

QUANTILE MARTINGALE DIFFERENCE DIVERGENCE FOR DIMENSION REDUCTION

Chung Eun Lee and Haileab Hilafu

Baruch College and University of Tennessee

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

Key words and phrases: Central subspace, dimension reduction, quantile dependence.

1. Introduction

Sufficient dimension reduction (SDR) (Li (1991); Cook (1998)) combines dimension reduction and 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 $\boldsymbol{\beta} \in \mathbb{R}^{p \times d'}$, for $d' \leq p$, is said to be a SDR subspace if $Y \perp\!\!\!\perp \mathbf{X} \mid \boldsymbol{\beta}^\top \mathbf{X}$, where $\perp\!\!\!\perp$ denotes independence. Furthermore, the minimal SDR subspace is called the central subspace, and is formally defined as

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

where $\mathbf{B} \in \mathbb{R}^{p \times d}$. The body of literature on SDR is growing; see the excellent reviews by Cook (1998), Li (2018), and Ma and Zhu (2013). Most methods have targeted the central subspace. These include, the sliced inverse regression (Li (1991)), sliced average variance estimation (Cook and Weisberg (1991)), parametric inverse regression (Bura and Cook (2001)), contour regression

Corresponding author: Chung Eun Lee, Paul H. Chook Department of Information Systems and Statistics, Baruch College, The City University of New York, New York, NY 10010, USA. E-mail: chungleun.lee@baruch.cuny.edu.

(Li, Zha and Chiaromonte (2005)), directional regression (Li and Wang (2007)), cumulative slicing estimation (Zhu, Zhu and Feng (2010)), Fourier method (Zhu and Zeng (2006)), ensemble of minimum average variance estimators method (Yin and Li (2011)), and SDR method via distance covariance (Sheng and Yin (2016)), among others.

In many applications, the focus of a regression analysis is a particular characteristic of Y given \mathbf{X} , rather than the entire conditional distribution of Y given \mathbf{X} . Such methods include the central mean subspace (Cook and Li (2002)), central k th moment subspace (Yin and Cook (2002)), central variance subspace (Zhu and Zhu (2009)), and T -central subspace (Luo, Li and Yin (2014)). Kong and Xia (2012) and Kong and Xia (2014) proposed the quantile outer-product of gradients (QOPG) method, which can be used to estimate the central quantile subspace (Luo, Li and Yin (2014)); see (3.1) for the definition of the central quantile subspace. Recently, Christou (2020) proposed an efficient algorithm for finding the central quantile subspace, and generalized the approach by considering any statistical functional of interest. However, the approach implicitly assumes a linear model for the conditional quantile of Y given \mathbf{X} , when the structural dimension of the central quantile subspace d_τ is one. In addition, both procedures in Kong and Xia (2014) and Christou (2020) rely on a nonparametric regression, and thus require several user-chosen quantities.

In this study, we consider the following semi-parametric model, and propose a new approach to estimating the central quantile subspace:

$$Q_\tau(Y | \mathbf{X}) = g(\mathbf{B}_\tau^\top \mathbf{X}), \quad (1.2)$$

where $Q_\tau(Y | \mathbf{X})$ 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 that spans the central quantile subspace. Throughout this paper, we assume that the central quantile subspace and the central subspace exist. We refer to Luo, Li and Yin (2014), Christou (2020), Cook (1998), and Yin, Li and Cook (2008) for discussions on the central quantile subspace and the central subspace. We first introduce a variant of the martingale difference divergence (MDD) (Shao and Zhang (2014)), the so-called quantile martingale difference divergence (QMDD). The QMDD measures the quantile dependence, and we apply it to estimate the central quantile subspace. An appealing feature of our approach is that it does not impose a parametric structure between the conditional quantile of Y given \mathbf{X} and $\mathbf{B}_\tau^\top \mathbf{X}$, nor does it involve any tuning parameters, making it simple and easy to implement. Moreover, we introduce a new bootstrap test to determine the dimension of the central quantile subspace.

We also propose an inverse regression method for estimating the central subspace called the quantile martingale difference divergence matrix (QMDDM) approach, which is as 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 MDD (Shao and Zhang (2014)) and the MDDM (Lee and Shao (2018)). We introduce the QMDD and its properties, and apply it to estimate 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 an application of the proposed method to a real data set. Section 7 concludes the paper. All proofs are relegated to the Supplementary Material.

