Statistica Sinica Preprint No: SS-2020-0006	
Title	Quantile Martingale Difference Divergence for Dimension Reduction
Manuscript ID	SS-2020-0006
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202020.0006
Complete List of Authors	Chung Eun Lee and Haileab Hilafu
Corresponding Author	Chung Eun Lee
E-mail	clee88@utk.edu
Notice: Accepted version subject to English editing.	

Quantile Martingale Difference Divergence for Dimension Reduction

Chung Eun Lee and Haileab Hilafu

The University of Tennessee, Knoxville, TN 37996.

Abstract: In this article, we utilize two metrics, the quantile martingale difference divergence (QMDD) and the quantile martingale difference divergence matrix (QMDDM) which measure the quantile dependence of a scalar response variable and a vector of predictors, to reduce the dimension of the predictors by considering the central quantile subspace or central subspace. The proposed dimension reduction methods do not involve user-chosen parameters or assume a parametric model, and are simple to implement. Extensive simulations and real data illustration are provided to demonstrate the usefulness of the proposed methods. The proposed methods are shown to yield competitive finite sample performances. Theoretical properties are also provided.

Key words and phrases: Central subspace, Dimension reduction, Quantile dependence.

We would like to thank Dr. Christou for providing the R code used in Christou (2019), and the editor, the associate editor, and the referee for thoughtful comments that led to a substantial improvement of the paper.

1. Introduction

The sufficient dimension reduction (SDR) in regression paradigm (Li 1991, Cook 1998) incorporates dimension reduction with the concept of sufficiency to attain low-dimensional predictors without loss of information on the regression of $Y \in \mathbb{R}$ on $\mathbf{X} \in \mathbb{R}^p$. A subspace spanned by the columns of a matrix $\boldsymbol{\beta} \in \mathbb{R}^{p \times d'}$, $d' \leq p$ is said to be a sufficient dimension reduction subspace if $Y \perp\!\!\!\perp \mathbf{X} \mid \boldsymbol{\beta}^\top \mathbf{X}$, where $\perp\!\!\!\perp$ denotes independence. Furthermore, the minimal sufficient dimension reduction subspace is called the central subspace and is formally defined as

$$\mathcal{S}_{Y|\mathbf{X}} = \cap\{\text{span}(\boldsymbol{\beta}) : Y \perp\!\!\!\perp \mathbf{X} \mid \boldsymbol{\beta}^\top \mathbf{X}\} := \text{span}(\mathbf{B}), \tag{1.1}$$

where $\mathbf{B} \in \mathbb{R}^{p \times d}$. During the past two decades, there has been a growing literature on SDR; see the excellent reviews by Cook (1998), Li (2018), and Ma & Zhu (2013). The developments have been mainly devoted to the approaches targeting the central subspace. For example, the sliced inverse regression (Li 1991), the sliced average variance estimation (Cook & Weisberg 1991), the parametric inverse regression (Bura & Cook 2001), the contour regression (Li et al. 2005), the directional regression (Li & Wang 2007), and the cumulative slicing estimation (Zhu et al. 2010), the Fourier

method (Zhu & Zeng 2006), the ensemble of minimum average variance estimators method (Yin & Li 2011), and the sufficient dimension reduction method via distance covariance (Sheng & Yin 2016), among others.

In many applications, the focus of a regression analysis is a particular characteristic of Y given $\mathbf{X}$, instead of the entire conditional distribution of Y given $\mathbf{X}$. Methods along this line of research include the central mean subspace (Cook & Li 2002), the central k-th moment subspace (Yin & Cook 2002) , the central variance subspace (Zhu & Zhu 2009), and the T-central subspace (Luo et al. 2014). Later, Kong & Xia (2012) and Kong & Xia (2014) proposed the quantile outer-product of gradients (QOPG)-based method which can be used to estimate the central quantile subspace (Luo et al. 2014); see (3.6) for the definition of the central quantile subspace. Recently, Christou (2019) proposed an efficient algorithm for finding the central quantile subspace and generalized the approach considering any statistical functional of interest. However, the approach implicitly assumes the linear model for the conditional quantile of Y given $\mathbf{X}$, when the structural dimension of the central quantile subspace d_τ is one. Also, the both procedures in Kong & Xia (2014) and Christou (2019) rely on the use of nonparametric regression, and thus require several user-chosen quantities.

In this paper, we consider the semi-parametric model below and propose

a new approach estimating the central quantile subspace.

$$Q_\tau(Y \mid \mathbf{X}) = g(\mathbf{B}_\tau^\top \mathbf{X}, \varepsilon), \tag{1.2}$$

where $Q_\tau(Y \mid \mathbf{X})$ is the conditional τ-th quantile of Y given $\mathbf{X}$, g is an arbitrary link function, $\mathbf{B}_\tau \in \mathbb{R}^{p \times d_\tau}$ is a matrix which spans the central quantile subspace, and ε is an error term independent of $\mathbf{X}$. Throughout this article, we assume that the central quantile subspace and central subspace exist. We refer to Luo et al. (2014), Christou (2019), Cook (1998), and Yin et al. (2008) for many interesting details on the central quantile subspace and the central subspace. We first introduce a variant of the martingale difference divergence (MDD) (Shao & Zhang 2014) the so-called quantile martingale difference divergence (QMDD) which measures the quantile dependence, and apply QMDD to estimate the central quantile subspace. An appealing feature of our approach is that it does not impose any parametric structure between the conditional quantile of Y given $\mathbf{X}$ and $\mathbf{B}_\tau^\top \mathbf{X}$, and it does not involve any tuning parameters, so it is simple and easy to implement. Moreover, we introduce a new bootstrap test to determine the dimension of the central quantile subspace. Further, we propose an inverse regression method to estimate the central subspace by developing the quantile

martingale difference divergence matrix (QMDDM)-based approach which is also computationally efficient as other inverse regression methods with numerically stable estimates.

The remainder of the paper is organized as follows. In Section 2, we give a brief review of the martingale difference divergence (Shao & Zhang 2014) and the martingale difference divergence matrix (MDDM) (Lee & Shao 2018). We introduce QMDD, its properties, and application to estimating the central quantile subspace in Section 3. In Section 4, we introduce the QMDDM-based approach to seek the central subspace. Section 5 presents numerical studies on synthetic data, and Section 6 presents application of the proposed method to a real data set. Section 7 contains summary and discussion. All proofs are relegated to the supplementary material.

A word on notation. Let $i = \sqrt{-1}$ be the imaginary unit. Scalar product of vectors x and y is $\langle x, y\rangle$. For a complex-valued function $f(\cdot)$, $|f|^2 = f\bar{f}$, where $\bar{f}$ is the complex conjugate of f. Denote the Euclidean norm of $\mathbf{X} = (x_1, \cdots, x_p) \in \mathbb{R}^p$ as $\|\mathbf{X}\|$, where $\|\mathbf{X}\|^2 = \sum_{i=1}^p x_i^2$. For a square matrix $\mathbf{A} = (\mathbf{A}_{i,j})_{i,j=1}^p \in \mathbb{R}^{p\times p}$, the spectral norm of $\mathbf{A}$ is denoted as $\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^\top\mathbf{A})}$, where $\lambda_{\max}(\mathbf{A}^\top\mathbf{A})$ is the largest eigenvalue of $\mathbf{A}^\top\mathbf{A}$, Forbenius norm is denoted as $\|\mathbf{A}\|_F = \sqrt{tr(\mathbf{A}^\top\mathbf{A})}$ and $tr(\mathbf{A}) = \sum_{i=1}^p \mathbf{A}_{i,i}$. For $\mathbf{X} \in \mathbb{R}^p$, $\mathbf{X} \in \mathcal{L}^2$ if $\mathbb{E}\|\mathbf{X}\|^2 < \infty$. The orthogonal complement of $\mathcal{S}$ is

$\mathcal{S}^{\perp}$.

2. Review of MDD and MDDM

To introduce the new approaches for dimension reduction, we briefly review the martingale difference divergence (MDD) and the martingale difference divergence matrix (MDDM). For $\mathbf{U} \in \mathbb{R}^r$ and $V \in \mathbb{R}$, where r is a fixed positive integer, Shao & Zhang (2014) proposed MDD to measure the mean dependence of V on $\mathbf{U}$, i.e.,

$$\mathbb{E}(V \mid \mathbf{U}) = \mathbb{E}(V), \text{ almost surely.} \tag{2.3}$$

Specifically if $\mathbb{E}(|V|^2 + \|\mathbf{U}\|^2) < \infty$, $\mathrm{MDD}^2(V \mid \mathbf{U})$ is defined as the non-negative number which is

$$\mathrm{MDD}^2(V \mid \mathbf{U}) = -\mathbb{E}\left[\{V - \mathbb{E}(V)\}\{V' - \mathbb{E}(V')\}\|\mathbf{U} - \mathbf{U}'\|\right], \tag{2.4}$$

where $(V', \mathbf{U}')$ is an independent copy of $(V, \mathbf{U})$. The key property is that $\mathrm{MDD}^2(V \mid \mathbf{U}) = 0$ if and only if (2.3) holds, so it fully characterizes the mean independence of V on $\mathbf{U}$.

