\lambda^{-1}(\tau)n^{\mu-1/2}. \end{aligned}\tag{64}$$

Moreover, by Cauchy–Schwarz inequality,

$$\sum_{i=1}^n E|Z_{i,k}|^3 I_{F_n^c} \le \lambda^{-1}(\tau)E\sum_{i=1}^n |Z_{i,k}|I_{F_n^c} \le \lambda^{-1}(\tau)\sqrt{E\sum_{i=1}^n Z_{i,k}^2 nP(F_n)} = \lambda^{-1}(\tau)n^{1/2}P^{1/2}(F_n^c) \tag{65}$$

and

$$\sum_{i=1}^n E|Z_{i,k}|^3 I_{Q_k} \le \lambda^{-1}(\tau)E(\sum_{i=1}^n |Z_{i,k}|)I_{Q_k} \le \lambda^{-1}(\tau)\sqrt{E\sum_{i=1}^n Z_{i,k}^2 nP(Q_k)} = \lambda^{-1}(\tau)n^{1/2}P^{1/2}(Q_k). \tag{66}$$

By $1-\Phi(\log^{1/2} n) \le \exp\{-\log n\}/\log^{1/2} n$ and condition 9,

$$\begin{aligned} & P(\sup_{k\in[s+1,q]} |(D^{(k)})^{1/2}h_{(k)}(\tau)| > (q-s)\log^{1/2} n) \le \sum_{k=s+1}^q P(|(D^{(k)})^{1/2}h_{(k)}(\tau)| > \log^{1/2} n) \\ \le\ & (q-s)(2\lambda^{-1}(\tau)n^{\mu-1/2} + \lambda^{-1}(\tau)n^{1/2}P^{1/2}(F_n^c) + \lambda^{-1}(\tau)n^{1/2}P^{1/2}(Q_q)) \\ & +(q-s)(1-\Phi(\log^{1/2} n)) = o(1) \end{aligned}\tag{67}$$

so

$$\sup_{k\in[s+1,q]} |QL_{k-1} - QL_k| = o_p((q-s)^2\log n) + O_p(q^2(q-s)n^{\mu-1/2}\log^2 n). \tag{68}$$

(iii) In view of the proof in (iii), it is easy to see there is constant $c''>0$ such that

$$\lim_{n\to\infty} P(\sup_{k\in[s+1,q]} \mathbf{W}_{n(k)}(\hat{\delta}_{(k)}(\tau),\tau) > c''q^2(q-s)n^{\mu-1/2}\log^2 n) = 0. \tag{69}$$

then

$$\lim_{n\to\infty} P(|2(QL_{k-1} - QL_k)/\{\lambda(\tau)w(\tau)\} - h_{(k)}(\tau)^T D^{(k)} h_{(k)}(\tau)| > 2c'' q^2(q-s)n^{\mu-1/2}\log^2 n) = 0.$$

Theorem 1.1 in [4] and (64)–(66) show that for $k \in (s, q]$, $\Delta_{CI}((\sum_{i=1}^n Z_{i,k})^2, \chi_1^2) \le c' \sum_{i=1}^n E|Z_{i,k}|^3$ and $\sum_{i=1}^n E|Z_{i,k}|^3 \le 2\lambda^{-1}(\tau)n^{\mu-1/2} + \lambda^{-1}(\tau)n^{1/2}P^{1/2}(F_n^c) + \lambda^{-1}(\tau)n^{1/2}P^{1/2}(Q_k)$, so by (67), $\sup_{k\in[s+1,q]} \Delta_{CI}((\sum_{i=1}^n Z_{i,k})^2, \chi_1^2) = o(1)$, as $n \to \infty$. Note that $h_{(k)}(\tau)^T D^{(k)} h_{(k)}(\tau) = (\sum_{i=1}^n Z_{i,k})^2$ and $q^2(q-s)n^{\mu-1/2}\log^2 n = o(1)$, we have

$$\sup_{k\in[s+1,q]} \Delta_{CI}(2(QL_{k-1} - QL_k)/\{\lambda(\tau)w(\tau)\}, \chi_1^2) = o(1), \quad \text{as } n \to \infty. \tag{70}$$

References

[1] Maria Florina Balcan, Avrim Blum, Shai Fine, and Yishay Mansour. Distributed learning, communication complexity and privacy. *Conference on Learning Theory*, 26(1):1–22, 2012.

[2] Heather Battey, Jianqing Fan, Han Liu, Junwei Lu, and Ziwei Zhu. Distributed testing and estimation under sparse high dimensional models. *The Annals of Statistics*, 46(3):1352–1382, 2018.

[3] A. Belloni and V. Chernozhukov. l_1-penalized quantile regression for high dimensional sparse models. *The Annals of Statistics*, 39:82–130, 2011.

[4] V. Bentkus. A lyapunov-type bound in $\mathbb{R}^d$. *Theory of Probability & Its Applications*, 49(2):311–323, 2005.

[5] Thierry Bertin-Mahieux, Daniel PW Ellis, Brian Whitman, and Paul Lamere. The million song dataset. *Proceedings of the 12th International Society for Music Information Retrieval Conference, ISMIR 2011*, 2:1–6.

[6] Xi Chen, Weidong Liu, and Yichen Zhang. Quantile regression under memory constraint. *The Annals of Statistics*, 47(6):3244–3273, 2019.

[7] Xueying Chen and Minge Xie. A split-and-conquer approach for analysis of extraordinarily large data. *Statistica Sinica*, 24(4):1655–1684, 2014.

[8] Gabriela Ciuperca. Model selection in high-dimensional quantile regression with seamless l_0 penalty. *Statistics and Probability Letters*, 107:313–323, 2015.

[9] Jianqing Fan, Yingying Fan, and Emre Barut. Adaptive robust variable selection. *Annals of statistics*, 42(1):324, 2014.

[10] Jianqing Fan, Fang Han, and Han Liu. Challenges of big data analysis. *National science review*, 1(2):293–314, 2014.

[11] Dan Garber, Ohad Shamir, and Nathan Srebro. Communication-efficient algorithms for distributed stochastic principal component analysis. *Proceedings of the 34th International Conference on Machine Learning*, 70:1203–1212, 2017.

[12] Yuwen Gu, Jun Fan, Lingchen Kong, Shiqian Ma, and Hui Zou. Admm for high-dimensional sparse penalized quantile regression. *Technometrics*, 60(3):319–331, 2018.

[13] C. Gutenbrunner, J. Jurecková, R. Koenker, and S. Portnoy. Tests of linear hypotheses based on regression rank scores. *Journal of Nonparametric Statistics*, 2(4):307–331, 1992.

[14] Xuming He, Lan Wang, and Hyokyoung Grace Hong. Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *The Annals of Statistics*, 41(1):342–369, 2013.

[15] K. Knight. Limiting distributions for l_1 regression estimators under general conditions. *The Annals of Statistics*, 26:755–770, 1998.

[16] Roger Koenker and Gilbert Bassett. Tests of linear hypotheses and l_1 estimation. *Econometrica*, 50(6):1577–1584, 1982.

[17] Roger Koenker and José A. F. Machado. Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94(448):1296–1310, 1999.

[18] Nan Lin and Ruibin Xi. Aggregated estimating equation estimation. *Statistics and Its Interface*, 4(1):73–83, 2011.

[19] X. Lu and L. Su. Jackknife model averaging for quantile regressions. *Journal of Econometrics*, 188:40–58, 2015.

[20] Chun Wang, Ming-Hui Chen, Elizabeth Schifano, Jing Wu, and Jun Yan. Statistical methods and computing for big data. *Statistics and Its Interface*, 9(4):399–411, 2016.

[21] Xiangyu Wang and Chenlei Leng. High-dimensional ordinary least-squares projection for screening variables. *Journal of the Royal Statistical Society*, 78(3):589–611, 2016.

[22] Ruibin Xi, Lin Nan, and Yixin Chen. Compression and aggregation for logistic regression analysis in data cubes. *IEEE Transactions on Knowledge and Data Engineering*, 21(4):479–492, 2009.

[23] Ganggang Xu, Zuofeng Shang, and Guang Cheng. Distributed generalized cross-validation for divide-and-conquer kernel ridge regression and its asymptotic optimality. *Journal of Computational and Graphical Statistics*, 28(4):891–908, 2019.

[24] Congrui Yi and Jian Huang. Semismooth newton coordinate descent algorithm for elastic-net penalized huber loss regression and quantile regression. *Journal of Computational and Graphical Statistics*, 26(3):547–557, 2017.

[25] Yuchen Zhang, John Duchi, and Martin Wainwright. Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates. *The Journal of Machine Learning Research*, 16(1):3299–3340, 2015.

[26] Yuchen Zhang, John C Duchi, and Martin J Wainwright. Communication-efficient algorithms for statistical optimization. *The Journal of Machine Learning Research*, 14(1):3321–3363, 2013.

Simulation Results

Table 3: Results in Section 4.1 for Cauchy error.

τ		0.3						0.5					
n=500	**Method**	**Loss_beta**	**Time**	**Model_Size**	**TP**	**FP**	**F1-Score**	**Loss_beta**	**Time**	**Model_Size**	**TP**	**FP**	**F1-Score**
L=10	**DLS**	1.49E-19	82.66	1001.00	6.00	995.00	0.01	7.51E-20	203.39	1001.00	6.00	995.00	0.01
	DHXS	2.24E-32	91.89	236.60	6.00	230.60	0.05	1E-32	221.75	233.96	6.00	227.96	0.05
	DHXV	2.01E-32	91.88	6.00	6.00	0.00	1.00	7.91E-33	221.73	6.00	6.00	0.00	1.00
	DlHXS	2.24E-32	17.96	236.60	6.00	230.60	0.05	1E-32	39.27	233.96	6.00	227.96	0.05
	DlHXV	2.01E-32	17.95	6.00	6.00	0.00	1.00	7.91E-33	39.25	6.00	6.00	0.00	1.00
	DLHS	2.11E-32	14.22	228.74	6.00	222.74	0.05	1.34E-32	31.16	239.53	6.00	233.53	0.05
	DLHV	1.89E-32	14.20	6.00	6.00	0.00	1.00	1.12E-32	31.13	6.00	6.00	0.00	1.00
	WL	2.31E-21	159.54	1001.00	6.00	995.00	0.01	2.52E-22	241.69	1001.00	6.00	995.00	0.01
	WQ	2.5E-31	237.63	1001.00	6.00	995.00	0.01	2.25E-31	547.66	1001.00	6.00	995.00	0.01
L=20	**DLS**	1.24E-19	176.84	1001.00	6.00	995.00	0.01	5.03E-20	297.46	1001.00	6.00	995.00	0.01
	DHXS	1.35E-06	195.60	410.76	6.00	404.76	0.03	1.33E-06	332.31	415.43	6.00	409.43	0.03
	DHXV	1.15E-06	195.59	6.00	6.00	0.00	1.00	1.07E-06	332.29	6.00	6.00	0.00	1.00
	DlHXS	1.35E-06	28.31	410.76	6.00	404.76	0.03	1.33E-06	52.08	415.43	6.00	409.43	0.03
	DlHXV	1.15E-06	28.30	6.00	6.00	0.00	1.00	1.07E-06	52.06	6.00	6.00	0.00	1.00
	DLHS	1.5E-32	28.61	403.94	6.00	397.94	0.03	5.07E-33	57.38	416.16	6.00	410.16	0.03
	DLHV	1.39E-32	28.59	6.00	6.00	0.00	1.00	4.01E-33	57.35	6.00	6.00	0.00	1.00
	WL	1.18E-22	633.73	1001.00	6.00	995.00	0.01	1.58E-23	934.23	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	690.15	1001.00	6.00	995.00	0.01	1.7E-31	1488.28	1001.00	6.00	995.00	0.01
L=40	**DLS**	1.21E-19	750.18	1001.00	6.00	995.00	0.01	5.48E-20	891.12	1001.00	6.00	995.00	0.01
	DHXS	1.36E-32	837.78	638.28	6.00	632.28	0.02	1.52E-33	1002.09	651.33	6.00	645.33	0.02
	DHXV	1.3E-32	837.75	6.00	6.00	0.00	1.00	9.94E-34	1002.06	6.00	6.00	0.00	1.00
	DlHXS	1.36E-32	114.38	638.28	6.00	632.28	0.02	1.52E-33	189.55	651.33	6.00	645.33	0.02
	DlHXV	1.3E-32	114.35	6.00	6.00	0.00	1.00	9.94E-34	189.49	6.00	6.00	0.00	1.00
	DLHS	1.42E-32	150.35	637.92	6.00	631.92	0.02	2.29E-33	358.44	654.57	6.00	648.57	0.02
	DLHV	1.37E-32	150.31	6.00	6.00	0.00	1.00	1.76E-33	358.37	6.00	6.00	0.00	1.00
	WL	4.13E-23	4672.31	1001.00	6.00	995.00	0.01	1.34E-26	4982.25	1001.00	6.00	995.00	0.01
	WQ	1.56E-31	3236.34	1001.00	6.00	995.00	0.01	1.45E-31	3408.56	1001.00	6.00	995.00	0.01

Table 4: Results in Section 4.1 for Exponential error.