A word on notation. Let $i = \sqrt{-1}$ be the imaginary unit. The 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 $\mathbf{X} = (x_1, \dots, x_p) \in \mathbb{R}^p$ 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 $\mathbf{A}^\top \mathbf{A}$, the Frobenius norm is denoted as $\|\mathbf{A}\|_F = \sqrt{\text{tr}(\mathbf{A}^\top \mathbf{A})}$, and $\text{tr}(\mathbf{A}) = \sum_{i=1}^p \mathbf{A}_{i,i}$. 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 the MDD and MDDM

Before introducing the new dimension-reduction approaches, we briefly review the MDD and the MDDM. For $\mathbf{U} \in \mathbb{R}^r$ and $V \in \mathbb{R}$, where r is a fixed positive integer, Shao and Zhang (2014) proposed using the MDD to measure the mean dependence of V on \mathbf{U} ; that is,

$$\mathbb{E}(V | \mathbf{U}) = \mathbb{E}(V), \text{ almost surely.} \quad (2.1)$$

Specifically, if $\mathbb{E}(|V|^2 + \|\mathbf{U}\|^2) < \infty$, $\text{MDD}^2(V | \mathbf{U})$ is defined as the nonnegative number

$$\text{MDD}^2(V | \mathbf{U}) = -\mathbb{E} \left[\{V - \mathbb{E}(V)\} \{V' - \mathbb{E}(V')\} \|\mathbf{U} - \mathbf{U}'\| \right], \quad (2.2)$$

where (V', \mathbf{U}') is an independent copy of (V, \mathbf{U}) . The key property is that $\text{MDD}^2(V | \mathbf{U}) = 0$ if and only if (2.1) holds, so it fully characterizes the mean independence of V on \mathbf{U} .

Recently, Lee and Shao (2018) introduced the MDDM, which can be viewed

as an extension of the MDD from a scalar to a matrix, and applied it to reduce the dimensions 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 $\mathbb{E}(\|\mathbf{V}\|^2 + \|\mathbf{U}\|^2) < \infty$, $\text{MDDM}(\mathbf{V} | \mathbf{U})$ is defined as

$$\text{MDDM}(\mathbf{V} | \mathbf{U}) = -\mathbb{E} \left[\{\mathbf{V} - \mathbb{E}(\mathbf{V})\} \{\mathbf{V}' - \mathbb{E}(\mathbf{V}')\}^\top \|\mathbf{U} - \mathbf{U}'\| \right], \quad (2.3)$$

where $(\mathbf{V}', \mathbf{U}')$ is an independent copy of (\mathbf{V}, \mathbf{U}) . From (2.3), it is easy to see that $\text{MDDM}(\mathbf{V} | \mathbf{U}) \in \mathbb{R}^{u \times u}$ is a real, symmetric, and positive semi-definite matrix.

3. Central Quantile Subspace

Oftentimes, 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, as introduced by Luo, Li and Yin (2014); its definition is provided below.

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

$$\begin{aligned} \mathcal{S}_{Q(Y|\mathbf{X})}(\tau) &= \cap \{ \text{span}(\boldsymbol{\beta}) : Q_\tau(Y | \mathbf{X}) = Q_\tau(Y | \boldsymbol{\beta}^\top \mathbf{X}) \text{ almostly surely} \} \\ &:= \text{span}(\mathbf{B}_\tau). \end{aligned} \quad (3.1)$$

Note that the space $\mathcal{S}_{Q(Y|\mathbf{X})}(\tau) = \text{span}(\mathbf{B}_\tau)$ is identifiable, but \mathbf{B}_τ is not. Thus, we seek the identifiable parameter, $\mathcal{S}_{Q(Y|\mathbf{X})}(\tau)$; see Luo, Li and Yin (2014), Sheng and 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 the MDD that is central to our new approach. Its definition and properties are introduced in the following section.

3.1. The QMDD

Using the MDD in (2.2), we state a formal definition of a natural analogue of the MDD 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 QMDD is defined as

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

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

The QMDD is a special case of the MDD, and so inherits the latter's key property, stated in the following proposition. The proof is omitted because the proposition is obtained as a direct consequence of Proposition 1 in Lee, Zhang and Shao (2020).

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. $\text{QMDD}_\tau(Y | \mathbf{X}) \geq 0$.
2. $\text{QMDD}_\tau(Y | \mathbf{X}) = 0$ if and only if $P(Y \leq y_\tau | \mathbf{X}) = P(Y \leq y_\tau)$, almost surely.

