# Supplementary Materials: Quantile Martingale Difference Divergence for Dimension Reduction

Chung Eun Lee and Haileab Hilafu
Department of Business Analytics and Statistics
University of Tennessee, Knoxville

The following supplementary materials contain proofs of Theorems 1, 2, 3, additional simulation, and additional figure.

## 1 Technical Appendix

Recall that

$$\mathcal{D}(\mathbf{B}_{0,1}, \mathbf{B}_{0,2}) = \sqrt{p - d_\tau - tr(\mathbf{B}_{0,1}\mathbf{B}_{0,1}^\top \mathbf{B}_{0,2}\mathbf{B}_{0,2}^\top)},$$

where $\mathbf{B}_{0,1},\ \mathbf{B}_{0,2} \in \mathcal{H}$, and $\mathcal{H}$ is the set containing all $p \times (p - d_\tau)$ semi-orthogonal matrices, $\mathbf{H}$, i.e., $\mathbf{H}^\top \mathbf{H} = I_{p-d_\tau}$. We shall first state the following lemma which is shown in Pan and Yao (2008).

Lemma **1.1**. *Let $\mathcal{H}_\mathcal{D} := \mathcal{H}/\mathcal{D}$ be the quotient space consisting of all equivalent classes; that is, $\mathbf{H}_1$ and $\mathbf{H}_2$ are treated as the same element in $\mathcal{H}_\mathcal{D}$ if and only if $\mathcal{D}(\mathbf{H}_1, \mathbf{H}_2) = 0$. The set $\mathcal{H}_\mathcal{D}$ and the distance function $\mathcal{D}$ form a metric space in the sense that $\mathcal{D}$ is a well-defined distance function on $\mathcal{H}_\mathcal{D}$.*

We shall use Lemma 1.1 and Lemma 1.2 below for the proof of Theorem 1. Next we shall prove Lemma 1.2.

Lemma **1.2**. *If $\mathbf{X} \in \mathcal{L}^2$, then for any $\mathbf{U},\ \mathbf{V} \in \mathcal{H}_\mathcal{D}$ and any $\tau \in (0, 1)$,*

$$|G_\tau(\mathbf{U}) - G_\tau(\mathbf{V})| \leq c \cdot \mathcal{D}(\mathbf{U}, \mathbf{V}), \tag{1}$$

$$|\widehat{G}_\tau(\mathbf{U}) - \widehat{G}_\tau(\mathbf{V})| \leq c \cdot \mathcal{D}(\mathbf{U}, \mathbf{V}), \tag{2}$$

*where $c > 0$ is a generic constant.*

***Proof of Lemma 1.2***. Due to the fact that $\mathbf{U}$ is a semi-orthogonal matrix, we can rewrite $G_\tau(\mathbf{U})$ as

$$G_\tau(\mathbf{U}) = -\mathbb{E}[(\mathbf{1}(Y \leq y_\tau) - \tau)(\mathbf{1}(Y' \leq y_\tau) - \tau)\|\mathbf{U}\mathbf{U}^\top\mathbf{X} - \mathbf{U}\mathbf{U}^\top\mathbf{X}'\|].$$

Then by the Cauchy-Schwarz inequality and the fact that $\mathcal{D}(\mathbf{U}, \mathbf{V})^2 = \frac{1}{2}\|\mathbf{U}\mathbf{U}^\top - \mathbf{V}\mathbf{V}^\top\|_F^2$ and $E\left[\|\mathbf{X} - \mathbf{X}'\|^2\right] < \infty$, we have