τ		0.3						0.5					
n=500	Method	Loss_beta	Time	Model_Size	TP	FP	F1-Score	Loss_beta	Time	Model_Size	TP	FP	F1-Score
L=10	**DLS**	1.5E-19	89.47	1001.00	6.00	995.00	0.01	7.63E-20	226.76	1001.00	6.00	995.00	0.01
	DHXS	2.24E-32	99.15	236.60	6.00	230.60	0.05	1E-32	254.69	233.96	6.00	227.96	0.05
	DHXV	2.01E-32	99.13	6.00	6.00	0.00	1.00	7.91E-33	254.66	6.00	6.00	0.00	1.00
	DlHXS	2.24E-32	18.99	236.60	6.00	230.60	0.05	1E-32	50.08	233.96	6.00	227.96	0.05
	DlHXV	2.01E-32	18.98	6.00	6.00	0.00	1.00	7.91E-33	50.04	6.00	6.00	0.00	1.00
	DLHS	2.11E-32	14.46	228.74	6.00	222.74	0.05	1.34E-32	41.82	239.53	6.00	233.53	0.05
	DLHV	1.89E-32	14.45	6.00	6.00	0.00	1.00	1.12E-32	41.79	6.00	6.00	0.00	1.00
	WL	2.18E-21	128.30	1001.00	6.00	995.00	0.01	1.75E-22	266.32	1001.00	6.00	995.00	0.01
	WQ	2.5E-31	252.20	1001.00	6.00	995.00	0.01	2.25E-31	645.56	1001.00	6.00	995.00	0.01
L=20	**DLS**	1.32E-19	378.19	1001.00	6.00	995.00	0.01	5.47E-20	410.94	1001.00	6.00	995.00	0.01
	DHXS	1.35E-06	418.82	410.76	6.00	404.76	0.03	1.33E-06	455.32	415.43	6.00	409.43	0.03
	DHXV	1.15E-06	418.79	6.00	6.00	0.00	1.00	1.07E-06	455.30	6.00	6.00	0.00	1.00
	DlHXS	1.35E-06	59.29	410.76	6.00	404.76	0.03	1.33E-06	63.15	415.43	6.00	409.43	0.03
	DlHXV	1.15E-06	59.26	6.00	6.00	0.00	1.00	1.07E-06	63.12	6.00	6.00	0.00	1.00
	DLHS	1.5E-32	57.03	403.94	6.00	397.94	0.03	5.07E-33	60.93	416.16	6.00	410.16	0.03
	DLHV	1.39E-32	57.00	6.00	6.00	0.00	1.00	4.01E-33	60.90	6.00	6.00	0.00	1.00
	WL	8.2E-23	861.17	1001.00	6.00	995.00	0.01	2.11E-23	854.34	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	1444.80	1001.00	6.00	995.00	0.01	1.7E-31	1302.31	1001.00	6.00	995.00	0.01
L=40	**DLS**	1.15E-19	864.91	1001.00	6.00	995.00	0.01	5.33E-20	972.96	1001.00	6.00	995.00	0.01
	DHXS	1.36E-32	967.69	638.28	6.00	632.28	0.02	1.52E-33	1124.17	651.33	6.00	645.33	0.02
	DHXV	1.3E-32	967.65	6.00	6.00	0.00	1.00	9.94E-34	1124.12	6.00	6.00	0.00	1.00
	DlHXS	1.36E-32	127.00	638.28	6.00	632.28	0.02	1.52E-33	184.06	651.33	6.00	645.33	0.02
	DlHXV	1.3E-32	126.97	6.00	6.00	0.00	1.00	9.94E-34	184.01	6.00	6.00	0.00	1.00
	DLHS	1.42E-32	166.43	637.92	6.00	631.92	0.02	2.29E-33	254.01	654.57	6.00	648.57	0.02
	DLHV	1.37E-32	166.40	6.00	6.00	0.00	1.00	1.76E-33	253.96	6.00	6.00	0.00	1.00
	WL	1.49E-23	4139.88	1001.00	6.00	995.00	0.01	2.12E-24	4312.62	1001.00	6.00	995.00	0.01
	WQ	1.56E-31	3508.34	1001.00	6.00	995.00	0.01	1.45E-31	3610.71	1001.00	6.00	995.00	0.01

Table 5: Results in Section 4.1 for Normal error.

τ		0.3						0.5					
n=500	**Method**	**Loss_beta**	**Time**	**Model_Size**	**TP**	**FP**	**F1-Score**	**Loss_beta**	**Time**	**Model_Size**	**TP**	**FP**	**F1-Score**
L=10	**DLS**	1.53E-19	93.50	1001.00	6.00	995.00	0.01	7.76E-20	218.95	1001.00	6.00	995.00	0.01
	DHXS	2.24E-32	102.77	236.60	6.00	230.60	0.05	1E-32	243.12	233.96	6.00	227.96	0.05
	DHXV	2.01E-32	102.76	6.00	6.00	0.00	1.00	7.91E-33	243.09	6.00	6.00	0.00	1.00
	DlHXS	2.24E-32	18.81	236.60	6.00	230.60	0.05	1E-32	45.10	233.96	6.00	227.96	0.05
	DlHXV	2.01E-32	18.79	6.00	6.00	0.00	1.00	7.91E-33	45.08	6.00	6.00	0.00	1.00
	DLHS	2.11E-32	13.52	228.74	6.00	222.74	0.05	1.34E-32	36.48	239.53	6.00	233.53	0.05
	DLHV	1.89E-32	13.51	6.00	6.00	0.00	1.00	1.12E-32	36.45	6.00	6.00	0.00	1.00
	WL	3.1E-21	98.99	1001.00	6.00	995.00	0.01	3.31E-22	251.91	1001.00	6.00	995.00	0.01
	WQ	2.5E-31	267.84	1001.00	6.00	995.00	0.01	2.25E-31	653.05	1001.00	6.00	995.00	0.01
L=20	**DLS**	1.31E-19	407.10	1001.00	6.00	995.00	0.01	6.12E-20	388.25	1001.00	6.00	995.00	0.01
	DHXS	1.35E-06	447.11	410.76	6.00	404.76	0.03	1.33E-06	430.81	415.43	6.00	409.43	0.03
	DHXV	1.15E-06	447.08	6.00	6.00	0.00	1.00	1.07E-06	430.79	6.00	6.00	0.00	1.00
	DlHXS	1.35E-06	58.83	410.76	6.00	404.76	0.03	1.33E-06	61.71	415.43	6.00	409.43	0.03
	DlHXV	1.15E-06	58.81	6.00	6.00	0.00	1.00	1.07E-06	61.68	6.00	6.00	0.00	1.00
	DLHS	1.5E-32	53.59	403.94	6.00	397.94	0.03	5.07E-33	61.14	416.16	6.00	410.16	0.03
	DLHV	1.39E-32	53.57	6.00	6.00	0.00	1.00	4.01E-33	61.12	6.00	6.00	0.00	1.00
	WL	2.6E-22	837.90	1001.00	6.00	995.00	0.01	3.39E-23	862.21	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	1325.19	1001.00	6.00	995.00	0.01	1.7E-31	1406.57	1001.00	6.00	995.00	0.01
L=40	**DLS**	1.14E-19	868.76	1001.00	6.00	995.00	0.01	5.67E-20	940.05	1001.00	6.00	995.00	0.01
	DHXS	1.36E-32	963.38	638.28	6.00	632.28	0.02	1.52E-33	1059.10	651.33	6.00	645.33	0.02
	DHXV	1.3E-32	963.35	6.00	6.00	0.00	1.00	9.94E-34	1059.06	6.00	6.00	0.00	1.00
	DlHXS	1.36E-32	115.04	638.28	6.00	632.28	0.02	1.52E-33	148.71	651.33	6.00	645.33	0.02
	DlHXV	1.3E-32	115.01	6.00	6.00	0.00	1.00	9.94E-34	148.67	6.00	6.00	0.00	1.00
	DLHS	1.42E-32	138.13	637.92	6.00	631.92	0.02	2.29E-33	198.41	654.57	6.00	648.57	0.02
	DLHV	1.37E-32	138.10	6.00	6.00	0.00	1.00	1.76E-33	198.37	6.00	6.00	0.00	1.00
	WL	1.38E-23	4046.98	1001.00	6.00	995.00	0.01	1.49E-26	4007.96	1001.00	6.00	995.00	0.01
	WQ	1.56E-31	3734.18	1001.00	6.00	995.00	0.01	1.45E-31	3751.92	1001.00	6.00	995.00	0.01

Table 6: Results in Section 4.1 for t distribution error.