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

Definition 3. Given the independent and identically distributed (i.i.d.) observations $(\mathbf{X}_i, Y_i)_{i=1}^n$ from the joint distribution of (\mathbf{X}, Y) , the sample τ th QMDD is a nonnegative number defined as $\widehat{\text{QMDD}}_\tau(Y | \mathbf{X}) = (-1/n^2) \sum_{i,j=1}^n \{\mathbf{1}(Y_i \leq \hat{y}_\tau) - \tau\} \{\mathbf{1}(Y_j \leq \hat{y}_\tau) - \tau\} \|\mathbf{X}_i - \mathbf{X}_j\|$, where \hat{y}_τ is the empirical unconditional τ th quantile of Y .

Using the arguments in Section 2 of Lee and Shao (2018), it can be shown that $\widehat{\text{QMDD}}_\tau(Y | \mathbf{X})$ is nonnegative and is a biased estimator of $\text{QMDD}_\tau(Y | \mathbf{X})$, where the bias is asymptotically negligible when p is fixed. It is possible to define an unbiased estimator of $\text{QMDD}_\tau(Y | \mathbf{X})$ by adopting a \mathcal{U} -centering approach (Székely and Rizzo (2014), Park, Shao and Yao (2015), Zhang, Yao and Shao (2018)). However, the nonnegativeness is preferred to the unbiasedness for the optimization step in Section 3.2.

3.2. Estimation of the central quantile subspace

Our specific goal is to find linear combinations of \mathbf{X} , say $\mathbf{B}_\tau^\top \mathbf{X}$, for $\mathbf{B}_\tau \in \mathbb{R}^{p \times d_\tau}$, that fully describe the conditional quantile of Y given \mathbf{X} , for a prespecified $\tau \in (0, 1)$. In other words, $Q_\tau(Y | \mathbf{X}) = Q_\tau(Y | \mathbf{B}_\tau^\top \mathbf{X})$, almost surely, where \mathbf{B}_τ constructs the central quantile subspace. This implies that modeling the 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 losing any regression information. Interestingly, under Condition 1, and using the results in Cook (1998) and Li, Li and Tsai (2015), we have that $\mathbf{1}(Y \leq y_\tau)$ 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 that

$$P(Y \leq y_\tau | \mathbf{B}_{0,\tau}^\top \mathbf{X}) = P(Y \leq y_\tau), \text{ almost surely.} \quad (3.2)$$

Condition 1. Suppose \mathbf{B}_τ and $\mathbf{B}_{0,\tau}$ are bases that span $\mathcal{S}_{Q(Y|\mathbf{X})}(\tau)$ and $\mathcal{S}_{Q(Y|\mathbf{X})}^\perp(\tau)$, 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 and Yin (2013) and Proposition 2 of Sheng and Yin (2016), and is made for the basis associated with the central quantile subspace and its orthogonal complement. As mentioned in Sheng and Yin (2013) and Sheng and 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 and Yin (2013) for further discussion.

Suppose Condition 1 holds and the structural dimension d_τ is known. Then, for a given τ , we have (3.2), which is equivalent to

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

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

$$\widehat{\mathbf{B}}_{0,\tau} = \underset{\boldsymbol{\beta}_0^\top \boldsymbol{\beta}_0 = I_{p-d_\tau}}{\text{argmin}} \widehat{G}_\tau(\boldsymbol{\beta}_0), \quad (3.3)$$

where $\widehat{G}_\tau(\cdot)$ is the sample counterpart of $G_\tau(\cdot)$. In order to optimize (3.3) with an efficient computational cost, we use the optimization solver with orthogonality constraint proposed by Wen and Yin (2013). They use an efficient first-order updating procedure that preserves the orthogonality constraint, and so achieves a substantial saving in computational time; see Wen and Yin (2013) for more details.

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

$$\mathcal{D}(\mathbf{H}_1, \mathbf{H}_2) = \sqrt{p - d_\tau - \text{tr}(\mathbf{H}_1 \mathbf{H}_1^\top \mathbf{H}_2 \mathbf{H}_2^\top)}. \quad (3.4)$$

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

Condition 2.

- (C1) *The cumulative distribution function (c.d.f.) 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 $\mathbf{B}_{0,\tau}$ that minimizes G_τ . Furthermore, the minimum value of G_τ is obtained at a semi-orthogonal matrix β_0 if and only if $\mathcal{D}(\beta_0, \mathbf{B}_{0,\tau}) = 0$.