Recently, Lee & Shao (2018) introduced MDDM, which can be viewed as an extension of martingale difference divergence from a scalar to a matrix

and further applied it to the dimension reduction for the conditional mean of a multivariate time series. For two random vectors $\mathbf{V} \in \mathbb{R}^u$ and $\mathbf{U} \in \mathbb{R}^r$ with $\mathbb{E}(\|\mathbf{V}\|^2 + \|\mathbf{U}\|^2) < \infty$, $\mathrm{MDDM}(\mathbf{V} \mid \mathbf{U})$ is defined as

$$\mathrm{MDDM}(\mathbf{V} \mid \mathbf{U}) = -\mathbb{E}\left[\{\mathbf{V} - \mathbb{E}(\mathbf{V})\}\{\mathbf{V}' - \mathbb{E}(\mathbf{V}')\}^\top \|\mathbf{U} - \mathbf{U}'\|\right], \tag{2.5}$$

where $(\mathbf{V}', \mathbf{U}')$ is an independent copy of $(\mathbf{V}, \mathbf{U})$. From (2.5), it is easy to see that $\mathrm{MDDM}(\mathbf{V} \mid \mathbf{U}) \in \mathbb{R}^{u \times u}$ is a real, symmetric, and positive semi-definite matrix.

3. Central Quantile Subspace

Often times, the interest of a regression analysis may be the conditional τ-th quantile of Y given $\mathbf{X}$, where $\tau \in (0, 1)$. To this end, we seek the central quantile subspace which is introduced by Luo et al. (2014) and its definition is provided below.

Definition 1. The central quantile subspace for a given $\tau \in (0, 1)$ is defined as

$$\begin{aligned} \mathcal{S}_{Q(Y|\mathbf{X})}(\tau) &= \cap\{\mathrm{span}(\boldsymbol{\beta}) : Q_\tau(Y \mid \mathbf{X}) = Q_\tau(Y \mid \boldsymbol{\beta}^\top \mathbf{X}) \text{ almosty surely}\} \\ &:= \mathrm{span}(\mathbf{B}_\tau). \end{aligned} \tag{3.6}$$

It is important to note that the space $\mathcal{S}_{Q(Y|\mathbf{X})}(\tau) = \text{span}(\mathbf{B}_\tau)$ is identifiable but not $\mathbf{B}_\tau$ itself. Thus, we shall seek the identifiable parameter, $\mathcal{S}_{Q(Y|\mathbf{X})}(\tau)$; see Luo et al. (2014), Sheng & Yin (2016), and Li (2018) for more background on the identifiability of the spaces spanning the central quantile subspace and the central subspace. We first introduce a variant of MDD which is central to our new approach. Its definition and properties are introduced in the following section.

3.1 Quantile Martingle Difference Divergence

By using MDD in (2.4), we state a formal definition of a natural analogue of martingale difference divergence that quantifies the quantile dependence between a random variable Y and a random vector $\mathbf{X}$.

Definition 2. For a continuous random variable Y, a random vector $\mathbf{X} \in \mathcal{L}^2$ and $\tau \in (0,1)$, the τ-th quantile martingale difference divergence is defined as

$$\text{QMDD}_\tau(Y \mid \mathbf{X}) = -\mathbb{E}\left[\{\mathbf{1}(Y \le y_\tau) - \tau\}\{\mathbf{1}(Y' \le y_\tau) - \tau\}\|\mathbf{X} - \mathbf{X}'\|\right],$$

where $(\mathbf{X}', Y')$ is an independent copy of $(\mathbf{X}, Y)$ and y_τ is the unconditional τ-th quantile of Y.

The quantile martingale difference divergence is a special case of MDD so inherits the key property of MDD which is stated in the following proposition. The proof is omitted since the proposition is obtained by the direct consequence of Proposition 1 in Lee et al. (2019).

Proposition 1. *For a continuous random variable $Y \in \mathbb{R}$, a random vector $\mathbf{X} \in \mathcal{L}^2$, and a given $\tau \in (0,1)$, we have*

1. $\mathrm{QMDD}_\tau(Y \mid \mathbf{X}) \geq 0$.
2. $\mathrm{QMDD}_\tau(Y \mid \mathbf{X}) = 0$ *if and only if* $P(Y \leq y_\tau \mid \mathbf{X}) = P(Y \leq y_\tau)$ *almost surely.*

Inspired by the sample estimation of MDD^2 in Shao & Zhang (2014), we construct the estimator of QMDD as below.

Definition 3. Given the iid observations $(\mathbf{X}_i, Y_i)_{i=1}^n$ from the joint distribution of $(\mathbf{X}, Y)$, the sample τ-th quantile martingale difference divergence is a nonnegative number defined as $\widehat{\mathrm{QMDD}}_\tau(Y \mid \mathbf{X}) = \frac{-1}{n^2}\sum_{i,j=1}^n \{\mathbf{1}(Y_i \leq \widehat{y}_\tau) - \tau\}\{\mathbf{1}(Y_j \leq \widehat{y}_\tau) - \tau\}\|\mathbf{X}_i - \mathbf{X}_j\|$, where $\widehat{y}_\tau$ is the empirical unconditional τ-th quantile of Y.

Using the arguments in Section 2 of Lee & Shao (2018), it can be shown that $\widehat{\mathrm{QMDD}}_\tau(Y \mid \mathbf{X})$ is nonnegative and is a biased estimator of

$\mathrm{QMDD}_\tau(Y \mid \mathbf{X})$, where the bias is asymptotically negligible when p is fixed. It is possible to define an unbiased estimator of $\mathrm{QMDD}_\tau(Y \mid \mathbf{X})$ by adopting $\mathcal{U}$-centering approach (Székely & Rizzo (2014), Park et al. (2015), Zhang et al. (2018)). However, the nonnegativeness is more preferred than the unbiasedness for the optimization step in Section 3.2.

3.2 Estimation of the Central Quantile Subspace

Our specific goal is to find the linear combinations of $\mathbf{X}$, say $\mathbf{B}_\tau^\top \mathbf{X}$, $\mathbf{B}_\tau \in \mathbb{R}^{p\times d_\tau}$, which can fully describe the conditional quantile of Y given $\mathbf{X}$ for a pre-specified $\tau \in (0,1)$. In other words, $Q_\tau(Y \mid \mathbf{X}) = Q_\tau(Y \mid \mathbf{B}_\tau^\top \mathbf{X})$ almost surely, where $\mathbf{B}_\tau$ constructs the central quantile subspace. This implies that modeling τ-th conditional quantile of Y as a function of $\mathbf{X}$ can be simplified by replacing $\mathbf{X}$ with a lower dimensional $\mathbf{B}_\tau^\top \mathbf{X}$ without loss of regression information. Interestingly, under the Condition 1 below and using results in Cook (1998) and Li et al. (2015), we have $\mathbf{1}(Y \le y_\tau)$ is independent of $\mathbf{B}_{0,\tau}^\top \mathbf{X}$, where $(\mathbf{B}_\tau, \mathbf{B}_{0,\tau}) \in \mathbb{R}^{p\times p}$ is an orthogonal matrix. This further implies

$$P(Y \le y_\tau \mid \mathbf{B}_{0,\tau}^\top \mathbf{X}) = P(Y \le y_\tau) \text{ almost surely.} \tag{3.7}$$

Condition 1. Suppose $\mathbf{B}_\tau$ and $\mathbf{B}_{0,\tau}$ are the basis which span $\mathcal{S}_{Q(Y|\mathbf{X})}(\tau)$ and $\mathcal{S}^{\perp}_{Q(Y|\mathbf{X})}(\tau)$, respectively. Assume that $\mathbf{B}_\tau^\top \mathbf{X}$ is independent of $\mathbf{B}_{0,\tau}^\top \mathbf{X}$.

Condition 1 is an analogue of the assumptions in Proposition 1 of Sheng & Yin (2013) and Proposition 2 of Sheng & Yin (2016), where the assumption is made for the basis associated with the central quantile subspace and its orthogonal complement. As it is mentioned in Sheng & Yin (2013) and Sheng & Yin (2016), this assumption is not as strong as it seems to be, and it could be satisfied asymptotically when p is reasonably large; see section 3.5 in Sheng & Yin (2013) for more discussion.

Suppose the Condition 1 holds and the structural dimension d_τ is known. Then for a given τ, we have (3.7) which is equivalent with

$$\mathrm{QMDD}_\tau(\mathrm{Y} \mid \mathbf{B}_{0,\tau}^\top \mathbf{X}) = 0.$$

Motivated by this fact, we propose the following optimization with the objective function $G_\tau(\boldsymbol{\beta}_0) = \mathrm{QMDD}_\tau(\mathrm{Y} \mid \boldsymbol{\beta}_0^\top \mathbf{X})$. Our estimator $\mathbf{B}_{0,\tau}$ is

$$\widehat{\mathbf{B}}_{0,\tau} = \underset{\boldsymbol{\beta}_0^\top \boldsymbol{\beta}_0 = I_{p-d_\tau}}{\operatorname{argmin}} \ \widehat{G}_\tau(\boldsymbol{\beta}_0), \tag{3.8}$$

where $\widehat{G}_\tau(\cdot)$ is the sample counterpart of $G_\tau(\cdot)$. In order to optimize (3.8)

with efficient computational cost, we use the optimization solver with orthogonality constraint proposed by Wen & Yin (2013). They use the efficient first-order updating procedure which preserves the orthogonality constraint, so it can achieve substantial saving in computational time; see Wen & Yin (2013) for more details.