τ		0.3						0.5					
n=500	**Method**	**Loss_beta**	**Time**	**Model_Size**	**TP**	**FP**	**F1-Score**	**Loss_beta**	**Time**	**Model_Size**	**TP**	**FP**	**F1-Score**
L=10	**DLS**	1.64E-19	96.76	1001.00	6.00	995.00	0.01	8.23E-20	183.55	1001.00	6.00	995.00	0.01
	DHXS	2.24E-32	105.76	236.60	6.00	230.60	0.05	1E-32	201.55	233.96	6.00	227.96	0.05
	DHXV	2.01E-32	105.75	6.00	6.00	0.00	1.00	7.91E-33	201.53	6.00	6.00	0.00	1.00
	DlHXS	2.24E-32	18.83	236.60	6.00	230.60	0.05	1E-32	36.15	233.96	6.00	227.96	0.05
	DlHXV	2.01E-32	18.81	6.00	6.00	0.00	1.00	7.91E-33	36.13	6.00	6.00	0.00	1.00
	DLHS	2.11E-32	13.04	228.74	6.00	222.74	0.05	1.34E-32	24.89	239.53	6.00	233.53	0.05
	DLHV	1.89E-32	13.03	6.00	6.00	0.00	1.00	1.12E-32	24.86	6.00	6.00	0.00	1.00
	WL	2.45E-21	84.52	1001.00	6.00	995.00	0.01	2.21E-22	200.61	1001.00	6.00	995.00	0.01
	WQ	2.5E-31	294.72	1001.00	6.00	995.00	0.01	2.25E-31	576.28	1001.00	6.00	995.00	0.01
L=20	**DLS**	1.36E-19	380.74	1001.00	6.00	995.00	0.01	6.2E-20	345.91	1001.00	6.00	995.00	0.01
	DHXS	1.35E-06	419.11	410.76	6.00	404.76	0.03	1.33E-06	379.81	415.43	6.00	409.43	0.03
	DHXV	1.15E-06	419.08	6.00	6.00	0.00	1.00	1.07E-06	379.79	6.00	6.00	0.00	1.00
	DlHXS	1.35E-06	58.31	410.76	6.00	404.76	0.03	1.33E-06	50.00	415.43	6.00	409.43	0.03
	DlHXV	1.15E-06	58.28	6.00	6.00	0.00	1.00	1.07E-06	49.98	6.00	6.00	0.00	1.00
	DLHS	1.5E-32	54.96	403.94	6.00	397.94	0.03	5.07E-33	44.97	416.16	6.00	410.16	0.03
	DLHV	1.39E-32	54.93	6.00	6.00	0.00	1.00	4.01E-33	44.95	6.00	6.00	0.00	1.00
	WL	8.46E-23	844.05	1001.00	6.00	995.00	0.01	2.09E-23	774.26	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	1260.41	1001.00	6.00	995.00	0.01	1.7E-31	1172.64	1001.00	6.00	995.00	0.01
L=40	**DLS**	1.3E-19	804.42	1001.00	6.00	995.00	0.01	5.38E-20	886.91	1001.00	6.00	995.00	0.01
	DHXS	1.36E-32	887.28	638.28	6.00	632.28	0.02	1.52E-33	999.56	651.33	6.00	645.33	0.02
	DHXV	1.3E-32	887.25	6.00	6.00	0.00	1.00	9.94E-34	999.52	6.00	6.00	0.00	1.00
	DlHXS	1.36E-32	106.02	638.28	6.00	632.28	0.02	1.52E-33	130.47	651.33	6.00	645.33	0.02
	DlHXV	1.3E-32	105.99	6.00	6.00	0.00	1.00	9.94E-34	130.43	6.00	6.00	0.00	1.00
	DLHS	1.42E-32	122.27	637.92	6.00	631.92	0.02	2.29E-33	150.48	654.57	6.00	648.57	0.02
	DLHV	1.37E-32	122.24	6.00	6.00	0.00	1.00	1.76E-33	150.44	6.00	6.00	0.00	1.00
	WL	7.19E-23	3775.02	1001.00	6.00	995.00	0.01	1.34E-26	3779.44	1001.00	6.00	995.00	0.01
	WQ	1.56E-31	3559.48	1001.00	6.00	995.00	0.01	1.45E-31	3410.90	1001.00	6.00	995.00	0.01

Table 7: Results in Section 4.1 for χ^2 distribution error.

τ		0.3						0.5					
n=500	Method	Loss_beta	Time	Model_Size	TP	FP	F1-Score	Loss_beta	Time	Model_Size	TP	FP	F1-Score
L=10	**DLS**	1.61E-19	144.47	1001.00	6.00	995.00	0.01	7.62E-20	186.01	1001.00	6.00	995.00	0.01
	DHXS	2.24E-32	159.32	236.60	6.00	230.60	0.05	1E-32	204.06	233.96	6.00	227.96	0.05
	DHXV	2.01E-32	159.30	6.00	6.00	0.00	1.00	7.91E-33	204.04	6.00	6.00	0.00	1.00
	DlHXS	2.24E-32	29.38	236.60	6.00	230.60	0.05	1E-32	36.56	233.96	6.00	227.96	0.05
	DlHXV	2.01E-32	29.36	6.00	6.00	0.00	1.00	7.91E-33	36.54	6.00	6.00	0.00	1.00
	DLHS	2.11E-32	21.12	228.74	6.00	222.74	0.05	1.34E-32	25.09	239.53	6.00	233.53	0.05
	DLHV	1.89E-32	21.10	6.00	6.00	0.00	1.00	1.12E-32	25.06	6.00	6.00	0.00	1.00
	WL	2.52E-21	151.00	1001.00	6.00	995.00	0.01	1.85E-22	209.58	1001.00	6.00	995.00	0.01
	WQ	2.5E-31	388.71	1001.00	6.00	995.00	0.01	2.25E-31	495.99	1001.00	6.00	995.00	0.01
L=20	**DLS**	1.29E-19	325.87	1001.00	6.00	995.00	0.01	5.64E-20	298.77	1001.00	6.00	995.00	0.01
	DHXS	1.35E-06	356.75	410.76	6.00	404.76	0.03	1.33E-06	325.00	415.43	6.00	409.43	0.03
	DHXV	1.15E-06	356.73	6.00	6.00	0.00	1.00	1.07E-06	324.99	6.00	6.00	0.00	1.00
	DlHXS	1.35E-06	45.70	410.76	6.00	404.76	0.03	1.33E-06	41.37	415.43	6.00	409.43	0.03
	DlHXV	1.15E-06	45.68	6.00	6.00	0.00	1.00	1.07E-06	41.35	6.00	6.00	0.00	1.00
	DLHS	1.5E-32	42.46	403.94	6.00	397.94	0.03	5.07E-33	36.01	416.16	6.00	410.16	0.03
	DLHV	1.39E-32	42.44	6.00	6.00	0.00	1.00	4.01E-33	35.99	6.00	6.00	0.00	1.00
	WL	1.2E-22	729.48	1001.00	6.00	995.00	0.01	3.23E-23	695.59	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	1008.84	1001.00	6.00	995.00	0.01	1.7E-31	961.92	1001.00	6.00	995.00	0.01
L=40	**DLS**	1.19E-19	673.66	1001.00	6.00	995.00	0.01	5.19E-20	619.91	1001.00	6.00	995.00	0.01
	DHXS	1.36E-32	732.06	638.28	6.00	632.28	0.02	1.52E-33	677.91	651.33	6.00	645.33	0.02
	DHXV	1.3E-32	732.04	6.00	6.00	0.00	1.00	9.94E-34	677.89	6.00	6.00	0.00	1.00
	DlHXS	1.36E-32	73.49	638.28	6.00	632.28	0.02	1.52E-33	74.26	651.33	6.00	645.33	0.02
	DlHXV	1.3E-32	73.47	6.00	6.00	0.00	1.00	9.94E-34	74.24	6.00	6.00	0.00	1.00
	DLHS	1.42E-32	80.82	637.92	6.00	631.92	0.02	2.29E-33	81.33	654.57	6.00	648.57	0.02
	DLHV	1.37E-32	80.80	6.00	6.00	0.00	1.00	1.76E-33	81.31	6.00	6.00	0.00	1.00
	WL	1.51E-23	2991.71	1001.00	6.00	995.00	0.01	1.53E-26	2676.54	1001.00	6.00	995.00	0.01
	WQ	1.56E-31	2456.80	1001.00	6.00	995.00	0.01	1.45E-31	2128.05	1001.00	6.00	995.00	0.01

Table 8: Results in Section 4.2 for Cauchy error.

τ		0.3						0.5					
	Method	Loss_beta	Time	Model_Size	TP	FP	F1-Score	Loss_beta	Time	Model_Size	TP	FP	F1-Score
$n = 250$ $L = 40$	**DLS**	6.75E-19	243.41	1001.00	6.00	995.00	0.01	2.71E-19	178.88	1001.00	6.00	995.00	0.01
	DHXS	0.001218	265.55	375.81	6.00	369.81	0.03	0.001162	191.74	376.75	6.00	370.75	0.03
	DHXV	0.001169	265.53	6.00	6.00	0.00	1.00	0.001115	191.73	6.00	6.00	0.00	1.00
	DlHXS	0.001218	30.07	375.81	6.00	369.81	0.03	0.001162	17.56	376.75	6.00	370.75	0.03
	DlHXV	0.001169	30.04	6.00	6.00	0.00	1.00	0.001115	17.54	6.00	6.00	0.00	1.00
	DLHS	1.39E-32	81.75	186.78	6.00	180.78	0.07	2.03E-33	35.41	188.79	6.00	182.79	0.07
	DLHV	1.37E-32	81.70	6.00	6.00	0.00	1.00	1.79E-33	35.39	6.00	6.00	0.00	1.00
	WL	1.26E-22	1197.85	1001.00	6.00	995.00	0.01	1.75E-23	1171.24	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	1453.41	1001.00	6.00	995.00	0.01	1.69E-31	1780.73	1001.00	6.00	995.00	0.01
$n = 500$ $L = 20$	**DLS**	1.24E-19	176.84	1001.00	6.00	995.00	0.01	5.03E-20	297.46	1001.00	6.00	995.00	0.01
	DHXS	1.35E-06	195.60	410.76	6.00	404.76	0.03	1.33E-06	332.31	415.43	6.00	409.43	0.03
	DHXV	1.15E-06	195.59	6.00	6.00	0.00	1.00	1.07E-06	332.29	6.00	6.00	0.00	1.00
	DlHXS	1.35E-06	28.31	410.76	6.00	404.76	0.03	1.33E-06	52.08	415.43	6.00	409.43	0.03
	DlHXV	1.15E-06	28.30	6.00	6.00	0.00	1.00	1.07E-06	52.06	6.00	6.00	0.00	1.00
	DLHS	1.5E-32	28.61	403.94	6.00	397.94	0.03	5.07E-33	57.38	416.16	6.00	410.16	0.03
	DLHV	1.39E-32	28.59	6.00	6.00	0.00	1.00	4.01E-33	57.35	6.00	6.00	0.00	1.00
	WL	1.18E-22	633.73	1001.00	6.00	995.00	0.01	1.58E-23	934.23	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	690.15	1001.00	6.00	995.00	0.01	1.7E-31	1488.28	1001.00	6.00	995.00	0.01
$n = 1000$ $L = 10$	**DLS**	9.36E-20	482.34	1001.00	6.00	995.00	0.01	1.1E-20	872.81	1001.00	6.00	995.00	0.01
	DHXS	2.35E-32	626.07	552.87	6.00	546.87	0.02	1.36E-32	1067.59	553.66	6.00	547.66	0.02
	DHXV	2.02E-32	626.05	6.02	6.00	0.02	1.00	1.07E-32	1067.57	6.07	6.00	0.07	1.00
	DlHXS	2.39E-32	255.59	134.55	6.00	128.55	0.09	1.26E-32	224.28	134.17	6.00	128.17	0.09
	DlHXV	2.31E-32	255.57	74.98	6.00	68.98	0.16	1.19E-32	224.26	73.84	6.00	67.84	0.16
	DLHS	2.71E-32	128.26	557.00	6.00	551.00	0.02	1.22E-32	155.33	554.03	6.00	548.03	0.02
	DLHV	2.37E-32	128.25	6.03	6.00	0.03	1.00	9.33E-33	155.31	6.05	6.00	0.05	1.00
	WL	5.58E-23	1064.31	1001.00	6.00	995.00	0.01	2.05E-23	946.88	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	1372.00	1001.00	6.00	995.00	0.01	1.7E-31	1586.07	1001.00	6.00	995.00	0.01

Table 9: Results in Section 4.2 for Exponential error.