The condition (C1) is used in Shao and Zhang (2014) and Zhang, Yao and Shao (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 the 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 $\mathcal{D}(\widehat{\mathbf{B}}_{0,\tau}, \mathbf{B}_{0,\tau}) \rightarrow^p 0$ as $n \rightarrow \infty$.

Remark 1. It is worth mentioning the difference between our method and those of Kong and Xia (2014) and Christou (2020). In the estimation procedure, Kong and Xia (2014) and Christou (2020) rely on a local smoothing quantile regression, whereas we use the 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 (2020) relies on the OLS estimate regressing the nonparametric estimate $\widehat{Q}_\tau(Y | \mathbf{X})$ on \mathbf{X} , which implicitly assumes that $Q_\tau(Y | \mathbf{X})$ is a linear function of $\mathbf{B}_\tau^\top \mathbf{X}$. In contrast, our approach does not impose any model assumptions between $Q_\tau(Y | \mathbf{X})$ 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, Zhang and Shao (2020) introduced a wild bootstrap test for testing the mean independence of functional data using an MDD-type test statistic. Because the QMDD is an analogue of the MDD, we follow the approach in Lee, Zhang and Shao (2020) and propose a new bootstrap test using the QMDD to estimate the structural dimension d_τ . In particular, we sequentially test

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

using the wild bootstrap procedure described below:

1. Compute the test statistic, $T_n = n \cdot \widehat{\text{QMDD}}_\tau(Y \mid \widehat{\mathbf{B}}_{0,\tau}^\top \mathbf{X})$, where $\widehat{\mathbf{B}}_{0,\tau}^\top$ is estimated using (3.3) and $\dim(\text{span}(\widehat{\mathbf{B}}_{0,\tau})) = k$.
2. Generate the bootstrap statistic using

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

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

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

3. Repeat step 2 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 to 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 apply the bootstrap test to determine the dimension d_τ in our simulation study, and observe 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 QMDDM.

4.1. The QMDDM

Most existing inverse regression methods assume the linearity condition, and the first-moment methods hinge on the fact that $\Sigma^{-1}\{\mathbb{E}(\mathbf{X}|Y = y) - \mathbb{E}(\mathbf{X})\} \in$

$\mathcal{S}_{Y|\mathbf{X}}$, for all y , where $\Sigma = \text{var}(\mathbf{X})$. Then, the central subspace, or its subspace, can be obtained by estimating $\mathbb{E}(\mathbf{X}|Y = y) - \mathbb{E}(\mathbf{X})$. Often, a nonparametric approach is used to estimate $\mathbb{E}(\mathbf{X}|Y = y)$, which involves a user-chosen quantity, such as the number of slices. Instead, we focus on the observation that

$$\Sigma^{-1}\{\mathbb{E}(\mathbf{X}|\mathbf{1}(Y \leq y_\tau) = y) - \mathbb{E}(\mathbf{X})\} \in \mathcal{S}_{Y|\mathbf{X}}, \quad \forall \tau \in (0, 1), \quad \forall y \in \{0, 1\}. \quad (4.1)$$

Because (4.1) hinges on the mean dependence, we use the MDDM in Lee and Shao (2018) to characterize this relationship, and define a quantile dependence analogue of the MDDM that 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 QMDDM is defined as

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

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

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

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

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

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

4.2. Estimation of the central subspace

As 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 $\mathbb{E}(\mathbf{X} | \mathbf{B}^\top \mathbf{X})$ 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 on $\mathcal{S}_{Y|\mathbf{X}}$ under different quantiles, we construct a new matrix by following the approaches in Kong and Xia (2014) and Christou (2020).

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 that encodes 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 performance; see Li (1991), Kong and Xia (2014), and Christou (2020). Because we have finite observations of $(\mathbf{X}_i, Y_i)_{i=1}^n$, we approximate Γ by $(\tau_i)_{i=1}^{n-1}$, $\tau_i = i/n$. 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 $\widehat{\text{QMDDM}}_{\mathbf{X}|Y}(\tau_i)$ 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 $\nu_1 > \nu_2 > \dots > \nu_d > 0 = \nu_{d+1} = \dots = \nu_p$.*

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

Theorem 3 suggests that the empirical eigenvalues and eigenvectors of $\hat{\Gamma}$ are reasonable estimators of the population counterparts for large sample sizes. This theorem is proved for fixed p ; the theory for increasing p is left to future research.