Next, we show that the estimator proposed above yields consistent estimate. Since $\mathbf{B}_\tau$ or $\mathbf{B}_{0,\tau}$ are only identifiable up to span($\mathbf{B}_\tau$) or span($\mathbf{B}_{0,\tau}$), we define the following distance and show the theoretical result. For semi-orthogonal matrices $\mathbf{H}_1,\ \mathbf{H}_2 \in \mathbb{R}^{p\times p-d_\tau}$, i.e., $\mathbf{H}_1^\top\mathbf{H}_1 = \mathbf{H}_2^\top\mathbf{H}_2 = I_{p-d_\tau}$,

$$\mathcal{D}(\mathbf{H}_1, \mathbf{H}_2) = \sqrt{p - d_\tau - tr(\mathbf{H}_1\mathbf{H}_1^\top\mathbf{H}_2\mathbf{H}_2^\top)}. \tag{3.9}$$

Note that $\mathcal{D}(\mathbf{H}_1, \mathbf{H}_2) = 0$ if and only if span($\mathbf{H}_1$) = span($\mathbf{H}_2$). In preparation, we make the following assumptions.

Condition 2. (C1) The cumulative distribution function of the continuous response variable Y, F_Y, is continuously differentiable in a small neighborhood of y_τ, say $[y_\tau - \delta_0, y_\tau + \delta_0]$ with $\delta_0 > 0$. Let $G_1(\delta_0) = inf_{y\in[y_\tau-\delta_0, y_\tau+\delta_0]} f_Y(y)$ and $G_2(\delta_0) = sup_{y\in[y_\tau-\delta_0, y_\tau+\delta_0]} f_Y(y)$, where f_Y is the density function of Y. Assume that $0 < G_1(\delta_0) \leq G_2(\delta_0) < \infty$.

(C2) There is a $p \times (p - d_\tau)$ semi-orthogonal matrix $\mathbf{B}_{0,\tau}$ which minimizes

G_τ. Furthermore, the minimum value of G_τ is obtained at a semi-orthogonal matrix $\boldsymbol{\beta}_0$ if and only if $\mathcal{D}(\boldsymbol{\beta}_0, \mathbf{B}_{0,\tau}) = 0$.

The condition (C1) is used in Shao & Zhang (2014) and Zhang et al. (2018) for conditional quantile screening and testing. This assumption is on the marginal distribution of Y and is quite mild. The last condition (C2) ensures that $\mathbf{B}_{0,\tau}$ is the unique minimizer of G_τ in a sense of $\mathcal{D}$; see Lemma 1.1 in the supplementary material.

Theorem 1. *Assume that d_τ is known. Under Condition 1, Condition 2, and $\mathbf{X} \in \mathcal{L}^2$, we have $\mathcal{D}(\widehat{\mathbf{B}}_{0,\tau}, \mathbf{B}_{0,\tau}) \to^p 0$, as $n \to \infty$.*

Remark 1. It is worth mentioning the difference among our method and methods of Kong & Xia (2014), Christou (2019). In estimation procedure, Kong & Xia (2014) and Christou (2019) rely on local smoothing quantile regression whereas our approach uses QMDD which is an unconditional mean. Thus, the existing methods require user-chosen parameters such as a kernel function, bandwidth parameter, and the order of the polynomial regression, whereas the QMDD-based approach involves no user-chosen parameters. Moreover, when the structural dimension is one, the approach of Christou (2019) relies on the OLS estimate regressing the nonparametric estimate $\widehat{Q}_\tau(Y \mid \mathbf{X})$ on $\mathbf{X}$ which implicitly assumes that $Q_\tau(Y \mid \mathbf{X})$ is a

linear function of $\mathbf{B}_\tau^\top \mathbf{X}$, while our approach does not impose any model assumptions between $Q_\tau(Y \mid \mathbf{X})$ and $\mathbf{B}_\tau^\top \mathbf{X}$.

3.3 Dimension Selection

In practice, the dimension of the central quantile subspace is unknown and needs to be adaptively estimated from the data. Recently, Lee et al. (2019) introduced a wild bootstrap test for testing the mean independence of functional data using MDD-type test statistic. Since QMDD is an analogue of MDD, we follow the approach in Lee et al. (2019) and propose a new bootstrap test with QMDD to estimate the structural dimension d_τ. In particular, we sequentially test

$$H_0^{(k)} : d_\tau = k, \ k = 1, \cdots, p$$

using the wild bootstrap procedure described below.

1. Compute the test statistic, $T_n = n \cdot \widehat{\mathrm{QMDD}}_\tau(Y \mid \widehat{\mathbf{B}}_{0,\tau}^\top \mathbf{X})$, where $\widehat{\mathbf{B}}_{0,\tau}^\top$ is estimated by (3.8) and $dim(\widehat{\mathbf{B}}_{0,\tau}) = k$.

2. Generate the bootstrap statistic by

$$T_{n,b}^* = \left| \frac{-1}{n} \sum_{i,j} w_i^{(b)} \{\mathbf{1}(Y_i \le \widehat{y}_\tau) - \tau\} \{\mathbf{1}(Y_j \le \widehat{y}_\tau) - \tau\} |\widehat{\mathbf{B}}_{0,\tau}^{*\top} \mathbf{X}_i - \widehat{\mathbf{B}}_{0,\tau}^{*\top} \mathbf{X}_j| w_j^{(b)} \right|,$$

where $(w_i^{(b)})_{i=1}^n$ are iid with zero mean and unit variance, e.g., standard normal variables, and

$$\widehat{\mathbf{B}}_{0,\tau}^{*\top} = \underset{\boldsymbol{\beta}_0^\top \boldsymbol{\beta}_0 = I_{p-k}}{\operatorname{argmin}} \left(\frac{-1}{n^2} \sum_{i,j} w_i^{(b)} \{\mathbf{1}(Y_i \le \widehat{y}_\tau) - \tau\} \{\mathbf{1}(Y_j \le \widehat{y}_\tau) - \tau\} |\boldsymbol{\beta}_0^\top \mathbf{X}_i - \boldsymbol{\beta}_0^\top \mathbf{X}_j| w_j^{(b)} \right)^2.$$

3. Repeat step 2 for B times and collect $(T_{n,b}^*)_{b=1}^B$.
4. Obtain the $(1-\alpha)$-th quantile from the collected $(T_{n,b}^*)_{b=1}^B$, say $Q_{(1-\alpha)}^*$, and set it as the critical value for the test with significance level α.
5. Reject the null hypothesis if T_n is greater than the critical value $Q_{(1-\alpha)}^*$.

If k is smaller than the true dimension, d_τ, then we expect to reject $H_0^{(k)}$ and accept $H_0^{(k)}$ when k is identical to d_τ. The theory to show the consistency of the wild bootstrap test seems very challenging and is left for future research. The main difficulty of showing the consistency of the test arises from using the estimate $\widehat{\mathbf{B}}_{0,\tau}$ to compute T_n and involves the optimization step. Nevertheless, we applied the bootstrap test to determine the dimension d_τ in our simulation study and observed that the bootstrap test with $\alpha = 10\%$ works reasonably well; see Section 5.1.

4. Central Subspace

In this section, we seek the central subspace $\mathcal{S}_{Y|\mathbf{X}}$ and propose an estimation method using the so-called quantile martingale difference divergence matrix.

4.1 Quantile Martingale Difference Divergence Matrix

Most existing inverse regression methods assume the linearity condition, and the first moment methods hinge on the fact that $\Sigma^{-1}\big\{\mathbb{E}(\mathbf{X}|Y=y) - \mathbb{E}(\mathbf{X})\big\} \in \mathcal{S}_{Y|\mathbf{X}}$, for all y, where $\Sigma = \mathrm{var}(\mathbf{X})$. Then, the central subspace, or its subspace, can be obtained by estimating $\mathbb{E}(\mathbf{X}|Y=y) - \mathbb{E}(\mathbf{X})$. Often times, a nonparametric approach is used to estimate $\mathbb{E}(\mathbf{X}|Y=y)$, which involves a user chosen quantity, e.g., a number of slices. Instead, we focus on the observation that

$$\Sigma^{-1}\big\{\mathbb{E}(\mathbf{X}|\mathbf{1}(Y \le y_\tau) = y) - \mathbb{E}(\mathbf{X})\big\} \in \mathcal{S}_{Y|\mathbf{X}},\ \forall \tau \in (0,1),\ \forall y \in \{0,1\}. \tag{4.10}$$

Since (4.10) hinges on the mean dependence, we use MDDM in Lee & Shao (2018) to characterize this relationship and define a quantile dependence analogue of MDDM which involves no user-chosen quantities.