τ		0.3						0.5					
	Method	Loss_beta	Time	Model_Size	TP	FP	F1-Score	Loss_beta	Time	Model_Size	TP	FP	F1-Score
	DLS	6.99E-19	276.26	1001.00	6.00	995.00	0.01	2.72E-19	271.08	1001.00	6.00	995.00	0.01
	DHXS	0.001218	300.42	375.81	6.00	369.81	0.03	0.001162	290.60	376.75	6.00	370.75	0.03
	DHXV	0.001169	300.38	6.00	6.00	0.00	1.00	0.001115	290.58	6.00	6.00	0.00	1.00
	DlHXS	0.001218	30.82	375.81	6.00	369.81	0.03	0.001162	26.23	376.75	6.00	370.75	0.03
$n = 250$	**DlHXV**	0.001169	30.79	6.00	6.00	0.00	1.00	0.001115	26.21	6.00	6.00	0.00	1.00
$L = 40$	**DLHS**	1.39E-32	59.70	186.78	6.00	180.78	0.07	2.03E-33	51.46	188.79	6.00	182.79	0.07
	DLHV	1.37E-32	59.67	6.00	6.00	0.00	1.00	1.79E-33	51.42	6.00	6.00	0.00	1.00
	WL	2.2E-22	924.18	1001.00	6.00	995.00	0.01	3.06E-23	923.17	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	1523.27	1001.00	6.00	995.00	0.01	1.69E-31	1721.54	1001.00	6.00	995.00	0.01
	DLS	1.32E-19	378.19	1001.00	6.00	995.00	0.01	5.47E-20	410.94	1001.00	6.00	995.00	0.01
	DHXS	1.35E-06	418.82	410.76	6.00	404.76	0.03	1.33E-06	455.32	415.43	6.00	409.43	0.03
	DHXV	1.15E-06	418.79	6.00	6.00	0.00	1.00	1.07E-06	455.30	6.00	6.00	0.00	1.00
	DlHXS	1.35E-06	59.29	410.76	6.00	404.76	0.03	1.33E-06	63.15	415.43	6.00	409.43	0.03
$n = 500$	**DlHXV**	1.15E-06	59.26	6.00	6.00	0.00	1.00	1.07E-06	63.12	6.00	6.00	0.00	1.00
$L = 20$	**DLHS**	1.5E-32	57.03	403.94	6.00	397.94	0.03	5.07E-33	60.93	416.16	6.00	410.16	0.03
	DLHV	1.39E-32	57.00	6.00	6.00	0.00	1.00	4.01E-33	60.90	6.00	6.00	0.00	1.00
	WL	8.2E-23	861.17	1001.00	6.00	995.00	0.01	2.11E-23	854.34	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	1444.80	1001.00	6.00	995.00	0.01	1.7E-31	1302.31	1001.00	6.00	995.00	0.01
	DLS	9.63E-20	683.28	1001.00	6.00	995.00	0.01	1.09E-20	672.19	1001.00	6.00	995.00	0.01
	DHXS	2.32E-32	821.50	646.39	6.00	640.39	0.02	1.27E-32	854.91	665.69	6.00	659.69	0.02
	DHXV	2E-32	821.48	6.05	6.00	0.05	1.00	9.81E-33	854.89	6.11	6.00	0.11	0.99
	DlHXS	2.41E-32	207.48	156.08	6.00	150.08	0.08	1.28E-32	257.72	163.13	6.00	157.13	0.07
$n = 1000$	**DlHXV**	2.38E-32	207.46	92.50	6.00	86.50	0.13	1.26E-32	257.70	92.48	6.00	86.48	0.13
$L = 10$	**DLHS**	2.71E-32	166.79	557.00	6.00	551.00	0.02	1.22E-32	220.98	554.03	6.00	548.03	0.02
	DLHV	2.37E-32	166.77	6.03	6.00	0.03	1.00	9.33E-33	220.96	6.05	6.00	0.05	1.00
	WL	2.22E-22	902.49	1001.00	6.00	995.00	0.01	1.73E-23	920.30	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	1399.63	1001.00	6.00	995.00	0.01	1.7E-31	1493.16	1001.00	6.00	995.00	0.01

Table 10: Results in Section 4.2 for Normal error.

τ		0.3						0.5					
	Method	Loss_beta	Time	Model_Size	TP	FP	F1-Score	Loss_beta	Time	Model_Size	TP	FP	F1-Score
$n = 250$ $L = 40$	**DLS**	6.79E-19	281.39	1001.00	6.00	995.00	0.01	2.68E-19	368.65	1001.00	6.00	995.00	0.01
	DHXS	0.001218	305.38	375.81	6.00	369.81	0.03	0.001162	399.05	376.75	6.00	370.75	0.03
	DHXV	0.001169	305.36	6.00	6.00	0.00	1.00	0.001115	399.01	6.00	6.00	0.00	1.00
	DlHXS	0.001218	30.42	375.81	6.00	369.81	0.03	0.001162	38.74	376.75	6.00	370.75	0.03
	DlHXV	0.001169	30.40	6.00	6.00	0.00	1.00	0.001115	38.71	6.00	6.00	0.00	1.00
	DLHS	1.39E-32	57.83	186.78	6.00	180.78	0.07	2.03E-33	76.72	188.79	6.00	182.79	0.07
	DLHV	1.37E-32	57.80	6.00	6.00	0.00	1.00	1.79E-33	76.69	6.00	6.00	0.00	1.00
	WL	1.57E-22	915.33	1001.00	6.00	995.00	0.01	2.14E-23	1012.80	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	1615.40	1001.00	6.00	995.00	0.01	1.69E-31	1600.58	1001.00	6.00	995.00	0.01
$n = 500$ $L = 20$	**DLS**	1.31E-19	407.10	1001.00	6.00	995.00	0.01	6.12E-20	388.25	1001.00	6.00	995.00	0.01
	DHXS	1.35E-06	447.11	410.76	6.00	404.76	0.03	1.33E-06	430.81	415.43	6.00	409.43	0.03
	DHXV	1.15E-06	447.08	6.00	6.00	0.00	1.00	1.07E-06	430.79	6.00	6.00	0.00	1.00
	DlHXS	1.35E-06	58.83	410.76	6.00	404.76	0.03	1.33E-06	61.71	415.43	6.00	409.43	0.03
	DlHXV	1.15E-06	58.81	6.00	6.00	0.00	1.00	1.07E-06	61.68	6.00	6.00	0.00	1.00
	DLHS	1.5E-32	53.59	403.94	6.00	397.94	0.03	5.07E-33	61.14	416.16	6.00	410.16	0.03
	DLHV	1.39E-32	53.57	6.00	6.00	0.00	1.00	4.01E-33	61.12	6.00	6.00	0.00	1.00
	WL	2.6E-22	837.90	1001.00	6.00	995.00	0.01	3.39E-23	862.21	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	1325.19	1001.00	6.00	995.00	0.01	1.7E-31	1406.57	1001.00	6.00	995.00	0.01
$n = 1000$ $L = 10$	**DLS**	8.8E-20	688.26	1001.00	6.00	995.00	0.01	1.26E-20	683.24	1001.00	6.00	995.00	0.01
	DHXS	2.24E-32	862.07	688.34	6.00	682.34	0.02	1.25E-32	844.89	720.14	6.00	714.14	0.02
	DHXV	1.91E-32	862.05	6.06	6.00	0.06	1.00	9.56E-33	844.87	6.15	6.00	0.15	0.99
	DlHXS	2.43E-32	229.10	166.26	6.00	160.26	0.07	1.13E-32	214.25	177.14	6.00	171.14	0.07
	DlHXV	2.43E-32	229.08	99.51	6.00	93.51	0.12	1.14E-32	214.23	102.44	6.00	96.44	0.12
	DLHS	2.71E-32	160.06	557.00	6.00	551.00	0.02	1.22E-32	152.66	554.03	6.00	548.03	0.02
	DLHV	2.37E-32	160.04	6.03	6.00	0.03	1.00	9.33E-33	152.64	6.05	6.00	0.05	1.00
	WL	1.32E-22	880.06	1001.00	6.00	995.00	0.01	2.47E-23	823.53	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	1244.21	1001.00	6.00	995.00	0.01	1.7E-31	1331.72	1001.00	6.00	995.00	0.01

Table 11: Results in Section 4.2 for t distribution error.

τ		0.3						0.5					
	Method	Loss_beta	Time	Model_Size	TP	FP	F1-Score	Loss_beta	Time	Model_Size	TP	FP	F1-Score
$n = 250$ $L = 40$	**DLS**	6.65E-19	226.02	1001.00	6.00	995.00	0.01	2.7E-19	303.77	1001.00	6.00	995.00	0.01
	DHXS	0.001218	242.74	375.81	6.00	369.81	0.03	0.001162	328.05	376.75	6.00	370.75	0.03
	DHXV	0.001169	242.72	6.00	6.00	0.00	1.00	0.001115	328.02	6.00	6.00	0.00	1.00
	DlHXS	0.001218	22.33	375.81	6.00	369.81	0.03	0.001162	32.39	376.75	6.00	370.75	0.03
	DlHXV	0.001169	22.31	6.00	6.00	0.00	1.00	0.001115	32.36	6.00	6.00	0.00	1.00
	DLHS	1.39E-32	40.01	186.78	6.00	180.78	0.07	2.03E-33	61.42	188.79	6.00	182.79	0.07
	DLHV	1.37E-32	39.99	6.00	6.00	0.00	1.00	1.79E-33	61.39	6.00	6.00	0.00	1.00
	WL	1.86E-22	776.95	1001.00	6.00	995.00	0.01	9.3E-24	949.73	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	1207.59	1001.00	6.00	995.00	0.01	1.69E-31	1436.91	1001.00	6.00	995.00	0.01
$n = 500$ $L = 20$	**DLS**	1.36E-19	380.74	1001.00	6.00	995.00	0.01	6.2E-20	345.91	1001.00	6.00	995.00	0.01
	DHXS	1.35E-06	419.11	410.76	6.00	404.76	0.03	1.33E-06	379.81	415.43	6.00	409.43	0.03
	DHXV	1.15E-06	419.08	6.00	6.00	0.00	1.00	1.07E-06	379.79	6.00	6.00	0.00	1.00
	DlHXS	1.35E-06	58.31	410.76	6.00	404.76	0.03	1.33E-06	50.00	415.43	6.00	409.43	0.03
	DlHXV	1.15E-06	58.28	6.00	6.00	0.00	1.00	1.07E-06	49.98	6.00	6.00	0.00	1.00
	DLHS	1.5E-32	54.96	403.94	6.00	397.94	0.03	5.07E-33	44.97	416.16	6.00	410.16	0.03
	DLHV	1.39E-32	54.93	6.00	6.00	0.00	1.00	4.01E-33	44.95	6.00	6.00	0.00	1.00
	WL	8.46E-23	844.05	1001.00	6.00	995.00	0.01	2.09E-23	774.26	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	1260.41	1001.00	6.00	995.00	0.01	1.7E-31	1172.64	1001.00	6.00	995.00	0.01
$n = 1000$ $L = 10$	**DLS**	9.08E-20	559.53	1001.00	6.00	995.00	0.01	1.2E-20	557.46	1001.00	6.00	995.00	0.01
	DHXS	2.51E-32	657.68	717.68	6.00	711.68	0.02	1.37E-32	663.22	754.53	6.00	748.53	0.02
	DHXV	2.17E-32	657.67	6.08	6.00	0.08	0.99	1.09E-32	663.20	6.17	6.00	0.17	0.99
	DlHXS	2.35E-32	153.17	172.80	6.00	166.80	0.07	1.18E-32	152.11	190.14	6.00	184.14	0.06
	DlHXV	2.36E-32	153.16	105.23	6.00	99.23	0.11	1.22E-32	152.09	112.20	6.00	106.20	0.11
	DLHS	2.71E-32	106.78	557.00	6.00	551.00	0.02	1.22E-32	101.57	554.03	6.00	548.03	0.02
	DLHV	2.37E-32	106.76	6.03	6.00	0.03	1.00	9.33E-33	101.56	6.05	6.00	0.05	1.00
	WL	5.21E-23	726.22	1001.00	6.00	995.00	0.01	2.26E-23	669.72	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	930.73	1001.00	6.00	995.00	0.01	1.7E-31	950.18	1001.00	6.00	995.00	0.01

Table 12: Results in Section 4.2 for χ^2 distribution error.