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

$$\hat{d} = \operatorname{argmax}_{d \in \{1, \dots, p\}} \left(\frac{n \sum_{m=1}^d (\log(\hat{\lambda}_m + 1) - \hat{\lambda}_m)}{2 \sum_{m=1}^p (\log(\hat{\lambda}_m + 1) - \hat{\lambda}_m)} - 2C_n \times \frac{d(d+1)/2}{p} \right), \quad (4.2)$$

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

5. Numerical Simulations

In this section, we evaluate the finite-sample performance of the proposed methods using simulations, and compare it with that of existing methods. In order to assess the estimation accuracy, we compute the trace correlation (Zhu, Zhu and Feng (2010)), that is, $R = \operatorname{tr}(\mathbf{P}_{\mathbf{B}}\mathbf{P}_{\hat{\mathbf{B}}})/d$, 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 $\operatorname{span}(\mathbf{B})$ is identical to $\operatorname{span}(\hat{\mathbf{B}})$, and $R = 0$ if $\operatorname{span}(\mathbf{B})$ is perpendicular to $\operatorname{span}(\hat{\mathbf{B}})$. Thus, larger values of R indicate more accurate estimates. For each example, we repeat the simulations 100 times and report the 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 analyses to assess the performance of estimating the dimension using the bootstrap procedure of Section 3.3, and using the BIC-type criterion (Feng et al. (2013)) of Section 4.2.

5.1. Central quantile subspace

In this section, we estimate the central quantile subspace. In particular, we compare our method with the QOPG of Kong and Xia (2014) and the SIQR/MIQR of Christou (2020). These two methods involve several user-chosen parameters. We follow the choices made in the code provided and the suggestions in Li (2018), Kong and Xia (2014), and Christou (2020). We try several 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,

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

Method	p	$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		
		$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$	
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)

and linear quantile regression. Throughout the simulations for the central quantile subspace, we consider the sample sizes $n = 200$, and 400 , the dimensions of the predictor vector $p = 5$, and 10 , and the quantiles $\tau = 0.25$, 0.5 , and 0.75 , unless otherwise specified. For the dimension selection, the bootstrap replicate $B = 400$, $\{w_i\}_{i=1}^n$ follows a standard normal distribution, and the significance level is $\alpha = 10\%$.

Example 1. This example is adopted from Christou (2020). The response variable Y is generated as $Y = 3x_1 + x_2 + \varepsilon$, where $\mathbf{X} = (x_1, \dots, x_p)$ and ε is generated independently from a standard normal distribution. For a given τ , $\mathbf{B}_\tau = (3, 1, 0, \dots, 0)^\top / \sqrt{10}$.

Table 1 reports the trace correlation R for each method. All methods estimate the central quantile subspace accurately in terms of a higher R . In addition, all three methods produce comparable results. We observe that when p decreases or n increases, all methods improve. Note that $Q_\tau(Y | \mathbf{X})$ depends on $\mathbf{B}_\tau^\top \mathbf{X}$ in a linear fashion. Therefore, the SIQR of Christou (2020) is expected to perform well because it uses an OLS estimate regressing $\widehat{Q}_\tau(Y | \mathbf{X})$ on \mathbf{X} . It is interesting that the QMDD-based and QOPG-based approaches outperform the SIQR in some cases, for example, when $p = 5$. It appears that the QOPG-based method performs slightly better than the QMDD and SIQR methods.

Table 2. Simulation results for the central τ th quantile subspace estimation. Reported results are the **mean(standard deviation)** of the trace correlation from 100 replications.

Method	p	$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		
		$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$	
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 2. This example is adopted from Kong and Xia (2012), with a slight modification that satisfies Condition 1. The data are generated by $Y = \exp(3\sqrt{2}x_1 + 3\sqrt{2}x_5 - 6 + 6x_3\varepsilon)/(1 + \exp(3\sqrt{2}x_1 + 3\sqrt{2}x_5 - 6 + 6x_3\varepsilon))$, where $\mathbf{X} = (x_1, \dots, x_p)$ are generated independently from $U(0, 1)$, and ε is from $U(-1, 1)$. For a given τ , $\mathbf{B}_\tau = (\sqrt{2}, 0, 2(2\tau - 1), 0, \sqrt{2}, 0, \dots, 0)^\top / \sqrt{4 + 4(2\tau - 1)^2}$.

From Table 2, we can see that the QMDD-based and QOPG-based methods outperform the SIQR method. Note 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 its QMDD-based counterpart for some cases, depending on (τ, h) . Furthermore, the performance of the existing methods changes substantially with different values of c_h , showing its sensitivity to the choice of h . Note that there seems no uniformly best choice of h that 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.