$$\begin{aligned}
& |G_\tau(\mathbf{U}) - G_\tau(\mathbf{V})| \\
= & \left|-\mathbb{E}\left[(\mathbf{1}(Y \leq y_\tau) - \tau)(\mathbf{1}(Y' \leq y_\tau) - \tau)\left\{\|\mathbf{U}\mathbf{U}^\top\mathbf{X} - \mathbf{U}\mathbf{U}^\top\mathbf{X}'\| - \|\mathbf{V}\mathbf{V}^\top\mathbf{X} - \mathbf{V}\mathbf{V}^\top\mathbf{X}'\|\right\}\right]\right| \\
\leq & \ \mathbb{E}\left[(\mathbf{1}(Y \leq y_\tau) - \tau)^2(\mathbf{1}(Y' \leq y_\tau) - \tau)^2\right]^{1/2}\mathbb{E}\left[\left\{\|\mathbf{U}\mathbf{U}^\top\mathbf{X} - \mathbf{U}\mathbf{U}^\top\mathbf{X}'\| - \|\mathbf{V}\mathbf{V}^\top\mathbf{X} - \mathbf{V}\mathbf{V}^\top\mathbf{X}'\|\right\}^2\right]^{1/2} \\
\leq & \ c \cdot \mathbb{E}\left[\left\{\|(\mathbf{U}\mathbf{U}^\top - \mathbf{V}\mathbf{V}^\top)(\mathbf{X} - \mathbf{X}')\|\right\}^2\right]^{1/2} \leq c \cdot \|\mathbf{U}\mathbf{U}^\top - \mathbf{V}\mathbf{V}^\top\|_F \cdot \mathbb{E}\left[\|(\mathbf{X} - \mathbf{X}')\|^2\right]^{1/2} \\
= & \ c \cdot D(\mathbf{U}, \mathbf{V}),
\end{aligned}$$

This completes the proof of (1), and (2) can be shown by using similar arguments.

$\diamondsuit$

***Proof of Theorem 1***. Simple calculation yields that

$$\begin{aligned}
\widehat{G}_\tau(\boldsymbol{\beta}_0) &= \frac{-1}{n^2}\sum_{i,j}(\mathbf{1}(Y_i \leq \widehat{y}_\tau) - \tau)(\mathbf{1}(Y_j \leq \widehat{y}_\tau) - \tau)\|\boldsymbol{\beta}_0^\top\mathbf{X}_i - \boldsymbol{\beta}_0^\top\mathbf{X}_j\| \\
&= \frac{-1}{n^2}\sum_{i,j}(\mathbf{1}(Y_i \leq \widehat{y}_\tau) - \tau)(\mathbf{1}(Y_j \leq \widehat{y}_\tau) - \tau)\widetilde{A}_{ij},
\end{aligned}$$

where $\widetilde{A}_{ij} = a_{ij} - a_{i\cdot} - a_{\cdot j} + a_{\cdot\cdot}$, $a_{ij} = \|\boldsymbol{\beta}_0^\top\mathbf{X}_i - \boldsymbol{\beta}_0^\top\mathbf{X}_j\|$, $a_{i\cdot} = \frac{1}{n}\sum_{k=1}^n a_{ik}$, $a_{\cdot j} =$

$\frac{1}{n}\sum_{l=1}^n a_{lj}$, $a_{\cdot\cdot} = \frac{1}{n^2}\sum_{k,l=1}^n a_{lk}$; see Section 2 in Lee and Shao (2018).

We shall show that

$$\left|\widehat{G}_\tau(\boldsymbol{\beta}_0) - \widetilde{G}_\tau(\boldsymbol{\beta}_0)\right| \to^p 0, \tag{3}$$

where

$$\widetilde{G}_\tau(\boldsymbol{\beta}_0) = \frac{-1}{n^2}\sum_{i,j}\widetilde{A}_{ij}(\mathbf{1}(Y_i \le y_\tau) - \tau)(\mathbf{1}(Y_j \le y_\tau) - \tau),$$

which is replacing $\widehat{y}_\tau$ with $y_\tau$ in $\widehat{G}_\tau(\boldsymbol{\beta}_0)$.

For the ease of notation, we write $\widehat{W}_i = \mathbf{1}(Y_i \le \widehat{y}_\tau)$ and $W_i = \mathbf{1}(Y_i \le y_\tau)$. Due to the fact that $\sum_i \widetilde{A}_{ij} = \sum_j \widetilde{A}_{ij} = 0$, we have

$$\begin{aligned}\left|\widehat{G}_\tau(\boldsymbol{\beta}_0) - \widetilde{G}_\tau(\boldsymbol{\beta}_0)\right| &= \left|\frac{-1}{n^2}\sum_{i,j}\widetilde{A}_{ij}\left\{\widehat{W}_i\widehat{W}_j - W_iW_j\right\}\right| \\ &\le \frac{1}{n^2}\sum_{i,j}\widetilde{C}_{i,j}\left|\widehat{W}_i\widehat{W}_j - W_iW_j\right|,\end{aligned}$$

where $\widetilde{C}_{i,j} = \left\{\|\mathbf{X}_i - \mathbf{X}_j\| + \frac{1}{n^2}\sum_{k,l}\|\mathbf{X}_k - \mathbf{X}_l\|\right\}$. Then