		0.3						0.5					
	Method	**Loss_beta**	**Time**	**Model_Size**	**TP**	**FP**	**F1-Score**	**Loss_beta**	**Time**	**Model_Size**	**TP**	**FP**	**F1-Score**
$n = 250$ $L = 40$	**DLS**	6.71E-19	186.46	1001.00	6.00	995.00	0.01	2.86E-19	241.83	1001.00	6.00	995.00	0.01
	DHXS	0.001218	199.85	375.81	6.00	369.81	0.03	0.001162	259.60	376.75	6.00	370.75	0.03
	DHXV	0.001169	199.83	6.00	6.00	0.00	1.00	0.001115	259.58	6.00	6.00	0.00	1.00
	DlHXS	0.001218	18.07	375.81	6.00	369.81	0.03	0.001162	23.48	376.75	6.00	370.75	0.03
	DlHXV	0.001169	18.06	6.00	6.00	0.00	1.00	0.001115	23.46	6.00	6.00	0.00	1.00
	DLHS	1.39E-32	32.50	186.78	6.00	180.78	0.07	2.03E-33	42.80	188.79	6.00	182.79	0.07
	DLHV	1.37E-32	32.48	6.00	6.00	0.00	1.00	1.79E-33	42.78	6.00	6.00	0.00	1.00
	WL	2.13E-22	672.81	1001.00	6.00	995.00	0.01	2.63E-23	843.87	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	876.90	1001.00	6.00	995.00	0.01	1.69E-31	1113.39	1001.00	6.00	995.00	0.01
$n = 500$ $L = 20$	**DLS**	1.29E-19	325.87	1001.00	6.00	995.00	0.01	5.64E-20	298.77	1001.00	6.00	995.00	0.01
	DHXS	1.35E-06	356.75	410.76	6.00	404.76	0.03	1.33E-06	325.00	415.43	6.00	409.43	0.03
	DHXV	1.15E-06	356.73	6.00	6.00	0.00	1.00	1.07E-06	324.99	6.00	6.00	0.00	1.00
	DlHXS	1.35E-06	45.70	410.76	6.00	404.76	0.03	1.33E-06	41.37	415.43	6.00	409.43	0.03
	DlHXV	1.15E-06	45.68	6.00	6.00	0.00	1.00	1.07E-06	41.35	6.00	6.00	0.00	1.00
	DLHS	1.5E-32	42.46	403.94	6.00	397.94	0.03	5.07E-33	36.01	416.16	6.00	410.16	0.03
	DLHV	1.39E-32	42.44	6.00	6.00	0.00	1.00	4.01E-33	35.99	6.00	6.00	0.00	1.00
	WL	1.2E-22	729.48	1001.00	6.00	995.00	0.01	3.23E-23	695.59	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	1008.84	1001.00	6.00	995.00	0.01	1.7E-31	961.92	1001.00	6.00	995.00	0.01
$n = 1000$ $L = 10$	**DLS**	8.64E-20	415.93	1001.00	6.00	995.00	0.01	1.2E-20	375.51	1001.00	6.00	995.00	0.01
	DHXS	2.26E-32	502.34	738.30	6.00	732.30	0.02	1.27E-32	454.90	775.34	6.00	769.34	0.02
	DHXV	1.93E-32	502.33	6.10	6.00	0.10	0.99	9.81E-33	454.89	6.17	6.00	0.17	0.99
	DlHXS	2.3E-32	116.16	179.21	6.00	173.21	0.07	1.11E-32	110.15	196.93	6.00	190.93	0.06
	DlHXV	2.34E-32	116.15	109.77	6.00	103.77	0.11	1.17E-32	110.14	118.39	6.00	112.39	0.10
	DLHS	2.71E-32	80.05	557.00	6.00	551.00	0.02	1.22E-32	74.77	554.03	6.00	548.03	0.02
	DLHV	2.37E-32	80.04	6.03	6.00	0.03	1.00	9.33E-33	74.75	6.05	6.00	0.05	1.00
	WL	2.71E-22	409.36	1001.00	6.00	995.00	0.01	2.17E-23	321.36	1001.00	6.00	995.00	0.01
	WQ	1.59E-31	646.74	1001.00	6.00	995.00	0.01	1.7E-31	632.68	1001.00	6.00	995.00	0.01

Table 13: Results in Section 4.3 for Cauchy error.

τ		0.3						0.5					
n=1000	Method	Loss_beta	Time	Model_Size	TP	FP	F1-Score	Loss_beta	Time	Model_Size	TP	FP	F1-Score
L=5	**DLS**	1.07E-20	236.50	1001.00	20.00	981.00	0.04	3.81E-21	250.73	1001.00	20.00	981.00	0.04
	DHXS	7.04E-29	326.56	227.98	20.00	207.98	0.18	3.5E-29	338.03	246.33	20.00	226.33	0.17
	DHXV	6.66E-29	326.55	21.90	20.00	1.90	0.96	3.2E-29	338.01	22.49	20.00	2.49	0.94
	DlHXS	6.69E-29	137.11	137.78	20.00	117.78	0.27	3.44E-29	137.19	142.99	20.00	122.99	0.26
	DlHXV	6.51E-29	137.10	56.33	20.00	36.33	0.60	3.31E-29	137.17	60.67	20.00	40.67	0.58
	DLHS	6.74E-29	96.57	230.17	20.00	210.17	0.18	3.36E-29	98.91	251.59	20.00	231.59	0.16
	DLHV	6.35E-29	96.54	21.03	20.00	1.03	0.98	3.07E-29	98.89	21.26	20.00	1.26	0.97
	WL	1.18E-20	164.68	1001.00	20.00	981.00	0.04	2.98E-22	220.95	1001.00	20.00	981.00	0.04
	WQ	5.92E-28	517.13	1001.00	20.00	981.00	0.04	5.46E-28	586.85	1001.00	20.00	981.00	0.04
L=10	**DLS**	1E-20	445.15	1001.00	20.00	981.00	0.04	3.65E-21	413.74	1001.00	20.00	981.00	0.04
	DHXS	5.27E-29	684.95	386.72	20.00	366.72	0.10	1.84E-29	636.29	410.89	20.00	390.89	0.10
	DHXV	5.05E-29	684.94	20.76	20.00	0.76	0.98	1.67E-29	636.28	21.05	20.00	1.05	0.97
	DlHXS	5.26E-29	276.73	173.60	20.00	153.60	0.22	1.88E-29	260.00	181.03	20.00	161.03	0.22
	DlHXV	5.14E-29	276.72	44.97	20.00	24.97	0.69	1.79E-29	259.99	49.85	20.00	29.85	0.64
	DLHS	5.17E-29	203.89	394.49	20.00	374.49	0.10	1.79E-29	192.04	396.96	20.00	376.96	0.10
	DLHV	4.96E-29	203.88	20.00	20.00	0.00	1.00	1.63E-29	192.04	20.00	20.00	0.00	1.00
	WL	3.5E-21	447.99	1001.00	20.00	981.00	0.04	4.5E-23	432.51	1001.00	20.00	981.00	0.04
	WQ	4.18E-28	1136.71	1001.00	20.00	981.00	0.04	3.49E-28	1115.95	1001.00	20.00	981.00	0.04
L=20	**DLS**	8.36E-21	1401.62	1001.00	20.00	981.00	0.04	3.33E-21	1445.29	1001.00	20.00	981.00	0.04
	DHXS	3.81E-29	1751.24	632.49	20.00	612.49	0.06	9.29E-30	1809.11	641.56	20.00	621.56	0.06
	DHXV	3.71E-29	1751.22	21.04	20.00	1.04	0.97	8.49E-30	1809.09	21.27	20.00	1.27	0.97
	DlHXS	3.8E-29	419.35	219.60	20.00	199.60	0.18	8.97E-30	440.53	216.25	20.00	196.25	0.18
	DlHXV	3.74E-29	419.33	49.63	20.00	29.63	0.61	8.54E-30	440.51	54.19	20.00	34.19	0.57
	DLHS	3.88E-29	444.14	650.85	20.00	630.85	0.06	9.57E-30	608.62	664.18	20.00	644.18	0.06
	DLHV	3.77E-29	444.12	20.00	20.00	0.00	1.00	8.8E-30	608.56	20.00	20.00	0.00	1.00
	WL	4.6E-23	4091.63	1001.00	20.00	981.00	0.04	9.6E-25	4739.99	1001.00	20.00	981.00	0.04
	WQ	3.91E-28	3550.08	1001.00	20.00	981.00	0.04	3.13E-28	3680.68	1001.00	20.00	981.00	0.04

Table 14: Results in Section 4.3 for Exponential error.

τ		0.3						0.5					
n=1000	Method	Loss_beta	Time	Model_Size	TP	FP	F1-Score	Loss_beta	Time	Model_Size	TP	FP	F1-Score
L=5	**DLS**	1.25E-20	223.04	1001.00	20.00	981.00	0.04	3.72E-21	223.98	1001.00	20.00	981.00	0.04
	DHXS	6.73E-29	306.35	310.90	20.00	290.90	0.13	3.56E-29	314.02	318.75	20.00	298.75	0.12
	DHXV	6.34E-29	306.33	22.85	20.00	2.85	0.94	3.27E-29	314.00	23.50	20.00	3.50	0.92
	DlHXS	6.77E-29	125.77	160.76	20.00	140.76	0.23	3.55E-29	131.12	161.74	20.00	141.74	0.23
	DlHXV	6.67E-29	125.75	74.25	20.00	54.25	0.48	3.46E-29	131.10	73.55	20.00	53.55	0.48
	DLHS	6.74E-29	88.91	230.17	20.00	210.17	0.18	3.36E-29	95.06	251.59	20.00	231.59	0.16
	DLHV	6.35E-29	88.89	21.03	20.00	1.03	0.98	3.07E-29	95.04	21.26	20.00	1.26	0.97
	WL	1.19E-20	129.56	1001.00	20.00	981.00	0.04	2.89E-22	132.04	1001.00	20.00	981.00	0.04
	WQ	5.92E-28	518.99	1001.00	20.00	981.00	0.04	5.46E-28	532.51	1001.00	20.00	981.00	0.04
L=10	**DLS**	9.47E-21	341.08	1001.00	20.00	981.00	0.04	3.6E-21	347.59	1001.00	20.00	981.00	0.04
	DHXS	5.3E-29	556.74	508.53	20.00	488.53	0.08	1.8E-29	562.73	518.79	20.00	498.79	0.08
	DHXV	5.09E-29	556.73	21.01	20.00	1.01	0.98	1.64E-29	562.72	21.34	20.00	1.34	0.97
	DlHXS	5.24E-29	239.46	195.30	20.00	175.30	0.20	1.84E-29	238.91	200.84	20.00	180.84	0.19
	DlHXV	5.14E-29	239.45	58.62	20.00	38.62	0.56	1.78E-29	238.90	60.07	20.00	40.07	0.55
	DLHS	5.17E-29	178.33	394.49	20.00	374.49	0.10	1.79E-29	175.29	396.96	20.00	376.96	0.10
	DLHV	4.96E-29	178.32	20.00	20.00	0.00	1.00	1.63E-29	175.28	20.00	20.00	0.00	1.00
	WL	3.03E-21	376.99	1001.00	20.00	981.00	0.04	4.11E-23	386.24	1001.00	20.00	981.00	0.04
	WQ	4.18E-28	1167.00	1001.00	20.00	981.00	0.04	3.49E-28	1188.55	1001.00	20.00	981.00	0.04
L=20	**DLS**	8.73E-21	1411.73	1001.00	20.00	981.00	0.04	3.04E-21	1506.37	1001.00	20.00	981.00	0.04
	DHXS	3.86E-29	1827.82	765.36	20.00	745.36	0.05	8.92E-30	1958.15	762.84	20.00	742.84	0.05
	DHXV	3.76E-29	1827.79	21.18	20.00	1.18	0.97	8.13E-30	1958.13	21.44	20.00	1.44	0.97
	DlHXS	3.86E-29	473.42	238.18	20.00	218.18	0.17	9.3E-30	585.98	233.65	20.00	213.65	0.17
	DlHXV	3.81E-29	473.40	60.32	20.00	40.32	0.53	8.98E-30	585.95	64.31	20.00	44.31	0.49
	DLHS	3.88E-29	427.65	650.85	20.00	630.85	0.06	9.57E-30	650.48	664.18	20.00	644.18	0.06
	DLHV	3.77E-29	427.62	20.00	20.00	0.00	1.00	8.8E-30	650.44	20.00	20.00	0.00	1.00
	WL	4.99E-23	3684.43	1001.00	20.00	981.00	0.04	9.12E-25	3813.93	1001.00	20.00	981.00	0.04
	WQ	3.91E-28	3714.88	1001.00	20.00	981.00	0.04	3.13E-28	3945.01	1001.00	20.00	981.00	0.04

Table 15: Results in Section 4.3 for Normal error.