Definition 4. For a continuous random variable Y, a random vector $\mathbf{X} \in$

$\mathcal{L}^2$ and $\tau \in (0,1)$, the τ-th quantile martingale difference divergence matrix is defined as

$$\mathrm{QMDDM}_{\mathbf{X}|Y}(\tau) = -\mathbb{E}\left[\{\mathbf{X} - \mathbb{E}(\mathbf{X})\}\{\mathbf{X}' - \mathbb{E}(\mathbf{X}')\}^\top |\mathbf{1}(Y \le y_\tau) - \mathbf{1}(Y' \le y_\tau)|\right],$$

where $(\mathbf{X}', Y')$ is an independent copy of $(\mathbf{X}, Y)$.

Note that $\mathrm{QMDDM}_{\mathbf{X}|Y}(\tau)$ is a real, symmetric, and positive semi-definite matrix. Moreover, QMDDM inherits the same useful property as MDDM which is $\mathrm{span}\{\mathrm{QMDDM}_{\mathbf{X}|Y}(\tau)\} = \mathrm{span}[cov\{\mathbb{E}(\mathbf{X} \mid \mathbf{1}(Y \le y_\tau))\}]$; see Theorem 1 in Lee & Shao (2018). In the following, we show that $\mathrm{QMDDM}_{\mathbf{X}|Y}(\tau)$ can only identify one direction of $\mathcal{S}_{Y|\mathbf{X}}$. Suppose $\mathbb{E}(\mathbf{X}) = \mathbf{0}$. The τ-th quantile martingale difference divergence matrix is defined as

$$\begin{aligned} & -\mathbb{E}\left[\mathbf{X}\mathbf{X}'^\top |\mathbf{1}(Y \le y_\tau) - \mathbf{1}(Y' \le y_\tau)|\right] \\ = \; & -\mathbb{E}\left[\mathbf{X}\mathbf{X}'^\top \{\mathbf{1}(Y \le y_\tau)\mathbf{1}(Y' > y_\tau) + \mathbf{1}(Y > y_\tau)\mathbf{1}(Y' \le y_\tau)\}\right] \\ = \; & c \cdot m_\tau(\mathbf{X}) m_\tau(\mathbf{X})^\top, \end{aligned}$$

where c is a positive number and $m_\tau(\mathbf{X}) = \mathbb{E}\left[\mathbf{X}\mathbf{1}(Y \le y_\tau)\right] \in \mathbb{R}^p$. This implies that the rank of $\mathrm{QMDDM}_{\mathbf{X}|Y}(\tau)$ is one. Similar to the sample estimation of MDDM in Lee & Shao (2018), we define the sample estimator

of QMDDM as given in the definition below.

Definition 5. Given iid observations $(\mathbf{X}_i, Y_i)_{i=1}^n$ from the joint distribution of $(\mathbf{X}, Y)$, the sample τ-th quantile martingale difference divergence matrix is defined as $\widehat{\text{QMDDM}}_{\mathbf{X}|Y}(\tau) = -\frac{1}{n^2}\sum_{h,l=1}^n (\mathbf{X}_h - \overline{\mathbf{X}})(\mathbf{X}_l - \overline{\mathbf{X}})^\top |\mathbf{1}(Y_h \leq \widehat{y}_\tau) - \mathbf{1}(Y_l \leq \widehat{y}_\tau)|$, where $\overline{\mathbf{X}} = \frac{1}{n}\sum_{h=1}^n \mathbf{X}_h$.

4.2 Estimation of the Central Subspace

As we mentioned in Section 4.1, for a given τ, $\text{QMDDM}_{\mathbf{X}|Y}(\tau)$ can only provide one direction. In Theorem 2, we show that this direction indeed belongs to $\mathcal{S}_{Y|\mathbf{X}}$.

Theorem 2. *Assume the linearity condition that $\mathbb{E}(\mathbf{X} \mid \mathbf{B}^\top\mathbf{X})$ is a linear function of $\mathbf{B}^\top\mathbf{X}$, where $\mathbf{B}$ is a $p \times d$ basis matrix for $\mathcal{S}_{Y|\mathbf{X}}$. Then for a continuous random variable Y, a random vector $\mathbf{X} \in \mathcal{L}^2$, and any $\tau \in (0,1)$, we have $\Sigma^{-1}\text{span}\{\text{QMDDM}_{\mathbf{X}|Y}(\tau)\} \subseteq \mathcal{S}_{Y|\mathbf{X}}$.*

In order to gather information of $\mathcal{S}_{Y|\mathbf{X}}$ under different quantiles, we construct a new matrix by following the approaches in Kong & Xia (2014) and Christou (2019).

Definition 6. Let $\gamma_{1,\tau}$ be the eigenvector of $\text{QMDDM}_{\mathbf{X}|Y}(\tau)$ associated

with the largest eigenvalue. We define a new matrix by

$$\Gamma = \int_0^1 \Gamma(\tau) d\tau, \text{ where } \Gamma(\tau) = \gamma_{1,\tau}\gamma_{1,\tau}^\top.$$

Note that the matrix Γ is a real, symmetric, positive semidefinite matrix and encodes the information about the directions of $\mathcal{S}_{Y|\mathbf{X}}$. More precisely, the eigenvectors of Γ corresponding to the largest d eigenvalues of Γ belong to $\mathcal{S}_{Y|\mathbf{X}}$. The technique of aggregating the valid directions of the central subspace in Γ is quite common in SDR and provides successful finite sample performances; see Li (1991), Kong & Xia (2014), Christou (2019). As we have finite observations of $(\mathbf{X}_i, Y_i)_{i=1}^n$, we shall approximate Γ by $(\tau_i)_{i=1}^{n-1}$, $\tau_i = \frac{i}{n}$. In other words, we define

$$\widehat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n-1} \widehat{\Gamma}(\tau_i), \text{ where } \widehat{\Gamma}(\tau_i) = \widehat{\gamma}_{1,\tau_i}\widehat{\gamma}_{1,\tau_i}^\top,$$

and $\widehat{\gamma}_{1,\tau_i}$ is the eigenvector of $\widehat{\mathrm{QMDDM}}_{\mathbf{X}|Y}(\tau_i)$ associated with the largest eigenvalue.

Denote $\{\nu_j, \eta_j\}_{j=1}^p$ and $\{\widehat{\nu}_j, \widehat{\eta}_j\}_{j=1}^p$ as the eigenvalues and eigenvectors of Γ and $\widehat{\Gamma}$, respectively. We make the following assumptions, under which we establish the consistency of $\{\widehat{\nu}_j, \widehat{\eta}_j\}_{j=1}^d$.

Condition 3. (D1) The eigenvalues of Γ are given by $\nu_1 > \nu_2 > \cdots > \nu_d > 0 = \nu_{d+1} = \cdots = \nu_p$.

(D2) All elements of $\Gamma(\tau)$ are absolutely continuous on [0,1].

Theorem 3. *Under Condition 3 and* $\mathbf{X} \in \mathcal{L}^2$*, we have*

1. $\widehat{\nu}_j - \nu_j = O_p(n^{-1/2})$ *for* $j = 1, \cdots, d$,
2. $\|\widehat{\eta}_j - \eta_j\| = O_p(n^{-1/2})$ *for* $j = 1, \cdots, d$.

Theorem 3 suggests that the empirical eigenvalues and eigenvectors of $\widehat{\Gamma}$ are reasonable estimators of the population counterparts for large sample size. This theorem is proved for fixed p, and the theory for growing p is left for future research.

In practice, the structural dimension d of the central subspace is unkown. To estimate it, we adopt the BIC-type criterion proposed by Feng et al. (2013).

$$\widehat{d} = \underset{d \in \{1, \cdots, p\}}{\arg\max} \left(\frac{n \sum_{m=1}^{d} (\log(\widehat{\lambda}_m + 1) - \widehat{\lambda}_m)}{2 \sum_{m=1}^{p} (\log(\widehat{\lambda}_m + 1) - \widehat{\lambda}_m)} - 2C_n \times \frac{d(d+1)/2}{p} \right), \quad (4.11)$$

where C_n is a penalty constant, and $d(d+1)/2$ equals to the number of free parameters.

5. Numerical Simulations

In this section, we evaluate the finite sample performance of the proposed methods through simulations, and compare their performances with existing ones. In order to assess the estimation accuracy, we compute the trace correlation (Zhu et al. 2010), i.e., $R = \mathrm{tr}(\mathbf{P}_{\mathbf{B}}\mathbf{P}_{\widehat{\mathbf{B}}})/d$, where $\mathbf{P}_{\mathbf{B}}$ represents the projection matrix onto the column space of $\mathbf{B}$. Note that $0 \leq R \leq 1$ and $R = 1$ if $\mathrm{span}(\mathbf{B})$ is identical to $\mathrm{span}(\widehat{\mathbf{B}})$, and $R = 0$ if $\mathrm{span}(\mathbf{B})$ is perpendicular to $\mathrm{span}(\widehat{\mathbf{B}})$. Thus, larger values of R indicate more accurate estimates. For each example, we repeat the simulations 100 times and report results in the form of `mean(standard deviation)` of R. When assessing the estimation performance, we treat the structural dimension as known. However, we carry out separate simulation analysis to assess the performance of estimating the dimension using the bootstrap procedure in Section 3.3 or BIC-type criterion (Feng et al. 2013) in Section 4.2.