Example 3. This example is taken from Luo, Li and Yin (2014) and addresses the correlated \mathbf{X} case. More specifically, the data are generated by $Y = 1 + x_1 + (1 + 0.4x_2)\varepsilon$, where $\mathbf{X} = (x_1, \dots, x_p)$ is from $N(0, \Sigma)$, $\Sigma = [\Sigma_{ij}]_{i,j=1}^p$, $\Sigma_{ij} = 0.5^{|i-j|}$, and ε follows a standard normal distribution. This is a heteroscedastic model,

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

Method	p	$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		
		$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$	
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)

where $\mathbf{B}_\tau = (1, 0.4\Phi^{-1}(\tau), 0, \dots, 0)^\top / \sqrt{1 + (0.4\Phi^{-1}(\tau))^2}$, and Φ is the c.d.f. of the standard normal distribution.

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

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

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

Table 4. Simulation results for the central τ th quantile subspace estimation. Reported results are the `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)

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

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

From Table 5, it appears that our approach outperforms the existing methods in all cases. Overall, our simulation evidence suggests that the QMDD-based approach can outperform existing methods for both forward and inverse models with quite stable performance. In contrast, the performance of the existing methods is sensitive to the choice of h . Note that our approach does not involve any user-chosen quantities and is simpler to implement.

Lastly, we apply the bootstrap test described in Section 3.3 to select the dimension of the central quantile subspace. Table 6 reports the percentages 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 the dimension selection has reasonable results for all models.

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

Method	p	$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		
		$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$	
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)

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

Model	τ	$\hat{d}_\tau < d_\tau$	$\hat{d}_\tau = d_\tau$	$\hat{d}_\tau > d_\tau$
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 existing inverse regression methods, including the sliced inverse regression

(SIR; Li (1991)), directional regression (DR; Li and Wang (2007)), and cumulative slicing (CUME; Zhu, Zhu and Feng (2010)). For the SIR and DR, the number of slices is five. We consider the sample sizes $n = 200$, and 400 , and dimensions of the predictor $p = 10$, and 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 using $Y = (\beta_1^\top \mathbf{X} + 1)^3 + \varepsilon$, 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 $\beta_1 = (1, 1, 1, 0, 0, \dots, 0)^\top$.

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 $\mathbf{P}_\mathbf{B}$ as R increases.

Example 7. In this example, we consider a heteroscedastic model. In particular, we generate the data using 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 $\beta_1 = (1, 1, 1, 0, 0, \dots, 0)^\top$ and $\beta_2 = (0, 0, 0, 1, 1, 1, 0, \dots, 0)^\top$.

From Table 7, we observe that our QMDDM-based approach is superior to existing methods in terms of having a higher R . This shows that the CUME and SIR are comparable, with the CUME performing slightly 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 from $Y = \beta_1^\top \mathbf{X} / (\beta_2^\top \mathbf{X} + 1.5)^2 + 0.5 + 0.5\beta_1^\top \mathbf{X}\varepsilon$, with \mathbf{X} 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 $\beta_1 = (1, 0, \dots, 0)^\top$, $\beta_2 = (0, 1, 0, \dots, 0)^\top$.

Table 7 also presents the means and standard errors of all approaches for this example. The results suggest that the QMDDM-based method generates higher R values than the other methods do, indicating that the QMDD-based approach outperforms the other methods in all cases.

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

Table 7. Simulation results for the central subspace estimation. Reported results are the 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	$\hat{d} < d$	$\hat{d} = d$	$\hat{d} > d$
Example 6	0	100	0
Example 7	0	100	0
Example 8	14	86	0
Example 9	0	100	0

According to Table 7, it seems that all methods produce accurate results because the R values are sufficiently high. More precisely, the CUME and 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 parameters.

Table 8 reports the percentages of correctly identifying the structural dimension of the central subspace under the previous simulation models when $n = 400$ and $p = 10$. We apply the BIC-type criterion in Section 4.2. We observe that the BIC-type criterion works fairly well under all models.