τ				0.3						0.5			
n=1000	Method	Loss_beta	Time	Model_Size	TP	FP	F1-Score	Loss_beta	Time	Model_Size	TP	FP	F1-Score
	DLS	8.52E-21	214.00	1001.00	20.00	981.00	0.04	3.93E-21	236.21	1001.00	20.00	981.00	0.04
	DHXS	6.94E-29	296.08	358.07	20.00	338.07	0.11	3.37E-29	320.30	356.34	20.00	336.34	0.11
	DHXV	6.53E-29	296.06	23.65	20.00	3.65	0.92	3.08E-29	320.28	24.31	20.00	4.31	0.91
	DlHXS	6.76E-29	123.06	171.13	20.00	151.13	0.22	3.52E-29	129.51	170.68	20.00	150.68	0.22
	DlHXV	6.73E-29	123.05	81.63	20.00	61.63	0.43	3.45E-29	129.49	82.55	20.00	62.55	0.43
L=5	**DLHS**	6.74E-29	85.43	230.17	20.00	210.17	0.18	3.36E-29	85.74	251.59	20.00	231.59	0.16
	DLHV	6.35E-29	85.42	21.03	20.00	1.03	0.98	3.07E-29	85.73	21.26	20.00	1.26	0.97
	WL	1.28E-20	113.11	1001.00	20.00	981.00	0.04	2.68E-22	121.36	1001.00	20.00	981.00	0.04
	WQ	5.92E-28	503.10	1001.00	20.00	981.00	0.04	5.46E-28	496.46	1001.00	20.00	981.00	0.04
	DLS	9.54E-21	321.09	1001.00	20.00	981.00	0.04	3.57E-21	324.22	1001.00	20.00	981.00	0.04
	DHXS	5.2E-29	524.51	567.41	20.00	547.41	0.07	1.8E-29	526.13	571.43	20.00	551.43	0.07
	DHXV	4.99E-29	524.50	21.09	20.00	1.09	0.97	1.64E-29	526.13	21.50	20.00	1.50	0.96
	DlHXS	5.29E-29	232.81	207.32	20.00	187.32	0.19	1.75E-29	231.41	210.59	20.00	190.59	0.19
	DlHXV	5.22E-29	232.81	66.36	20.00	46.36	0.51	1.7E-29	231.40	67.18	20.00	47.18	0.50
L=10	**DLHS**	5.17E-29	181.60	394.49	20.00	374.49	0.10	1.79E-29	180.44	396.96	20.00	376.96	0.10
	DLHV	4.96E-29	181.59	20.00	20.00	0.00	1.00	1.63E-29	180.44	20.00	20.00	0.00	1.00
	WL	3.93E-21	390.58	1001.00	20.00	981.00	0.04	4.16E-23	389.71	1001.00	20.00	981.00	0.04
	WQ	4.18E-28	1211.53	1001.00	20.00	981.00	0.04	3.49E-28	1228.60	1001.00	20.00	981.00	0.04
	DLS	8.91E-21	1477.95	1001.00	20.00	981.00	0.04	3.08E-21	1565.34	1001.00	20.00	981.00	0.04
	DHXS	3.85E-29	1874.20	821.31	20.00	801.31	0.05	9.61E-30	2045.35	811.84	20.00	791.84	0.05
	DHXV	3.75E-29	1874.17	21.24	20.00	1.24	0.97	8.84E-30	2045.32	21.54	20.00	1.54	0.96
	DlHXS	3.81E-29	434.17	249.51	20.00	229.51	0.16	9.51E-30	431.16	241.16	20.00	221.16	0.17
	DlHXV	3.77E-29	434.15	67.57	20.00	47.57	0.48	9.23E-30	431.14	69.85	20.00	49.85	0.46
L=20	**DLHS**	3.88E-29	390.60	650.85	20.00	630.85	0.06	9.57E-30	408.72	664.18	20.00	644.18	0.06
	DLHV	3.77E-29	390.57	20.00	20.00	0.00	1.00	8.8E-30	408.69	20.00	20.00	0.00	1.00
	WL	3.79E-23	3461.14	1001.00	20.00	981.00	0.04	9.94E-25	3595.24	1001.00	20.00	981.00	0.04
	WQ	3.91E-28	3768.78	1001.00	20.00	981.00	0.04	3.13E-28	3815.70	1001.00	20.00	981.00	0.04

Table 16: Results in Section 4.3 for t distribution error.

τ		0.3						0.5					
n=1000	**Method**	**Loss_beta**	**Time**	**Model_Size**	**TP**	**FP**	**F1-Score**	**Loss_beta**	**Time**	**Model_Size**	**TP**	**FP**	**F1-Score**
L=5	**DLS**	1.1E-20	203.36	1001.00	20.00	981.00	0.04	3.9E-21	183.85	1001.00	20.00	981.00	0.04
	DHXS	6.75E-29	279.71	389.40	20.00	369.40	0.10	3.52E-29	260.26	381.96	20.00	361.96	0.10
	DHXV	6.36E-29	279.69	24.28	20.00	4.28	0.91	3.22E-29	260.24	24.92	20.00	4.92	0.89
	DlHXS	6.85E-29	116.64	178.55	20.00	158.55	0.21	3.59E-29	112.43	174.99	20.00	154.99	0.22
	DlHXV	6.86E-29	116.62	87.54	20.00	67.54	0.40	3.57E-29	112.41	88.08	20.00	68.08	0.40
	DLHS	6.74E-29	78.91	230.17	20.00	210.17	0.18	3.36E-29	80.03	251.59	20.00	231.59	0.16
	DLHV	6.35E-29	78.90	21.03	20.00	1.03	0.98	3.07E-29	80.01	21.26	20.00	1.26	0.97
	WL	1.26E-20	101.06	1001.00	20.00	981.00	0.04	2.84E-22	94.49	1001.00	20.00	981.00	0.04
	WQ	5.92E-28	465.77	1001.00	20.00	981.00	0.04	5.46E-28	432.38	1001.00	20.00	981.00	0.04
L=10	**DLS**	9.9E-21	310.26	1001.00	20.00	981.00	0.04	3.74E-21	303.18	1001.00	20.00	981.00	0.04
	DHXS	5.17E-29	500.71	605.84	20.00	585.84	0.06	1.88E-29	495.02	602.84	20.00	582.84	0.06
	DHXV	4.96E-29	500.70	21.16	20.00	1.16	0.97	1.71E-29	495.01	21.58	20.00	1.58	0.96
	DlHXS	5.18E-29	223.12	215.14	20.00	195.14	0.18	1.8E-29	221.54	217.62	20.00	197.62	0.18
	DlHXV	5.12E-29	223.11	71.26	20.00	51.26	0.47	1.75E-29	221.53	72.41	20.00	52.41	0.47
	DLHS	5.17E-29	178.42	394.49	20.00	374.49	0.10	1.79E-29	179.27	396.96	20.00	376.96	0.10
	DLHV	4.96E-29	178.42	20.00	20.00	0.00	1.00	1.63E-29	179.27	20.00	20.00	0.00	1.00
	WL	3.88E-21	394.98	1001.00	20.00	981.00	0.04	4.1E-23	394.73	1001.00	20.00	981.00	0.04
	WQ	4.18E-28	1238.05	1001.00	20.00	981.00	0.04	3.49E-28	1251.96	1001.00	20.00	981.00	0.04
L=20	**DLS**	8.66E-21	1398.83	1001.00	20.00	981.00	0.04	3.24E-21	1369.48	1001.00	20.00	981.00	0.04
	DHXS	3.84E-29	1712.20	853.84	20.00	833.84	0.05	9.49E-30	1679.84	840.55	20.00	820.55	0.05
	DHXV	3.73E-29	1712.18	21.27	20.00	1.27	0.97	8.71E-30	1679.82	21.57	20.00	1.57	0.96
	DlHXS	3.72E-29	387.89	255.39	20.00	235.39	0.16	8.81E-30	394.00	247.83	20.00	227.83	0.16
	DlHXV	3.69E-29	387.87	71.93	20.00	51.93	0.45	8.56E-30	393.98	73.21	20.00	53.21	0.44
	DLHS	3.88E-29	355.66	650.85	20.00	630.85	0.06	9.57E-30	357.25	664.18	20.00	644.18	0.06
	DLHV	3.77E-29	355.64	20.00	20.00	0.00	1.00	8.8E-30	357.23	20.00	20.00	0.00	1.00
	WL	4.18E-23	3210.79	1001.00	20.00	981.00	0.04	1.35E-24	3193.46	1001.00	20.00	981.00	0.04
	WQ	3.91E-28	3430.26	1001.00	20.00	981.00	0.04	3.13E-28	3207.33	1001.00	20.00	981.00	0.04

Table 17: Results in Section 4.3 for χ^2 distribution error.

τ		0.3						0.5					
n=1000	**Method**	**Loss_beta**	**Time**	**Model_Size**	**TP**	**FP**	**F1-Score**	**Loss_beta**	**Time**	**Model_Size**	**TP**	**FP**	**F1-Score**
L=5	**DLS**	9.48E-21	187.74	1001.00	20.00	981.00	0.04	4.15E-21	184.24	1001.00	20.00	981.00	0.04
	DHXS	6.56E-29	255.07	409.06	20.00	389.06	0.10	3.52E-29	247.81	399.89	20.00	379.89	0.10
	DHXV	6.2E-29	255.05	24.85	20.00	4.85	0.90	3.22E-29	247.80	25.48	20.00	5.48	0.88
	DlHXS	6.84E-29	104.00	184.47	20.00	164.47	0.20	3.51E-29	98.86	178.55	20.00	158.55	0.21
	DlHXV	6.87E-29	103.98	93.48	20.00	73.48	0.38	3.51E-29	98.84	93.21	20.00	73.21	0.38
	DLHS	6.74E-29	66.31	230.17	20.00	210.17	0.18	3.36E-29	63.26	251.59	20.00	231.59	0.16
	DLHV	6.35E-29	66.30	21.03	20.00	1.03	0.98	3.07E-29	63.25	21.26	20.00	1.26	0.97
	WL	1.18E-20	93.82	1001.00	20.00	981.00	0.04	2.58E-22	93.33	1001.00	20.00	981.00	0.04
	WQ	5.92E-28	382.56	1001.00	20.00	981.00	0.04	5.46E-28	372.78	1001.00	20.00	981.00	0.04
L=10	**DLS**	9.07E-21	338.20	1001.00	20.00	981.00	0.04	3.69E-21	299.10	1001.00	20.00	981.00	0.04
	DHXS	5.13E-29	554.34	630.21	20.00	610.21	0.06	1.81E-29	487.25	625.91	20.00	605.91	0.06
	DHXV	4.93E-29	554.33	21.19	20.00	1.19	0.97	1.65E-29	487.24	21.61	20.00	1.61	0.96
	DlHXS	5.31E-29	245.17	220.82	20.00	200.82	0.18	1.77E-29	217.61	222.41	20.00	202.41	0.18
	DlHXV	5.27E-29	245.16	75.68	20.00	55.68	0.45	1.74E-29	217.60	75.63	20.00	55.63	0.45
	DLHS	5.17E-29	202.12	394.49	20.00	374.49	0.10	1.79E-29	176.32	396.96	20.00	376.96	0.10
	DLHV	4.96E-29	202.11	20.00	20.00	0.00	1.00	1.63E-29	176.31	20.00	20.00	0.00	1.00
	WL	3.71E-21	413.72	1001.00	20.00	981.00	0.04	4.15E-23	384.92	1001.00	20.00	981.00	0.04
	WQ	4.18E-28	1252.05	1001.00	20.00	981.00	0.04	3.49E-28	1184.61	1001.00	20.00	981.00	0.04
L=20	**DLS**	8.8E-21	1191.97	1001.00	20.00	981.00	0.04	3.33E-21	1147.64	1001.00	20.00	981.00	0.04
	DHXS	3.87E-29	1458.99	874.12	20.00	854.12	0.04	1.01E-29	1373.83	859.17	20.00	839.17	0.05
	DHXV	3.77E-29	1458.97	21.31	20.00	1.31	0.97	9.29E-30	1373.81	21.61	20.00	1.61	0.96
	DlHXS	3.79E-29	293.28	262.29	20.00	242.29	0.15	9.17E-30	265.06	251.72	20.00	231.72	0.16
	DlHXV	3.76E-29	293.27	75.57	20.00	55.57	0.43	8.97E-30	265.04	75.94	20.00	55.94	0.43
	DLHS	3.88E-29	253.82	650.85	20.00	630.85	0.06	9.57E-30	235.33	664.18	20.00	644.18	0.06
	DLHV	3.77E-29	253.80	20.00	20.00	0.00	1.00	8.8E-30	235.31	20.00	20.00	0.00	1.00
	WL	4.24E-23	2442.38	1001.00	20.00	981.00	0.04	1.19E-24	2374.21	1001.00	20.00	981.00	0.04
	WQ	3.91E-28	2638.48	1001.00	20.00	981.00	0.04	3.13E-28	2371.14	1001.00	20.00	981.00	0.04