5.1 Central Quantile Subspace

In this section, we estimate the central quantile subspace. In particular, we compare with the methods of QOPG in Kong & Xia (2014) and SIQR/MIQR in Christou (2019). These two methods involve several user-chosen parameters. We follow the choices made in the code provided and

the suggestions in Li (2018), Kong & Xia (2014), Christou (2019). We tried several different bandwidth parameters, $h = c_h n^{-1/(p+4)}$ or $h = c_h n^{-1/(d_\tau+4)}$ depending on the predictor used for smoothing, $c_h = 0.7,\ 1.5,\ 2.34$, and use the Gaussian kernel, linear quantile regression. Throughout the simulations for central quantile subspace, we consider the sample sizes $n = 200,\ 400$, the dimensions of the predictor vector $p = 5,\ 10$, and the quantiles $\tau = 0.25,\ 0.5,\ 0.75$ unless otherwise specified. For the dimension selection, the bootstrap replicate $B = 400$, $\{w_i\}_{i=1}^n$ are from the standard normal distribution, and the significance level is $\alpha = 10\%$.

Example 1. This example is adopted from Christou (2019). The response variable Y is generated as $Y = 3x_1 + x_2 + \varepsilon$, where $\mathbf{X} = (x_1, \cdots, x_p)$ and ε are independently generated from standard normal distribution. For a given τ, $\mathbf{B}_\tau = (3, 1, 0, \cdots, 0)^\top/\sqrt{10}$.

Table 1 reports the trace correlation R for each method. All methods estimate the central quantile subspace accurately in terms of a higher R. Also, all three methods produce very comparable results. We observe that when p decreases or n increases, all methods improve. Note that $Q_\tau(Y \mid \mathbf{X})$ depends on $\mathbf{B}_\tau^\top \mathbf{X}$ in a linear fashion and so, the SIQR of Christou (2019) is expected to perform well since it uses OLS estimate regressing $\widehat{Q}_\tau(Y \mid \mathbf{X})$ on $\mathbf{X}$. It is interesting that the QMDD-based and QOPG-based approaches

can perform better than SIQR in some cases, e.g., when $p = 5$. It appears that the QOPG-based method performs slightly better than the QMDD and SIQR counterparts.

Table 1: Simulation results for the central $\tau-$th quantile subspace estimation. Reported results are `mean(standard deviation)` of the trace correlation from 100 replications.

			$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$	
Method		p	$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
QMDD		5	0.97 (0.02)	0.99 (0.01)	0.98 (0.01)	0.99 (0.01)	0.97 (0.02)	0.99 (0.01)
		10	0.95 (0.02)	0.98 (0.01)	0.96 (0.02)	0.98 (0.01)	0.95 (0.02)	0.98 (0.01)
QOPG	$c_h = 0.7$	5	0.98 (0.01)	0.99 (0.01)	0.98 (0.01)	0.99 (0.01)	0.98 (0.01)	0.99 (0.01)
		10	0.97 (0.01)	0.99 (0.01)	0.97 (0.01)	0.99 (0.01)	0.97 (0.01)	0.99 (0.01)
	$c_h = 1.5$	5	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.00)
		10	0.98 (0.01)	0.99 (0.01)	0.98 (0.01)	0.99 (0.01)	0.98 (0.01)	0.99 (0.01)
	$c_h = 2.34$	5	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.00)	0.99 (0.01)	1.00 (0.00)
		10	0.98 (0.01)	0.99 (0.00)	0.98 (0.01)	0.99 (0.00)	0.98 (0.01)	0.99 (0.01)
SIQR	$c_h = 0.7$	5	0.97 (0.03)	0.98 (0.04)	0.97 (0.02)	0.98 (0.04)	0.97 (0.03)	0.98 (0.04)
		10	0.97 (0.02)	0.98 (0.01)	0.97 (0.02)	0.98 (0.01)	0.97 (0.02)	0.98 (0.01)
	$c_h = 1.5$	5	0.97 (0.03)	0.98 (0.04)	0.97 (0.03)	0.98 (0.04)	0.97 (0.03)	0.98 (0.04)
		10	0.97 (0.02)	0.98 (0.01)	0.97 (0.02)	0.98 (0.01)	0.97 (0.02)	0.98 (0.01)
	$c_h = 2.34$	5	0.97 (0.03)	0.98 (0.04)	0.97 (0.03)	0.98 (0.04)	0.97 (0.03)	0.98 (0.04)
		10	0.96 (0.02)	0.98 (0.01)	0.96 (0.02)	0.98 (0.01)	0.97 (0.02)	0.98 (0.01)

Example 2. This example is adopted from Kong & Xia (2012) with slight modification which satisfies Condition 1. The data is generated by $Y = \frac{\exp\left(3\sqrt{2}x_1+3\sqrt{2}x_5-6+6x_3\varepsilon\right)}{1+\exp\left(3\sqrt{2}x_1+3\sqrt{2}x_5-6+6x_3\varepsilon\right)}$, where $\mathbf{X} = (x_1, \cdots, x_p)$ are independently generated from $U(0, 1)$, ε is from $U(-1, 1)$. For a given τ, $\mathbf{B}_\tau = (\sqrt{2}, 0, 2(2\tau - 1), 0, \sqrt{2}, 0, \cdots, 0)^\top / \sqrt{4 + 4(2\tau - 1)^2}$.

From Table 2, we can see that QMDD-based and QOPG-based methods outperform the SIQR method. Notice that the model has a strong nonlinear

quantile dependence between Y and $\mathbf{B}_\tau^\top \mathbf{X}$. We speculate that this could be a part of the reason for the inferior performance of the SIQR approach. It also shows that the QOPG-based approach performs better than the QMDD-based counterpart for some cases depending on (τ, h). Furthermore, the performance of the existing methods substantially changes based on different values of c_h, showing its sensitivity with respect to the choice of h. It is also worth noting that there seems no uniformly best choice of h which gives the best performance for all cases. In other words, for different combinations of (τ, n, p), different values of h yield the best performance.

Table 2: Simulation results for the central $\tau-$th quantile subspace estimation. Reported results are `mean(standard deviation)` of the trace correlation from 100 replications.

			$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$	
Method		p	$n = 200$	n =400	$n = 200$	n =400	$n = 200$	n =400
QMDD		5	0.89 (0.06)	0.93 (0.04)	0.95 (0.04)	0.98 (0.02)	0.91 (0.06)	0.94 (0.04)
		10	0.83 (0.08)	0.90 (0.05)	0.90 (0.05)	0.95 (0.02)	0.84 (0.07)	0.91 (0.04)
QOPG	$c_h = 0.7$	5	0.66 (0.18)	0.75 (0.12)	0.83 (0.15)	0.89 (0.10)	0.81 (0.17)	0.88 (0.10)
		10	0.56 (0.11)	0.63 (0.09)	0.83 (0.09)	0.91 (0.05)	0.80 (0.09)	0.86 (0.06)
	$c_h = 1.5$	5	0.93 (0.04)	0.97 (0.03)	0.95 (0.03)	0.97 (0.02)	0.97 (0.02)	0.99 (0.01)
		10	0.79 (0.09)	0.87 (0.05)	0.89 (0.05)	0.94 (0.03)	0.91 (0.04)	0.95 (0.02)
	$c_h = 2.34$	5	0.94 (0.04)	0.97 (0.02)	0.95 (0.03)	0.97 (0.02)	0.97 (0.02)	0.99 (0.01)
		10	0.88 (0.06)	0.93 (0.04)	0.91 (0.04)	0.95 (0.02)	0.94 (0.03)	0.97 (0.02)
SIQR	$c_h = 0.7$	5	0.44 (0.25)	0.42 (0.22)	0.52 (0.25)	0.53 (0.21)	0.45 (0.28)	0.47 (0.25)
		10	0.41 (0.19)	0.45 (0.17)	0.56 (0.20)	0.64 (0.20)	0.51 (0.21)	0.59 (0.21)
	$c_h = 1.5$	5	0.36 (0.25)	0.37 (0.22)	0.43 (0.27)	0.47 (0.24)	0.38 (0.30)	0.43 (0.27)
		10	0.36 (0.19)	0.41 (0.19)	0.50 (0.23)	0.60 (0.23)	0.47 (0.24)	0.58 (0.24)
	$c_h = 2.34$	5	0.34 (0.25)	0.35 (0.22)	0.40 (0.28)	0.44 (0.25)	0.36 (0.30)	0.42 (0.27)
		10	0.35 (0.19)	0.40 (0.19)	0.49 (0.24)	0.60 (0.24)	0.47 (0.25)	0.58 (0.24)

Example 3. This example is taken from Luo et al. (2014) and it ad-

dresses the correlated $\mathbf{X}$ case. More specifically the data is generated by $Y = 1 + x_1 + (1 + 0.4x_2)\varepsilon$, where $\mathbf{X} = (x_1, \cdots, x_p)$ is from $N(0, \Sigma)$, $\Sigma = [\Sigma_{ij}]_{i,j=1}^p$, $\Sigma_{ij} = 0.5^{|i-j|}$ and ε is from standard normal distribution. This is a heteroscedastic model, where $\mathbf{B}_\tau = (1, 0.4\Phi^{-1}(\tau), 0, \cdots, 0)^\top / \sqrt{1 + (0.4\Phi^{-1}(\tau))^2}$, Φ is the c.d.f. of the standard normal distribution.