$$\begin{aligned}\frac{1}{n^2}\sum_{i,j}\widetilde{C}_{ij}\left|\widehat{W}_i\widehat{W}_j - W_iW_j\right| &= \frac{1}{n^2}\sum_{i,j}\left|\widehat{W}_i - W_i\right|\widehat{W}_j\widetilde{C}_{ij} + \frac{1}{n^2}\sum_{i,j}\left|\widehat{W}_j - W_j\right|W_i\widetilde{C}_{ij} \\ &= \frac{1}{n^2}\sum_{i,j}\mathbf{1}_{(y_\tau < Y_i \le \widehat{y}_\tau)}\widehat{W}_j\widetilde{C}_{ij} + \frac{1}{n^2}\sum_{i,j}\mathbf{1}_{(\widehat{y}_\tau < Y_i \le y_\tau)}\widehat{W}_j\widetilde{C}_{ij} \\ &+ \frac{1}{n^2}\sum_{i,j}\mathbf{1}_{(y_\tau < Y_j \le \widehat{y}_\tau)}W_i\widetilde{C}_{ij} + \frac{1}{n^2}\sum_{i,j}\mathbf{1}_{(\widehat{y}_\tau < Y_j \le y_\tau)}W_i\widetilde{C}_{ij} \\ &= I_{n,1} + I_{n,2} + I_{n,3} + I_{n,4},\end{aligned}$$

where $I_{n,i},\ i = 1,2,3,4$ are defined implicitly.

Let $\mathcal{A}_\delta = \{\widehat{y}_\tau - y_\tau > \delta\} \cup \{\widehat{y}_\tau - y_\tau \le -\delta\}$. For any $\epsilon > 0$, choose $0 < \delta < \delta_0$ such that $P(\mathcal{A}_\delta) \le \epsilon/2$ for large $n$. For any $\varepsilon > 0$,

$$\begin{aligned}
& P(|I_{n,1}| > \varepsilon) \\
\le\ & P(|I_{n,1}| > \varepsilon, \mathcal{A}_\delta^c) + P(\mathcal{A}_\delta) \le P(|I_{n,1}| > \varepsilon, \mathcal{A}_\delta^c) + \epsilon/2 \\
\le\ & P\left(\frac{1}{n^2}\left|\sum_{i,j} \mathbf{1}(y_\tau < Y_i \le \widehat{y}_\tau, \mathcal{A}_\delta^c)\widehat{W}_j \widetilde{C}_{ij}\right| > \varepsilon\right) + \epsilon/2 \\
\le\ & \frac{1}{\varepsilon n^2}\mathbb{E}\left|\sum_{i,j} \mathbf{1}(y_\tau < Y_i \le \widehat{y}_\tau, \mathcal{A}_\delta^c)\widehat{W}_j \widetilde{C}_{ij}\right| + \epsilon/2 \\
\le\ & \frac{1}{\varepsilon n^2}\sum_{i,j} \mathbb{E}[\mathbf{1}(y_\tau < Y_i \le \widehat{y}_\tau, \mathcal{A}_\delta^c)\widehat{W}_j]^{1/2}\mathbb{E}[\widetilde{C}_{ij}^2]^{1/2} + \epsilon/2 \\
\le\ & \frac{c\delta^{1/2}}{\varepsilon} + \epsilon/2. \qquad (4)
\end{aligned}$$

Here (4) is derived by using the fact that $\mathbb{E}[\widetilde{C}_{ij}^2] < \infty$, and

$$\begin{aligned}
& \mathbb{E}[\mathbf{1}(y_\tau < Y_i \le \widehat{y}_\tau, \mathcal{A}_\delta^c)\widehat{W}_j] \\
\le\ & c\mathbb{E}[\mathbf{1}(y_\tau < Y_{i_1} \le y_\tau + \delta)] \le cG_2(\delta_0)\delta.
\end{aligned}$$

Therefore, we choose a small $\delta$ such that

$$P(|I_{n,1}| > \varepsilon) \le \epsilon,$$

which implies $I_{n,1} = o_p(1)$. Similarly, we also have $I_{n,i} = o_p(1)$, $i = 2, 3, 4$ and obtain (3).