Table 18: Results in Section 4.4 for Cauchy error.

τ		0.3						0.5					
	Method	Loss_beta	Time	Model_Size	TP	FP	F1-Score	Loss_beta	Time	Model_Size	TP	FP	F1-Score
$n = 500$ $L = 20$	DLS	1.44E-19	298.97	1001.00	20.00	981.00	0.04	1.41E-20	310.50	1001.00	20.00	981.00	0.04
	DHXS	0.171129	376.01	378.10	20.00	358.10	0.10	0.133782	387.72	382.71	20.00	362.71	0.10
	DHXV	0.264732	376.00	21.00	20.00	1.00	0.98	0.227644	387.72	21.00	20.00	1.00	0.98
	DlHXS	0.171129	90.92	378.10	20.00	358.10	0.10	0.133782	85.77	382.71	20.00	362.71	0.10
	DlHXV	0.264732	90.91	21.00	20.00	1.00	0.98	0.227644	85.76	21.00	20.00	1.00	0.98
	DLHS	3.56E-29	84.90	169.17	20.00	149.17	0.23	7.55E-30	82.69	175.54	20.00	155.54	0.23
	DLHV	3.53E-29	84.89	20.00	20.00	0.00	1.00	7.31E-30	82.68	20.01	20.00	0.01	1.00
	WL	2.95E-21	453.44	1001.00	20.00	981.00	0.04	4.52E-23	441.83	1001.00	20.00	981.00	0.04
	WQ	4.18E-28	1281.13	1001.00	20.00	981.00	0.04	3.49E-28	1193.25	1001.00	20.00	981.00	0.04
$n = 1000$ $L = 10$	DLS	1E-20	445.15	1001.00	20.00	981.00	0.04	3.65E-21	413.74	1001.00	20.00	981.00	0.04
	DHXS	5.27E-29	684.95	386.72	20.00	366.72	0.10	1.84E-29	636.29	410.89	20.00	390.89	0.10
	DHXV	5.05E-29	684.94	20.76	20.00	0.76	0.98	1.67E-29	636.28	21.05	20.00	1.05	0.97
	DlHXS	5.26E-29	276.73	173.60	20.00	153.60	0.22	1.88E-29	260.00	181.03	20.00	161.03	0.22
	DlHXV	5.14E-29	276.72	44.97	20.00	24.97	0.69	1.79E-29	259.99	49.85	20.00	29.85	0.64
	DLHS	5.17E-29	203.89	394.49	20.00	374.49	0.10	1.79E-29	192.04	396.96	20.00	376.96	0.10
	DLHV	4.96E-29	203.88	20.00	20.00	0.00	1.00	1.63E-29	192.04	20.00	20.00	0.00	1.00
	WL	3.5E-21	447.99	1001.00	20.00	981.00	0.04	4.5E-23	432.51	1001.00	20.00	981.00	0.04
	WQ	4.18E-28	1136.71	1001.00	20.00	981.00	0.04	3.49E-28	1115.95	1001.00	20.00	981.00	0.04
$n = 2000$ $L = 5$	DLS	3.74E-20	1017.22	1001.00	20.00	981.00	0.04	4.63E-23	535.46	1001.00	20.00	981.00	0.04
	DHXS	8.73E-29	2434.47	573.10	20.00	553.10	0.07	4.26E-29	1063.19	604.24	20.00	584.24	0.06
	DHXV	8.02E-29	2434.46	50.58	20.00	30.58	0.59	3.72E-29	1063.18	55.15	20.00	35.15	0.56
	DlHXS	8.53E-29	1279.94	238.21	20.00	218.21	0.16	4.15E-29	633.26	243.83	20.00	223.83	0.15
	DlHXV	8.44E-29	1279.93	182.42	20.00	162.42	0.21	4.08E-29	633.25	188.40	20.00	168.40	0.20
	DLHS	8.26E-29	1206.60	593.93	20.00	573.93	0.07	4.03E-29	570.31	587.11	20.00	567.11	0.07
	DLHV	7.53E-29	1206.59	51.90	20.00	31.90	0.58	3.52E-29	570.30	51.78	20.00	31.78	0.59
	WL	5E-21	444.79	1001.00	20.00	981.00	0.04	4.54E-23	527.22	1001.00	20.00	981.00	0.04
	WQ	4.52E-28	1030.94	1001.00	20.00	981.00	0.04	3.11E-28	712.03	1001.00	20.00	981.00	0.04

Table 19: Results in Section 4.4 for Exponential error.

τ		0.3						0.5					
	Method	Loss_beta	Time	Model_Size	TP	FP	F1-Score	Loss_beta	Time	Model_Size	TP	FP	F1-Score
$n = 500$ $L = 20$	**DLS**	1.39E-19	276.88	1001.00	20.00	981.00	0.04	1.28E-20	269.44	1001.00	20.00	981.00	0.04
	DHXS	0.171129	348.20	378.10	20.00	358.10	0.10	0.133782	340.93	382.71	20.00	362.71	0.10
	DHXV	0.264732	348.19	21.00	20.00	1.00	0.98	0.227644	340.92	21.00	20.00	1.00	0.98
	DlHXS	0.171129	85.81	378.10	20.00	358.10	0.10	0.133782	84.47	382.71	20.00	362.71	0.10
	DlHXV	0.264732	85.80	21.00	20.00	1.00	0.98	0.227644	84.46	21.00	20.00	1.00	0.98
	DLHS	3.56E-29	80.71	169.17	20.00	149.17	0.23	7.55E-30	79.45	175.54	20.00	155.54	0.23
	DLHV	3.53E-29	80.70	20.00	20.00	0.00	1.00	7.31E-30	79.44	20.01	20.00	0.01	1.00
	WL	3.79E-21	420.51	1001.00	20.00	981.00	0.04	4.41E-23	403.41	1001.00	20.00	981.00	0.04
	WQ	4.18E-28	1295.64	1001.00	20.00	981.00	0.04	3.49E-28	1209.19	1001.00	20.00	981.00	0.04
$n = 1000$ $L = 10$	**DLS**	9.47E-21	341.08	1001.00	20.00	981.00	0.04	3.6E-21	347.59	1001.00	20.00	981.00	0.04
	DHXS	5.3E-29	556.74	508.53	20.00	488.53	0.08	1.8E-29	562.73	518.79	20.00	498.79	0.08
	DHXV	5.09E-29	556.73	21.01	20.00	1.01	0.98	1.64E-29	562.72	21.34	20.00	1.34	0.97
	DlHXS	5.24E-29	239.46	195.30	20.00	175.30	0.20	1.84E-29	238.91	200.84	20.00	180.84	0.19
	DlHXV	5.14E-29	239.45	58.62	20.00	38.62	0.56	1.78E-29	238.90	60.07	20.00	40.07	0.55
	DLHS	5.17E-29	178.33	394.49	20.00	374.49	0.10	1.79E-29	175.29	396.96	20.00	376.96	0.10
	DLHV	4.96E-29	178.32	20.00	20.00	0.00	1.00	1.63E-29	175.28	20.00	20.00	0.00	1.00
	WL	3.03E-21	376.99	1001.00	20.00	981.00	0.04	4.11E-23	386.24	1001.00	20.00	981.00	0.04
	WQ	4.18E-28	1167.00	1001.00	20.00	981.00	0.04	3.49E-28	1188.55	1001.00	20.00	981.00	0.04
$n = 2000$ $L = 5$	**DLS**	3.74E-20	605.01	1001.00	20.00	981.00	0.04	3.95E-23	409.97	1001.00	20.00	981.00	0.04
	DHXS	8.69E-29	1734.60	640.67	20.00	620.67	0.06	4.57E-29	927.24	688.09	20.00	668.09	0.06
	DHXV	7.99E-29	1734.59	61.93	20.00	41.93	0.51	4.06E-29	927.23	71.92	20.00	51.92	0.45
	DlHXS	8.49E-29	1272.25	254.70	20.00	234.70	0.15	4.37E-29	617.33	270.70	20.00	250.70	0.14
	DlHXV	8.45E-29	1272.24	199.98	20.00	179.98	0.19	4.38E-29	617.33	217.30	20.00	197.30	0.17
	DLHS	8.26E-29	1174.76	593.93	20.00	573.93	0.07	4.03E-29	574.19	587.11	20.00	567.11	0.07
	DLHV	7.53E-29	1174.75	51.90	20.00	31.90	0.58	3.52E-29	574.18	51.78	20.00	31.78	0.59
	WL	5.5E-21	333.46	1001.00	20.00	981.00	0.04	4.53E-23	312.78	1001.00	20.00	981.00	0.04
	WQ	4.52E-28	925.91	1001.00	20.00	981.00	0.04	3.11E-28	641.45	1001.00	20.00	981.00	0.04

Table 20: Results in Section 4.4 for Normal error.