The trace correlation results are reported in Table 3. The QMDD-based and the QOPG-based methods are superior to the SIQR approach in terms of a higher R in all cases. For $\tau = 0.25,\ 0.5$, we observe that our method is comparable or outperforms the QOPG method when $c_h = 0.7,\ 1.5$. Note that $\mathbf{X}$ is correlated which indicates that Condition 1 is not valid in this example. It shows that our QMDD-based approach could still work for correlated $\mathbf{B}_\tau^\top \mathbf{X}$ and $\mathbf{B}_{0,\tau}^\top \mathbf{X}$ to some extent.

Table 3: Simulation results for the central $\tau-$th quantile subspace estimation. Reported results are `mean(standard deviation)` of the trace correlation from 100 replications.

			$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$	
Method		p	$n = 200$	$n = 400$	$n = 200$	$n = 400$		
QMDD		5	0.93 (0.06)	0.95 (0.04)	0.93 (0.06)	0.97 (0.03)	0.88 (0.08)	0.95 (0.04)
		10	0.86 (0.08)	0.92 (0.04)	0.86 (0.07)	0.93 (0.04)	0.77 (0.11)	0.88 (0.05)
QOPG	$c_h = 0.7$	5	0.70 (0.25)	0.83 (0.14)	0.75 (0.21)	0.86 (0.12)	0.75 (0.17)	0.86 (0.10)
		10	0.71 (0.15)	0.82 (0.09)	0.77 (0.11)	0.89 (0.07)	0.74 (0.10)	0.84 (0.07)
	$c_h = 1.5$	5	0.93 (0.05)	0.97 (0.03)	0.94 (0.05)	0.97 (0.02)	0.94 (0.04)	0.97 (0.02)
		10	0.83 (0.11)	0.91 (0.05)	0.86 (0.08)	0.93 (0.04)	0.85 (0.07)	0.92 (0.04)
	$c_h = 2.34$	5	0.95 (0.05)	0.97 (0.02)	0.95 (0.04)	0.98 (0.02)	0.94 (0.04)	0.97 (0.02)
		10	0.89 (0.07)	0.94 (0.04)	0.90 (0.05)	0.95 (0.03)	0.89 (0.05)	0.94 (0.03)
SIQR	$c_h = 0.7$	5	0.46 (0.28)	0.53 (0.29)	0.46 (0.28)	0.54 (0.30)	0.32 (0.23)	0.39 (0.23)
		10	0.38 (0.24)	0.56 (0.24)	0.46 (0.24)	0.65 (0.22)	0.30 (0.17)	0.44 (0.19)
	$c_h = 1.5$	5	0.42 (0.28)	0.52 (0.30)	0.42 (0.29)	0.54 (0.30)	0.30 (0.23)	0.37 (0.24)
		10	0.36 (0.23)	0.55 (0.25)	0.44 (0.25)	0.64 (0.22)	0.28 (0.17)	0.43 (0.19)
	$c_h = 2.34$	5	0.42 (0.29)	0.51 (0.30)	0.41 (0.30)	0.54 (0.30)	0.29 (0.23)	0.36 (0.24)
		10	0.35 (0.24)	0.55 (0.25)	0.43 (0.25)	0.64 (0.22)	0.27 (0.18)	0.43 (0.19)

Example 4. This example considers an inverse model. In particular, we generate the data by $\mathbf{X} = \beta_1 log(Y^2 + 1.5) + \beta_2(sign(Y)) + \varepsilon$, where Y is generated from $U(-3, 3)$, ε is from Beta(1,2), and $\beta_1 = b_1$, $\beta_2 = b_2$ or $\beta_1 = (b_1^\top, b_1^\top)^\top/\sqrt{2}$, $\beta_2 = (b_2^\top, b_2^\top)^\top/\sqrt{2}$, where $b_1 = (2, 0, -1, 0, 2)^\top/3$, $b_2 = (0, 1, 0, 1, 0)^\top/\sqrt{2}$. There are two directions and $\mathbf{B}_\tau = (\beta_1, \beta_2)$ for $\tau = 0.25,\ 0.75$.

Table 4 summarizes the performance of our method, the QOPG and MIQR approaches. The QMDD-based approach is superior to the method of QOPG and MIQR. As the data is generated from the inverse model, this example seems to be more complicated than the previous examples. Due

to this fact, we presume that this could have affected the nonparametric modeling step in QOPG and MIQR, and lead to some loss of accuracy.

Table 4: Simulation results for the central $\tau-$th quantile subspace estimation. Reported results are `mean(standard deviation)` of the trace correlation from 100 replications.

Method		p	$\tau = 0.25$		$\tau = 0.75$	
			$n = 200$	$n = 400$	$n = 200$	$n = 400$
QMDD		5	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)
		10	0.97 (0.02)	0.98 (0.01)	0.97 (0.02)	0.98 (0.01)
QOPG	$c_h = 0.7$	5	0.57 (0.09)	0.53 (0.05)	0.58 (0.10)	0.53 (0.05)
		10	0.51 (0.06)	0.53 (0.07)	0.52 (0.08)	0.54 (0.08)
	$c_h = 1.5$	5	0.94 (0.04)	0.96 (0.02)	0.94 (0.03)	0.97 (0.02)
		10	0.87 (0.04)	0.93 (0.02)	0.87 (0.04)	0.93 (0.02)
	$c_h = 2.34$	5	0.94 (0.04)	0.97 (0.02)	0.94 (0.03)	0.97 (0.02)
		10	0.82 (0.05)	0.91 (0.03)	0.82 (0.06)	0.91 (0.03)
MIQR	$c_h = 0.7$	5	0.54 (0.07)	0.55 (0.07)	0.54 (0.07)	0.54 (0.06)
		10	0.49 (0.03)	0.51 (0.03)	0.49 (0.03)	0.51 (0.03)
	$c_h = 1.5$	5	0.55 (0.08)	0.54 (0.07)	0.54 (0.06)	0.54 (0.07)
		10	0.49 (0.03)	0.51 (0.02)	0.49 (0.02)	0.51 (0.03)
	$c_h = 2.34$	5	0.53 (0.06)	0.54 (0.06)	0.53 (0.06)	0.52 (0.04)
		10	0.49 (0.03)	0.51 (0.02)	0.49 (0.02)	0.51 (0.02)

Example 5. In this example, we generate the response variable Y by $Y = \sqrt{x_1 + 1} + \sqrt{x_2 + 1} + \varepsilon$, where $\mathbf{X} = (x_1, \cdots, x_p)$ is generated by $\chi^2(2)$, and ε is from Beta(1,2). Here $\mathbf{B}_\tau = (\beta_1, \beta_2)$, where $\beta_1 = (1, 0, \cdots, 0)^\top$ and $\beta_2 = (0, 1, 0, \cdots, 0)^\top$.

From Table 5, it appears that our approach outperforms the existing methods in all cases. Overall, our simulation evidence seems to suggest that the QMDD-based approach can perform better than the existing methods for both forward and inverse models with quite stable performance while

the performance of existing ones show sensitivity with respect to the choice of h. It is worth mentioning that our approach does not involve any user-chosen quantities and is simpler to implement.

Table 5: Simulation results for the central $\tau-$th quantile subspace estimation. Reported results are `mean(standard deviation)` of the trace correlation from 100 replications.

			$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$	
Method		p	$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
QMDD		5	0.90 (0.07)	0.94 (0.05)	0.92 (0.06)	0.96 (0.04)	0.93 (0.06)	0.96 (0.03)
		10	0.80 (0.09)	0.86 (0.07)	0.86 (0.08)	0.91 (0.05)	0.87 (0.06)	0.92 (0.05)
QOPG	$c_h = 0.7$	5	0.61 (0.13)	0.62 (0.12)	0.60 (0.13)	0.62 (0.11)	0.61 (0.13)	0.63 (0.14)
		10	0.54 (0.06)	0.54 (0.07)	0.55 (0.07)	0.55 (0.07)	0.55 (0.07)	0.56 (0.08)
	$c_h = 1.5$	5	0.65 (0.16)	0.72 (0.15)	0.68 (0.16)	0.76 (0.15)	0.70 (0.16)	0.77 (0.16)
		10	0.53 (0.04)	0.55 (0.06)	0.56 (0.08)	0.61 (0.11)	0.60 (0.10)	0.67 (0.13)
	$c_h = 2.34$	5	0.77 (0.17)	0.89 (0.10)	0.75 (0.17)	0.86 (0.12)	0.74 (0.15)	0.86 (0.12)
		10	0.58 (0.10)	0.69 (0.13)	0.65 (0.11)	0.75 (0.12)	0.67 (0.13)	0.77 (0.12)
MIQR	$c_h = 0.7$	5	0.64 (0.14)	0.64 (0.15)	0.64 (0.14)	0.66 (0.15)	0.63 (0.14)	0.64 (0.15)
		10	0.54 (0.06)	0.54 (0.07)	0.54 (0.06)	0.54 (0.06)	0.54 (0.07)	0.54 (0.06)
	$c_h = 1.5$	5	0.66 (0.14)	0.65 (0.13)	0.65 (0.14)	0.65 (0.14)	0.66 (0.15)	0.65 (0.15)
		10	0.54 (0.07)	0.54 (0.07)	0.55 (0.07)	0.55 (0.08)	0.55 (0.08)	0.55 (0.07)
	$c_h = 2.34$	5	0.67 (0.14)	0.65 (0.13)	0.67 (0.15)	0.66 (0.14)	0.67 (0.15)	0.65 (0.15)
		10	0.55 (0.07)	0.55 (0.07)	0.55 (0.08)	0.56 (0.08)	0.56 (0.09)	0.55 (0.07)

Lastly, we apply the bootstrap test described in Section 3.3 to select the dimension of the central quantile subspace. Table 6 reports the percentage of correctly identifying the structural dimension of the central quantile subspace under the previous simulation models when $n = 400$ and $p = 5$. The bootstrap test for dimension selection has reasonable results for all models.