Notice that $\widetilde{G}_\tau(\boldsymbol{\beta}_0)$ can be rewritten as

$$\widetilde{G}_\tau(\boldsymbol{\beta}_0) = \frac{(n-1)}{n}\frac{1}{\binom{n}{2}}\sum_{i<j} H(Z_i, Z_j),$$

where $Z_i = (\mathbf{X}_i, Y_i)$ and

$$H(Z_i, Z_j) = (\mathbf{1}(Y_i < y_\tau) - \tau)(\mathbf{1}(Y_j < y_\tau) - \tau)\left\{-\|\boldsymbol{\beta}_0^\top \mathbf{X}_i - \boldsymbol{\beta}_0^\top \mathbf{X}_j\|\right\}.$$

Under the assumption that $\mathbf{X} \in \mathcal{L}^2$, we have $E[H(Z, Z')^2] < \infty$. Applying Lemma 5.2.1.A (page 183) in Serfling (1980), we obtain $|\widetilde{G}_\tau(\boldsymbol{\beta}_0) - G_\tau(\boldsymbol{\beta}_0)| = O_p(n^{-1/2})$, for any fixed matrix $\boldsymbol{\beta}_0 \in \mathcal{H}$. Furthermore, with (3), we obtain $|\widehat{G}_\tau(\boldsymbol{\beta}_0) - G_\tau(\boldsymbol{\beta}_0)| \rightarrow^p 0$ for any matrix $\boldsymbol{\beta}_0$. Since $\widehat{G}_\tau(\boldsymbol{\beta}_0)$ is equicontinuous, this implies that

$$sup_{\boldsymbol{\beta}_0 \in \mathcal{H}} |\widehat{G}_\tau(\boldsymbol{\beta}_0) - G_\tau(\boldsymbol{\beta}_0)| \rightarrow^p 0. \tag{5}$$

Since (C2) in Condition 2, Lemma 1.1, Lemma 1.2, and (5) are satisfied, we obtain the following result by applying the argmax mapping theorem (Theorem 3.2.2 and Corollary 3.2.3) in van der Vaart and Wellner (1996). Therefore, for a fixed $\tau$, we have

$$\mathcal{D}(\widehat{\mathbf{B}}_{\tau,0}, \mathbf{B}_{\tau,0}) \rightarrow^p 0, \text{ as } n \rightarrow \infty.$$

$\diamondsuit$

***Proof of Theorem 2.*** For simplicity, we assume that $\mathbb{E}(\mathbf{X}) = \mathbf{0}$. Let $\alpha \in \mathbb{R}^p$ be any vector that falls in the orthogonal complement of $\Sigma\text{span}(\mathbf{B})$, i.e., $\alpha^\top \Sigma \mathbf{B} = \mathbf{0}$. We need to show that $\alpha^\top \mathbb{E}(\mathbf{X} \mid \mathbf{1}(Y \leq y_\tau)) = 0, \ \forall \tau \in (0,1)$ which is equivalent to $\alpha^\top \text{QMDDM}_{\mathbf{X}|Y}(\tau)\alpha = 0, \ \forall \tau \in (0,1)$; see Section 4.1. For any $\tau \in (0,1)$,

$$\begin{aligned} \alpha^\top \mathbb{E}\{\mathbf{X} \mid \mathbf{1}(Y \leq y_\tau)\} &= \alpha^\top \mathbb{E}\left[\mathbb{E}\{\mathbf{X} \mid \mathbf{B}^\top \mathbf{X}, \mathbf{1}(Y \leq y_\tau)\} \mid \mathbf{1}(Y \leq y_\tau)\right] \\ &= \mathbb{E}\left[\mathbb{E}(\alpha^\top \mathbf{X} \mid \mathbf{B}^\top \mathbf{X}) \mid \mathbf{1}(Y \leq y_\tau)\right], \end{aligned}$$

where the last equality follows from the fact that $Y \perp\!\!\!\perp \mathbf{X} | \mathbf{B}^\top \mathbf{X}$. We claim that

$\mathbb{E}(\alpha^\top \mathbf{X} \mid \mathbf{B}^\top \mathbf{X}) = 0$ which is equivalent to

$$\mathbb{E}[\{\mathbb{E}(\alpha^\top \mathbf{X} \mid \mathbf{B}^\top \mathbf{X})\}^2] = 0.$$