τ		0.3						0.5					
	Method	**Loss_beta**	**Time**	**Model_Size**	**TP**	**FP**	**F1-Score**	**Loss_beta**	**Time**	**Model_Size**	**TP**	**FP**	**F1-Score**
$n = 500$ $L = 20$	**DLS**	1.39E-19	268.83	1001.00	20.00	981.00	0.04	1.43E-20	253.57	1001.00	20.00	981.00	0.04
	DHXS	0.171129	337.93	378.10	20.00	358.10	0.10	0.133782	319.59	382.71	20.00	362.71	0.10
	DHXV	0.264732	337.92	21.00	20.00	1.00	0.98	0.227644	319.58	21.00	20.00	1.00	0.98
	DlHXS	0.171129	82.67	378.10	20.00	358.10	0.10	0.133782	79.25	382.71	20.00	362.71	0.10
	DlHXV	0.264732	82.66	21.00	20.00	1.00	0.98	0.227644	79.24	21.00	20.00	1.00	0.98
	DLHS	3.56E-29	77.38	169.17	20.00	149.17	0.23	7.55E-30	75.35	175.54	20.00	155.54	0.23
	DLHV	3.53E-29	77.37	20.00	20.00	0.00	1.00	7.31E-30	75.34	20.01	20.00	0.01	1.00
	WL	3.41E-21	412.99	1001.00	20.00	981.00	0.04	3.91E-23	402.77	1001.00	20.00	981.00	0.04
	WQ	4.18E-28	1291.33	1001.00	20.00	981.00	0.04	3.49E-28	1261.48	1001.00	20.00	981.00	0.04
$n = 1000$ $L = 10$	**DLS**	9.54E-21	321.09	1001.00	20.00	981.00	0.04	3.57E-21	324.22	1001.00	20.00	981.00	0.04
	DHXS	5.2E-29	524.51	567.41	20.00	547.41	0.07	1.8E-29	526.13	571.43	20.00	551.43	0.07
	DHXV	4.99E-29	524.50	21.09	20.00	1.09	0.97	1.64E-29	526.13	21.50	20.00	1.50	0.96
	DlHXS	5.29E-29	232.81	207.32	20.00	187.32	0.19	1.75E-29	231.41	210.59	20.00	190.59	0.19
	DlHXV	5.22E-29	232.81	66.36	20.00	46.36	0.51	1.7E-29	231.40	67.18	20.00	47.18	0.50
	DLHS	5.17E-29	181.60	394.49	20.00	374.49	0.10	1.79E-29	180.44	396.96	20.00	376.96	0.10
	DLHV	4.96E-29	181.59	20.00	20.00	0.00	1.00	1.63E-29	180.44	20.00	20.00	0.00	1.00
	WL	3.93E-21	390.58	1001.00	20.00	981.00	0.04	4.16E-23	389.71	1001.00	20.00	981.00	0.04
	WQ	4.18E-28	1211.53	1001.00	20.00	981.00	0.04	3.49E-28	1228.60	1001.00	20.00	981.00	0.04
$n = 2000$ $L = 5$	**DLS**	3.77E-20	517.61	1001.00	20.00	981.00	0.04	3.92E-23	391.02	1001.00	20.00	981.00	0.04
	DHXS	8.7E-29	1386.30	673.80	20.00	653.80	0.06	4.22E-29	930.15	724.88	20.00	704.88	0.05
	DHXV	8.01E-29	1386.29	69.78	20.00	49.78	0.46	3.75E-29	930.14	83.10	20.00	63.10	0.40
	DlHXS	8.29E-29	902.02	263.68	20.00	243.68	0.14	4.26E-29	618.18	282.03	20.00	262.03	0.13
	DlHXV	8.31E-29	902.01	211.93	20.00	191.93	0.18	4.3E-29	618.17	230.71	20.00	210.71	0.16
	DLHS	8.26E-29	672.48	593.93	20.00	573.93	0.07	4.03E-29	559.05	587.11	20.00	567.11	0.07
	DLHV	7.53E-29	672.47	51.90	20.00	31.90	0.58	3.52E-29	559.04	51.78	20.00	31.78	0.59
	WL	5.72E-21	224.09	1001.00	20.00	981.00	0.04	4.06E-23	212.74	1001.00	20.00	981.00	0.04
	WQ	4.52E-28	623.16	1001.00	20.00	981.00	0.04	3.11E-28	594.56	1001.00	20.00	981.00	0.04

Table 21: Results in Section 4.4 for t distribution error.

τ		0.3						0.5					
	Method	Loss_beta	Time	Model_Size	TP	FP	F1-Score	Loss_beta	Time	Model_Size	TP	FP	F1-Score
$n = 500$ $L = 20$	**DLS**	1.41E-19	270.47	1001.00	20.00	981.00	0.04	1.41E-20	275.37	1001.00	20.00	981.00	0.04
	DHXS	0.171129	338.79	378.10	20.00	358.10	0.10	0.133782	344.72	382.71	20.00	362.71	0.10
	DHXV	0.264732	338.78	21.00	20.00	1.00	0.98	0.227644	344.71	21.00	20.00	1.00	0.98
	DlHXS	0.171129	81.29	378.10	20.00	358.10	0.10	0.133782	84.57	382.71	20.00	362.71	0.10
	DlHXV	0.264732	81.28	21.00	20.00	1.00	0.98	0.227644	84.56	21.00	20.00	1.00	0.98
	DLHS	3.56E-29	76.24	169.17	20.00	149.17	0.23	7.55E-30	79.80	175.54	20.00	155.54	0.23
	DLHV	3.53E-29	76.23	20.00	20.00	0.00	1.00	7.31E-30	79.79	20.01	20.00	0.01	1.00
	WL	3.42E-21	417.17	1001.00	20.00	981.00	0.04	4.18E-23	419.87	1001.00	20.00	981.00	0.04
	WQ	4.18E-28	1320.06	1001.00	20.00	981.00	0.04	3.49E-28	1361.77	1001.00	20.00	981.00	0.04
$n = 1000$ $L = 10$	**DLS**	9.9E-21	310.26	1001.00	20.00	981.00	0.04	3.74E-21	303.18	1001.00	20.00	981.00	0.04
	DHXS	5.17E-29	500.71	605.84	20.00	585.84	0.06	1.88E-29	495.02	602.84	20.00	582.84	0.06
	DHXV	4.96E-29	500.70	21.16	20.00	1.16	0.97	1.71E-29	495.01	21.58	20.00	1.58	0.96
	DlHXS	5.18E-29	223.12	215.14	20.00	195.14	0.18	1.8E-29	221.54	217.62	20.00	197.62	0.18
	DlHXV	5.12E-29	223.11	71.26	20.00	51.26	0.47	1.75E-29	221.53	72.41	20.00	52.41	0.47
	DLHS	5.17E-29	178.42	394.49	20.00	374.49	0.10	1.79E-29	179.27	396.96	20.00	376.96	0.10
	DLHV	4.96E-29	178.42	20.00	20.00	0.00	1.00	1.63E-29	179.27	20.00	20.00	0.00	1.00
	WL	3.88E-21	394.98	1001.00	20.00	981.00	0.04	4.1E-23	394.73	1001.00	20.00	981.00	0.04
	WQ	4.18E-28	1238.05	1001.00	20.00	981.00	0.04	3.49E-28	1251.96	1001.00	20.00	981.00	0.04
$n = 2000$ $L = 5$	**DLS**	3.84E-20	377.70	1001.00	20.00	981.00	0.04	4.61E-23	372.66	1001.00	20.00	981.00	0.04
	DHXS	8.61E-29	937.97	693.38	20.00	673.38	0.06	4.14E-29	911.22	746.45	20.00	726.45	0.05
	DHXV	7.95E-29	937.96	74.78	20.00	54.78	0.44	3.69E-29	911.21	90.51	20.00	70.51	0.37
	DlHXS	8.55E-29	640.26	269.14	20.00	249.14	0.14	4.21E-29	623.14	292.01	20.00	272.01	0.13
	DlHXV	8.58E-29	640.25	219.36	20.00	199.36	0.17	4.28E-29	623.13	241.63	20.00	221.63	0.16
	DLHS	8.26E-29	541.05	593.93	20.00	573.93	0.07	4.03E-29	556.92	587.11	20.00	567.11	0.07
	DLHV	7.53E-29	541.04	51.90	20.00	31.90	0.58	3.52E-29	556.91	51.78	20.00	31.78	0.59
	WL	5.41E-21	157.54	1001.00	20.00	981.00	0.04	4.14E-23	196.48	1001.00	20.00	981.00	0.04
	WQ	4.52E-28	499.41	1001.00	20.00	981.00	0.04	3.11E-28	613.30	1001.00	20.00	981.00	0.04

Table 22: Results in Section 4.4 for χ^2 distribution error.

τ		0.3						0.5					
	Method	Loss_beta	Time	Model_Size	TP	FP	F1-Score	Loss_beta	Time	Model_Size	TP	FP	F1-Score
$n = 500$ $L = 20$	DLS	1.39E-19	265.40	1001.00	20.00	981.00	0.04	1.29E-20	283.82	1001.00	20.00	981.00	0.04
	DHXS	0.171129	331.87	378.10	20.00	358.10	0.10	0.133782	355.11	382.71	20.00	362.71	0.10
	DHXV	0.264732	331.86	21.00	20.00	1.00	0.98	0.227644	355.10	21.00	20.00	1.00	0.98
	DlHXS	0.171129	79.78	378.10	20.00	358.10	0.10	0.133782	84.60	382.71	20.00	362.71	0.10
	DlHXV	0.264732	79.77	21.00	20.00	1.00	0.98	0.227644	84.59	21.00	20.00	1.00	0.98
	DLHS	3.56E-29	75.13	169.17	20.00	149.17	0.23	7.55E-30	80.03	175.54	20.00	155.54	0.23
	DLHV	3.53E-29	75.12	20.00	20.00	0.00	1.00	7.31E-30	80.02	20.01	20.00	0.01	1.00
	WL	3.03E-21	409.01	1001.00	20.00	981.00	0.04	4.11E-23	422.92	1001.00	20.00	981.00	0.04
	WQ	4.18E-28	1277.28	1001.00	20.00	981.00	0.04	3.49E-28	1322.00	1001.00	20.00	981.00	0.04
$n = 1000$ $L = 10$	DLS	9.07E-21	338.20	1001.00	20.00	981.00	0.04	3.69E-21	299.10	1001.00	20.00	981.00	0.04
	DHXS	5.13E-29	554.34	630.21	20.00	610.21	0.06	1.81E-29	487.25	625.91	20.00	605.91	0.06
	DHXV	4.93E-29	554.33	21.19	20.00	1.19	0.97	1.65E-29	487.24	21.61	20.00	1.61	0.96
	DlHXS	5.31E-29	245.17	220.82	20.00	200.82	0.18	1.77E-29	217.61	222.41	20.00	202.41	0.18
	DlHXV	5.27E-29	245.16	75.68	20.00	55.68	0.45	1.74E-29	217.60	75.63	20.00	55.63	0.45
	DLHS	5.17E-29	202.12	394.49	20.00	374.49	0.10	1.79E-29	176.32	396.96	20.00	376.96	0.10
	DLHV	4.96E-29	202.11	20.00	20.00	0.00	1.00	1.63E-29	176.31	20.00	20.00	0.00	1.00
	WL	3.71E-21	413.72	1001.00	20.00	981.00	0.04	4.15E-23	384.92	1001.00	20.00	981.00	0.04
	WQ	4.18E-28	1252.05	1001.00	20.00	981.00	0.04	3.49E-28	1184.61	1001.00	20.00	981.00	0.04
$n = 2000$ $L = 5$	DLS	3.74E-20	309.29	1001.00	20.00	981.00	0.04	4.04E-23	357.78	1001.00	20.00	981.00	0.04
	DHXS	8.43E-29	799.26	706.64	20.00	686.64	0.06	4.4E-29	876.85	759.74	20.00	739.74	0.05
	DHXV	7.79E-29	799.25	78.69	20.00	58.69	0.42	3.97E-29	876.85	96.51	20.00	76.51	0.35
	DlHXS	8.36E-29	543.90	273.41	20.00	253.41	0.14	4.16E-29	590.98	300.14	20.00	280.14	0.13
	DlHXV	8.42E-29	543.89	224.09	20.00	204.09	0.17	4.27E-29	590.97	249.29	20.00	229.29	0.15
	DLHS	8.26E-29	466.61	593.93	20.00	573.93	0.07	4.03E-29	523.52	587.11	20.00	567.11	0.07
	DLHV	7.53E-29	466.61	51.90	20.00	31.90	0.58	3.52E-29	523.52	51.78	20.00	31.78	0.59
	WL	5.7E-21	133.61	1001.00	20.00	981.00	0.04	4.4E-23	171.33	1001.00	20.00	981.00	0.04
	WQ	4.52E-28	428.69	1001.00	20.00	981.00	0.04	3.11E-28	530.15	1001.00	20.00	981.00	0.04