6. Real-Data Illustration

In this section, we focus on the central quantile subspace, and consider the riboflavin data analyzed by Buhlmann, Kalisch and Meier (2014) and Zhang, Lee and Shao (2020). These data contain 71 samples of the riboflavin production rate and the expression levels of 4,088 genes. The response is the logarithm of the riboflavin production rate, and the predictors are the logarithms of the expression levels of the genes. Owing to the high dimensionality and relatively small sample size, we apply variable screening to the predictors, similarly to Buhlmann, Kalisch and Meier (2014) and Zhang, Lee and Shao (2020). In particular, we select g genes that are strongly related to the riboflavin production rate by using the quantile dependence analogue of the martingale difference correlation in Shao and 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 all of the data with $g = 5$; the estimated direction with $g = 10$ is similar to that in the figure; see the Supplementary Material. The results show some curvatures for $\tau = 0.25, 0.5$, and 0.75 . To evaluate the estimation stability of the central quantile subspace, we consider the bootstrap variability $B^{-1} \sum_{b=1}^B \|\mathbf{P}_{\hat{\mathbf{B}}_\tau} - \mathbf{P}_{\hat{\mathbf{B}}_\tau^b}\|_F$, where $\hat{\mathbf{B}}_\tau$ is the estimated semi-orthogonal matrix on the whole data set, and $\hat{\mathbf{B}}_\tau^b$, for $b = 1, \dots, B$, is the estimated semi-orthogonal matrix on 71 bootstrap samples from $B = 100$ bootstrap replicates. The bootstrap variability is summarized in Table 9, which includes the best results of the existing methods for different bandwidth parameters. The results show that the proposed method generates estimations of the central quantile subspace that are more stable than those of existing methods. Furthermore, the QMDD-based approach requires comparable or less computational time than existing methods do for this data set. In particular, when $g = 5$, $c_h = 0.7$, and $\tau = 0.25$, the QMDD, QOPG, and SIQR methods take 4.64, 14.54, and 5.17 (seconds), respectively, for $B = 100$ bootstrap replicates. The computation was performed on a Windows 10 computer with an Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz processor, 32.0 GB installed memory (RAM), and a 64-bit operating system.

7. Conclusion

In this study, we use two metrics, the QMDD and the QMDDM to estimate the central quantile subspace and the central subspace, respectively. We also introduce a new bootstrap test to select the structural dimension for the cen-

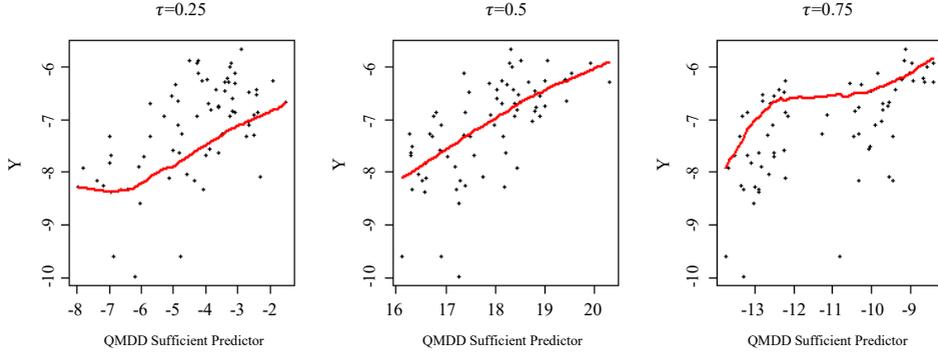


Figure 1. Sufficient summary plots of the central τ th-quantile subspace direction for QMDD approach, with $g = 5$. The solid lines refer to the local quantile regressions for each quantile.

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

	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)

tral quantile subspace, and use a BIC-type criterion to choose the dimension of the central subspace. The finite-sample performance and a real-data application suggest that our QMDD(M)-based approach performs relatively well, and produces results that are more accurate, with comparable or less computational time. In contrast to existing methods for the central quantile subspace or the central subspace, the QMDD(M)-based approach includes no user-chosen parameters, making it convenient and simple to implement. Theoretical results are obtained under suitable conditions, and justify the validity of our methods.

We conclude by mentioning several future research topics. The bootstrap test used to select the structural dimension is worth investigating. A rigorous theoretical study is needed. It is important to understand the theoretical behavior of the proposed approaches when the dimension p is large, and to examine whether we can extend the method to the large p case; however, this seems to be very challenging. Another issue is that we assume Y is univariate. It would be useful to extend our methods to multivariate Y , which seems nontrivial. Lastly, it would be interesting to extend the idea to estimate the envelope quantile regression in Ding et al. (2021), considering the connection between SDR and the predictor envelope model.

Supplementary Material

The online Supplementary Material contains proofs of Theorems 1, 2, 3, an additional simulation, and an additional figure.