Table 6: Percentages of correctly selected dimension, under-selection, and over-selection over 100 replicates for each example.

Model	τ	$\widehat{d}_\tau < d_\tau$	$\widehat{d}_\tau = d_\tau$	$\widehat{d}_\tau > d_\tau$
Example 1	0.25	0	100	0
	0.50	0	100	0
	0.75	0	100	0
Example 2	0.25	0	99	1
	0.50	0	100	0
	0.75	0	100	0
Example 3	0.25	0	100	0
	0.50	0	100	0
	0.75	0	100	0
Example 4	0.25	0	100	0
	0.75	0	100	0
Example 5	0.25	1	90	9
	0.50	0	88	12
	0.75	0	99	1

5.2 Central Subspace

In this section, we estimate the central subspace. We compare our method with the existing inverse regression methods including the sliced inverse regression (SIR; Li 1991), the directional regression (DR; Li & Wang 2007), and the cumulative slicing (CUME; Zhu et al. 2010). For SIR and DR, the number of slices is 5. We consider the sample size $n = 200,\ 400$, and the dimension of the predictor $p = 10,\ 20$. When we estimate the structural dimension using the BIC-type criterion, we use $C_n = n^{1/3}p^{2/3}$ following the recommendation in Feng et al. (2013).

Example 6. This example considers a single index model. More specifically, the response Y is generated by $Y = (\beta_1^\top \mathbf{X}+1)^3+\varepsilon$, where $\mathbf{X}$ and ε are generated from $U(0,5)$ and $U(-1,1)$, respectively. The central subspace is spanned by $\mathbf{B} = \beta_1$, where $\beta_1 = (1,1,1,0,0,\cdots,0)^\top$.

Table 7 summarizes the performance of all methods under different models. For this example, it appears that all approaches perform reasonably well, where our QMDDM-based method performs slightly better than the existing ones. Overall, when n increases and p decreases, all methods produce better estimates of $\mathbf{P_B}$ as R increases.

Example 7. In this example, we consider a heteroscedastic model. In particular, we generate the data by the following model. $Y = \exp(\beta_1^\top \mathbf{X}+1+\beta_2^\top \mathbf{X}\varepsilon)$, where $\mathbf{X}$ and ε are defined in Example 6. The structural dimension is equal to two and $\mathbf{B} = (\beta_1, \beta_2)$, where $\beta_1 = (1,1,1,0,0,\cdots,0)^\top$ and $\beta_2 = (0,0,0,1,1,1,0,\cdots,0)^\top$.

From Table 7, we observe that our QMDDM-based approach is superior to the existing methods in terms of a higher R. It shows that CUME and SIR are comparable with the CUME slightly performing better than the SIR and outperforming the DR.

Example 8. In this example, we examine a model with correlated $\mathbf{X}$.

We generate the response by $Y = \frac{\beta_1^\top \mathbf{X}}{(\beta_2^\top \mathbf{X}+1.5)^2+0.5} + 0.5\beta_1^\top \mathbf{X}\varepsilon$, where $\mathbf{X}$ is generated from $N(\mathbf{0}, \Sigma)$, where $\Sigma = [\sigma_{ij}]_{i,j=1}^p$, $\sigma_{ij} = 0.5^{|i-j|}$, and ε is generated from $N(0,1)$. Here, $\mathbf{B} = (\beta_1, \beta_2)$, where $\beta_1 = (1, 0, \cdots, 0)^\top$, $\beta_2 = (0, 1, 0 \cdots, 0)^\top$.

Table 7 also presents the means and standard errors of all approaches for this example. It suggests that the QMDDM-based method generates higher R values than the other ones which indicates that the QMDD-based approach outperforms the other methods in all cases.

Example 9. In this example, we consider an inverse model introduced in Example 4. In particular, we generate $\mathbf{X} = \beta_1 log(Y^2+1.5)+\beta_2(sign(Y))+\varepsilon$, where Y, ε, and $\mathbf{B} = \mathbf{B}_\tau$ are defined in Example 4.

According to Table 7, it seems that all methods produce accurate results since R values are high enough. More precisely, the CUME and the QMDDM-based methods are very comparable and outperform the other existing methods. To summarize, the simulation results clearly demonstrate the usefulness of the proposed approach which is easy to implement and includes no user-chosen parameter.

Furthermore, Table 8 reports the percentage of correctly identifying the structural dimension of the central subspace under the previous simulation

models when $n = 400$ and $p = 10$. We apply BIC-type criterion in Section 4.2. We observe that BIC-type criterion works fairly well under all models.

Table 7: Simulation results for the central subspace estimation. Reported results are `mean(standard deviation)` of the trace correlation from 100 replications.

Model	p	n	CUME	SIR	DR	QMDDM
Example 6	10	200	0.98 (0.01)	0.99 (0.00)	0.99 (0.00)	1.00 (0.00)
		400	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	20	200	0.97 (0.01)	0.99 (0.00)	0.98 (0.01)	1.00 (0.00)
		400	0.98 (0.01)	1.00 (0.00)	0.99 (0.00)	1.00 (0.00)
Example 7	10	200	0.79 (0.07)	0.75 (0.10)	0.51 (0.14)	0.83 (0.06)
		400	0.88 (0.04)	0.85 (0.06)	0.65 (0.15)	0.91 (0.03)
	20	200	0.63 (0.07)	0.57 (0.10)	0.26 (0.12)	0.66 (0.07)
		400	0.78 (0.05)	0.73 (0.07)	0.38 (0.12)	0.79 (0.05)
Example 8	10	200	0.79 (0.07)	0.74 (0.11)	0.65 (0.13)	0.88 (0.08)
		400	0.86 (0.06)	0.86 (0.09)	0.80 (0.13)	0.94 (0.05)
	20	200	0.63 (0.09)	0.52 (0.11)	0.40 (0.11)	0.75 (0.10)
		400	0.76 (0.06)	0.72 (0.08)	0.63 (0.12)	0.87 (0.05)
Example 9	10	200	0.99 (0.00)	0.86 (0.05)	0.83 (0.05)	0.99 (0.00)
		400	0.99 (0.00)	0.92 (0.02)	0.91 (0.03)	1.00 (0.00)
	20	200	0.98 (0.01)	0.72 (0.05)	0.66 (0.05)	0.99 (0.00)
		400	0.99 (0.00)	0.84 (0.03)	0.81 (0.04)	0.99 (0.00)

Table 8: Percentages of correctly selected dimension, under-selection, and over-selection over 100 replicates for each example.

Model	$\widehat{d} < d$	$\widehat{d} = d$	$\widehat{d} > d$
Example 6	0	100	0
Example 7	0	100	0
Example 8	14	86	0
Example 9	0	100	0

6. Real Data Illustration

In this section, we focus on the central quantile subspace and consider the riboflavin data which has been analyzed by Buhlmann et al. (2014) and Zhang et al. (2019). This data contains 71 samples of the riboflavin production rate and the expression level of 4,088 genes. The response is the logarithm of the riboflavin production rate and the predictors are the logarithm of the expression level of genes. Due to the high dimensionality and relatively small sample size, we apply variable screening to the predictors similar to Buhlmann et al. (2014) and Zhang et al. (2019). In particular, we select g strongly related genes to the riboflavin production rate by using the quantile dependence analogue of the martingale difference correlation in Shao & Zhang (2014). After applying the bootstrap test in Section 3.3, we determine $d_\tau = 1$ for $\tau = 0.25,\ 0.5,\ 0.75$ with the selected genes. We apply the proposed method and the existing methods with the same user-chosen quantities in Section 5.1. Figure 1 reports the estimated direction using the whole data with $g = 5$ and the estimated direction with $g = 10$ is very similar to the ones in the figure; see the supplementary material. It shows some curvatures for $\tau = 0.25,\ 0.5,\ 0.75$. To evaluate the estimation stability of the central quantile subspace, we consider the boostrap variability $B^{-1}\sum_{b=1}^{B}\|\mathbf{P}_{\widehat{\mathbf{B}}_\tau} - \mathbf{P}_{\widehat{\mathbf{B}}_\tau^b}\|_F$, where $\widehat{\mathbf{B}}_\tau$ is the estimated semi-orthogonal matrix

on the whole data and $\widehat{\mathbf{B}}_\tau^b$, $b = 1, \cdots, B$, is the estimated semi-orthogonal matrix on 71 bootstrap samples from $B = 100$ bootstrap replicates. The boostrap variability is summarized in Table 9 which includes the best results of the existing methods among different bandwidth parameters. It shows that the proposed method generates more stable estimation of the central quantile subspace than the existing methods. In terms of the computational time, the QMDD-based approach requires comparable or less computational time than the existing counterparts for this data set. In particular, when $g = 5$ and $c_h = 0.7$, and $\tau = 0.25$, the QMDD, the QOPG, and the SIQR methods take 4.64, 14.54, and 5.17 (seconds) for $B = 100$ bootstrap replicates. The computation has been done by Windows 10 computer with Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz processor, 32.0 GB installed memory (RAM), 64-bit Operating System.