By the linearity condition,

$$\begin{aligned}
\mathbb{E}[\{\mathbb{E}(\alpha^\top \mathbf{X} \mid \mathbf{B}^\top \mathbf{X})\}^2] &= \mathbb{E}[\mathbb{E}(\{\mathbb{E}(\alpha^\top \mathbf{X} \mid \mathbf{B}^\top \mathbf{X})\}\mathbf{X}^\top \alpha \mid \mathbf{B}^\top \mathbf{X})] \\
&= \mathbb{E}[\{\mathbb{E}(\alpha^\top \mathbf{X} \mid \mathbf{B}^\top \mathbf{X})\}\mathbf{X}^\top \alpha] \\
&= C^\top \mathbf{B}^\top \mathbb{E}[\mathbf{X}\mathbf{X}^\top]\alpha \qquad (6) \\
&= C^\top \mathbf{B}^\top \Sigma \alpha = 0,
\end{aligned}$$

where $C$ is a $d$-dimensional vector such that $\mathbb{E}(\alpha^\top \mathbf{X} \mid \mathbf{B}^\top \mathbf{X}) = C^\top \mathbf{B}^\top \mathbf{X}$. This completes the proof. $\diamondsuit$

***Proof of Theorem 3***. For simplicity, we still assume that $\mathbb{E}(\mathbf{X}) = \mathbf{0}$. Since we consider $\tau_i = \frac{i}{n}, i = 1, \cdots, n-1$, we shall set $y_i = Y_{(i)}$, where $Y_{(i)}$ is the $i$-th smallest observation among $(Y_j)_{j=1}^n$. We denote $\mathrm{QMDDM}_{\mathbf{X}|Y}^{(i)}$ and $\widehat{\mathrm{QMDDM}}_{\mathbf{X}|Y}^{(i)}$ as the QMDDM between $Y$ and $\mathbf{X}$ and its sample counterpart with $y_i$ as the marginal quantile of $Y$. We first show that

$$\|\widehat{\Sigma}^{-1}\widehat{\mathrm{QMDDM}}_{\mathbf{X}|Y}^{(i)} - \Sigma^{-1}\mathrm{QMDDM}_{\mathbf{X}|Y}^{(i)}\|_2^2 = O_p(n^{-1}),\ i = 1, \cdots, n-1. \qquad (7)$$

From (A.1) in Zhu, Zhu, and Feng (2010), we have the following result for all

$i = 1, \cdots, n-1$.

$$
\begin{aligned}
& \widehat{\Sigma}^{-1}\widehat{\mathrm{QMDDM}}^{(i)}_{\mathbf{X}|Y} - \Sigma^{-1}\mathrm{QMDDM}^{(i)}_{\mathbf{X}|Y} \\
= \ & \Sigma^{-1}(\Sigma - \widehat{\Sigma})\widehat{\Sigma}^{-1}\left(\widehat{\mathrm{QMDDM}}^{(i)}_{\mathbf{X}|Y} - \mathrm{QMDDM}^{(i)}_{\mathbf{X}|Y}\right) \\
& + \Sigma^{-1}(\Sigma - \widehat{\Sigma})\widehat{\Sigma}^{-1}\mathrm{QMDDM}^{(i)}_{\mathbf{X}|Y} + \Sigma^{-1}\left(\widehat{\mathrm{QMDDM}}^{(i)}_{\mathbf{X}|Y} - \mathrm{QMDDM}^{(i)}_{\mathbf{X}|Y}\right)
\end{aligned}
$$

Since $\|\Sigma - \widehat{\Sigma}\|_2^2 = O_p(n^{-1})$, we focus on showing $\|\widehat{\mathrm{QMDDM}}^{(i)}_{\mathbf{X}|Y} - \mathrm{QMDDM}^{(i)}_{\mathbf{X}|Y}\|_2^2 = O_p(n^{-1})$ which implies (7).

For any $i = 1, \cdots, n-1$, we have

$$
\|\widehat{\mathrm{QMDDM}}^{(i)}_{\mathbf{X}|Y} - \mathrm{QMDDM}^{(i)}_{\mathbf{X}|Y}\|_2 \leq \|\widehat{\mathrm{QMDDM}}^{(i)}_{\mathbf{X}|Y} - \mathrm{QMDDM}^{(i)}_{\mathbf{X}|Y}\|_F
$$

where $\widehat{\mathrm{QMDDM}}^{(i)}_{\mathbf{X}|Y} = \frac{-1}{n^2}\sum_{h,l=1}^{n} \mathbf{X}_h \mathbf{X}_l^\top |\mathbf{1}(Y_h \leq y_i) - \mathbf{1}(Y_l \leq y_i)| = \left[\left(\widehat{\mathrm{QMDDM}}^{(i)}_{\mathbf{X}|Y}\right)_{s,t}\right]_{s,t=1}^{p}$.