Acknowledgments

We would like to thank Dr. Christou for providing the R code used in Christou (2020), and the editor, associate editor, and referee for their thoughtful comments and suggestions.

References

- Buhlmann, P., Kalisch, M. and 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. and Cook, R. D. (2001). Extending sliced inverse regression: The weighted chi-squared test. *Journal of the American Statistical Association* **96**, 996–1003.
- Christou, E. (2020). Central quantile subspace. *Statistics and Computing* **30**, 677–695.
- Cook, R. (1998). *Regression Graphics: Ideas for studying regressions through graphics*. Wiley, New York.
- Cook, R. and Li, B. (2002). Dimension reduction for conditional mean in regression. *The Annals of Statistics* **30**, 455–474.
- Cook, R. and 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. and Wang, L. (2021). Envelope quantile regression. *Statistica Sinica* **31**, 79–106.
- Feng, Z., Wen, X., Yu, Z. and 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. and Xia, Y. (2012). A single-index quantile regression model and its estimation. *Econometric Theory* **28**, 730–768.
- Kong, E. and Xia, Y. (2014). An adaptive composite quantile approach to dimension reduction. *The Annals of Statistics* **42**, 1657–1688.
- Lee, C. E. and 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. and Shao, X. (2020). Testing the conditional mean independence for functional data. *Biometrika* **107**, 331–346.
- Li, B. (2018). *Sufficient Dimension Reduction: Methods and Applications with R*. Monographs on Statistics and Applied Probability. Chapman and Hall/CRC.
- Li, B. and Wang, S. (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association* **102**, 997–1008.
- Li, B., Zha, H. and Chiaromonte, F. (2005). Contour regression: A general approach to dimension reduction. *The Annals of Statistics* **33**, 1580–1616.
- Li, G., Li, Y. and 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. and Yin, X. (2014). On efficient dimension reduction with respect to a statistical functional of interest. *The Annals of Statistics* **42**, 382–412.
- Ma, Y. and Zhu, L. (2013). A review on dimension reduction. *International Statistical Review* **81**, 134–150.
- Park, T., Shao, S. and Yao, S. (2015). Partial martingale difference correlation. *Electronic Journal of Statistics* **9**, 3392–3416.
- Shao, X. and 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. and Yin, X. (2013). Direction estimation in single-index models via distance covariance. *Journal of Multivariate Analysis* **122**, 148–161.
- Sheng, W. and Yin, X. (2016). Sufficient dimension reduction via distance covariance. *Journal of Computational and Graphical Statistics* **25**, 91–104.
- Székely, G. and Rizzo, M. (2014). Partial distance correlation with methods for dissimilarities. *The Annals of Statistics* **42**, 2382–2412.
- Wen, Z. and Yin, W. (2013). A feasible method for optimization with orthogonality constraints. *Mathematical Programming* **142**, 397–434.
- Yin, X. and Cook, R. D. (2002). Dimension reduction for the conditional k th moment in regression. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **64**, 159–175.
- Yin, X. and Li, B. (2011). Sufficient dimension reduction based on an ensemble of minimum average variance estimators. *The Annals of Statistics* **39**, 3392–3416.
- Yin, X., Li, B. and 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. and Shao, X. (2020). Envelopes in multivariate regression models with nonlinearity and heteroscedasticity. *Biometrika* **107**, 965–981.
- Zhang, X., Yao, S. and Shao, X. (2018). Conditional mean and quantile dependence testing in high dimension. *The Annals of Statistics* **46**, 219–246.
- Zhu, L.-P. and Zhu, L.-X. (2009). Dimension reduction for conditional variance in regression. *Statistica Sinica* **19**, 869–883.
- Zhu, L.-P., Zhu, L.-X. and Feng, Z.-H. (2010). Dimension reduction in regressions through cumulative slicing estimation. *Journal of the American Statistical Association* **105**, 1455–1466.
- Zhu, Y. and 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.

Chung Eun Lee

Paul H. Chook Department of Information Systems and Statistics, Baruch College, The City University of New York, 55 Lexington Ave at 24th Street, New York, NY 10010, USA.

E-mail: chung Eun.lee@baruch.cuny.edu

Haileab Hilafu

Business Analytics and Statistics Department, Stokely Management Center, 916 Volunteer Blvd,
Knoxville, TN 37996, USA.

E-mail: hhilafu@utk.edu

(Received January 2020; accepted May 2020)