	QMDD			QOPG			SIQR		
	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
$g = 5$	0.20 (0.02)	0.16 (0.01)	0.15 (0.01)	0.45 (0.02)	0.72 (0.03)	0.55 (0.03)	0.50 (0.03)	0.85 (0.03)	0.64 (0.02)
$g = 10$	0.28 (0.02)	0.20 (0.01)	0.17 (0.01)	0.69 (0.02)	0.89 (0.03)	0.87 (0.03)	0.72 (0.02)	0.85 (0.03)	0.93 (0.03)

Table 9: Comparison of methods for the central quantile subspace in estimation stability.

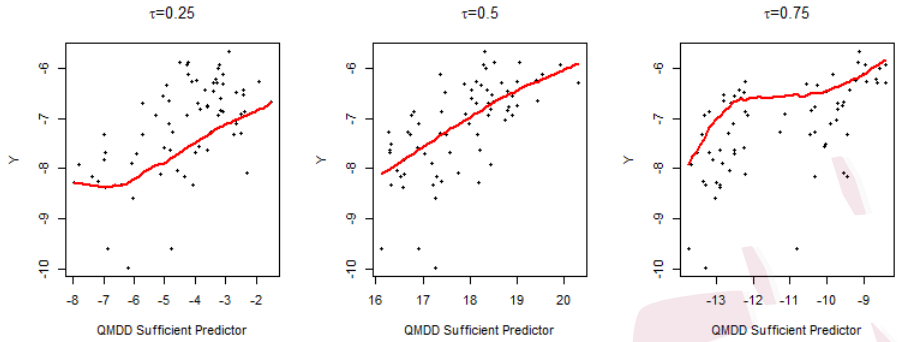

Figure 1: Sufficient summary plots of the central τth-quantile subspace direction for QMDD approach with $g = 5$. The solid lines refer to the local quantile regressions for each quantile.

7. Conclusion

In this paper, we use two metrics, the quantile martingale difference divergence and quantile martingale difference divergence matrix to estimate the central quantile subspace and the central subspace, respectively. We also introduce a new bootstrap test to select the structural dimension for central quantile subspace and use BIC-type criterion to choose the dimension of central subspace. Finite sample performances and real data application suggest that our QMDD(M)-based approach performs comparably well and can produce more accurate results with comparable or less computational time. In contrast to the existing methods for the central quantile subspace or the central subspace, the QMDD(M)-based approach includes no user-chosen numbers so the QMDD(M)-based approach is more convenient

and simple to implement. Theoretical results are obtained under suitable conditions and they justify the validity of our methods.

We shall conclude by mentioning several future research topics. The bootstrap test to select the structural dimension is worth investigating. A rigorous theoretical study is needed. It would be important to understand the behavior of the proposed approaches when the dimension p is large from theoretical aspects and examine if we can extend the method to large p case however it seems to be very challenging. Another issue is that we assume Y is a univariate throughout the article. It would be useful to extend our methods to a multivariate Y which seems to be nontrivial. Lastly, it would be interesting to extend the idea to estimate the envelope quantile regression in Ding et al. (2019) considering the connection between SDR and the predictor envelope model. The research along these directions are well underway.

References

Buhlmann, P., Kalisch, M. & Meier, L. (2014), 'High-dimensional statistics with a view toward applications in biology,', *Annual Review of Statistics and Its Application* **1**, 255–78.

Bura, E. & Cook, R. D. (2001), 'Extending sliced inverse regression: The weighted chi-squared test', *Journal of the American Statistical Association* **96**, 996–1003.

Christou, E. (2019), 'Central quantile subspace', *Statistics and Computing* **to appear**.

Cook, R. (1998), *Regression Graphics: Ideas for studying regressions through graphics*, Wiley: New York.

Cook, R. & Li, B. (2002), 'Dimension reduction for conditional mean in regression', *The Annals of Statistics* **30**, 455–474.

Cook, R. & Weisberg, S. (1991), 'Discussion of "sliced inverse regression for dimension reduction"', *Journal of the American Statistical Association* **86**, 328–332.

Ding, S., Su, Z., Zhu, G. & Wang, L. (2019), 'Envelope quantile regression', *Statistica Sinica* **to appear**.

Feng, Z., Wen, X., Yu, Z. & Zhu, L. (2013), 'On partial sufficient dimension reduction with applications to partially linear multi-index models', *Journal of the American Statistical Association* **108**, 237–246.

Kong, E. & Xia, Y. (2012), 'A single-index quantile regression model and its estimation', *Econometric Theory* **28**, 730–768.

Kong, E. & Xia, Y. (2014), 'An adaptive composite quantile approach to dimension reduction', *The Annals of Statistics* **42**, 1657–1688.

Lee, C. E. & Shao, X. (2018), 'Martingale difference divergence matrix and its application to dimension reduction for stationary multivariate time series', *Journal of the American Statistical Association* **113**, 216–229.

Lee, C. E., Zhang, X. & Shao, X. (2019), 'Testing the conditional mean independence for functional data', *Biometrika* **to appear**.

Li, B. (2018), *Sufficient Dimension Reduction: Methods and Applications with R*, Monographs on Statistics and Applied Probability, Chapman and Hall/CRC.

Li, B. & Wang, S. (2007), 'On directional regression for dimension reduction', *Journal of the American Statistical Association* **102**, 997–1008.

Li, B., Zha, H. & Chiaromonte, F. (2005), 'Contour regression: A general approach to dimension reduction', *The Annals of Statistics* **33**, 1580–1616.

Li, G., Li, Y. & Tsai, C.-L. (2015), 'Quantile correlations and quantile autoregressive modeling', *Journal of the American Statistical Association* **110**, 246–261.

Li, K.-C. (1991), 'Sliced inverse regression for dimension reduction', *Journal of the American Statistical Association* **86**, 316–342.

Luo, W., Li, B. & Yin, X. (2014), 'On efficient dimension reduction with respect to a statistical functional of interest', *The Annals of Statistics* **42**, 382–412.

Ma, Y. & Zhu, L. (2013), 'A review on dimension reduction.', *International Statistical Review* **81**, 134–150.

Park, T., Shao, S. & Yao, S. (2015), 'Partial martingale difference correlation', *Electronic Journal of Statistics* **9**, 3392–3416.

Shao, X. & Zhang, J. (2014), 'Martingale difference correlation and its use in high-dimensional

variable screening', *Journal of the American Statistical Association* **109**, 1302–1318.

Sheng, W. & Yin, X. (2013), 'Direction estimation in single-index models via distance covariance', *Journal of Multivariate Analysis* **122**, 148–161.

Sheng, W. & Yin, X. (2016), 'Sufficient dimension reduction via distance covariance', *Journal of Computational and Graphical Statistics* **25**, 91–104.

Székely, G. & Rizzo, M. (2014), 'Partial distance correlation with methods for dissimilarities', *The Annals of Statistics* **42**, 2382–2412.

Wen, Z. & Yin, W. (2013), 'A feasible method for optimization with orthogonality constraints', *Mathematical Programming* **142**, 397–434.

Yin, X. & Cook, R. D. (2002), 'Dimension reduction for the conditional kth moment in regression', *Journal of the Royal Statistical Society. Series B* **64**, 159–175.

Yin, X. & Li, B. (2011), 'Sufficient dimension reduction based on an ensemble of minimum average variance estimators', *Annals of Statistics* **39**, 3392–3416.

Yin, X., Li, B. & Cook, R. D. (2008), 'Successive direction extraction for estimating the central subspace in a multiple-index regression', *Journal of Multivariate Analysis* **99**, 1733 – 1757.

Zhang, X., Lee, C. E. & Shao, X. (2019), 'Envelopes in multivariate regression models with nonlinearity and heteroscedasticity', *Biometrika* **to appear**.

Zhang, X., Yao, S. & Shao, X. (2018), 'Conditional mean and quantile dependence testing in high dimension', *The Annals of Statistics* **46**, 219–246.

Zhu, L.-P. & Zhu, L.-X. (2009), 'Dimension reduction for conditional variance in regression', *Statistica Sinica* **19**, 869–883.

Zhu, L.-P., Zhu, L.-X. & Feng, Z.-H. (2010), 'Dimension reduction in regressions through cumulative slicing estimation', *Journal of the American Statistical Association* **105**, 1455–1466.

Zhu, Y. & Zeng, P. (2006), 'Fourier methods for estimating the central subspace and the central mean subspace in regression', *Journal of the American Statistical Association* **101**, 1638–1651.