For any $(s,t)$, we can rewrite $\left(\widehat{\mathrm{QMDDM}}^{(i)}_{\mathbf{X}|Y}\right)_{s,t}$ as

$$
\begin{aligned}
\left(\widehat{\mathrm{QMDDM}}^{(i)}_{\mathbf{X}|Y}\right)_{s,t} &= \frac{(n-1)}{n}\frac{1}{\binom{n}{2}}\sum_{j<m} h(Z_j, Z_m) \\
&:= \frac{(n-1)}{n}\mathcal{U}_n,
\end{aligned}
$$

where $\mathcal{U}_n$ is defined implicitly and

$$
h(Z_j, Z_m) = \frac{-1}{2!}\sum_{(v,u)}^{(j,m)} \mathbf{X}_{v,s}\mathbf{X}_{u,t}|\mathbf{1}(Y_v \leq y_i) - \mathbf{1}(Y_u \leq y_i)|,
$$

with $Z_j = (\mathbf{X}_j, Y_j)$ and $\sum_{(v,u)}^{(j,m)}$ is the summation over all permutations of the 2-tuple of indices $(j,m)$.

Under the assumption that $\mathbf{X} \in \mathcal{L}^2$, we have $\mathbb{E}[h(Z, Z')^2] < \infty$. Then by applying Lemma 5.2.1.A (page 183) in Serfling (1980) to $\mathcal{U}_n$, we obtain

$$\left| \left( \widehat{\text{QMDDM}}^{(i)}_{\mathbf{X}|Y} \right)_{s,t} - \left( \text{QMDDM}^{(i)}_{\mathbf{X}|Y} \right)_{s,t} \right|^2 = O_p(n^{-1}),$$

and implies that

$$\begin{aligned} & \|\widehat{\text{QMDDM}}^{(i)}_{\mathbf{X}|Y} - \text{QMDDM}^{(i)}_{\mathbf{X}|Y}\|_F^2 \\ \leq & \sum_{s,t=1}^{p} \left| \left( \widehat{\text{QMDDM}}^{(i)}_{\mathbf{X}|Y} \right)_{s,t} - \left( \text{QMDDM}^{(i)}_{\mathbf{X}|Y} \right)_{s,t} \right|^2 = O_p(n^{-1}) \end{aligned} \tag{8}$$

Thus we have $\|\widehat{\text{QMDDM}}^{(i)}_{\mathbf{X}|Y} - \text{QMDDM}^{(i)}_{\mathbf{X}|Y}\|_2 = O_p(n^{-1/2})$ and obtain (7).

We let $\{\lambda_{j,i}, \gamma_{j,i}\}_{j=1}^{p}$ be the eigenvalues and eigenvectors of $\text{QMDDM}^{(i)}_{\mathbf{X}|Y}$. Also, we let $\{\widehat{\lambda}_{j,i}, \widehat{\gamma}_{j,i}\}_{j=1}^{p}$ be the sample counterparts. Next, we use Lemma A.1. in Kneip and Utikal (2001) to show $\|\widehat{\gamma}_{1,i} - \gamma_{1,i}\| = O_p(n^{-1/2})$ for $i = 1, \cdots, n-1$.

By applying part (b) of Lemma A.1. in Kneip and Utikal (2001), we have

$$\widehat{\gamma}_{1,i} - \gamma_{1,i} = -\left\{ \sum_{h \neq 1} \frac{1}{\lambda_{h,i} - \lambda_{1,i}} \gamma_{h,i} \gamma_{h,i}^T \right\} \left( \widehat{\Sigma}^{-1} \widehat{\text{QMDDM}}^{(i)}_{\mathbf{X}|Y} - \Sigma^{-1} \text{QMDDM}^{(i)}_{\mathbf{X}|Y} \right) \gamma_{1,i} + \mathcal{R}_2,$$

where

$$\|\mathcal{R}_2\|_2 \leq \frac{6\|\widehat{\Sigma}^{-1}\widehat{\text{QMDDM}}^{(i)}_{\mathbf{X}|Y} - \Sigma^{-1}\text{QMDDM}^{(i)}_{\mathbf{X}|Y}\|_2^2}{min_{\lambda \in E(\Sigma^{-1}\text{QMDDM}^{(i)}_{\mathbf{X}|Y}), \lambda \neq \lambda_{1,i}} |\lambda - \lambda_{1,i}|^2} = O_p(n^{-1}),$$

and $E(A)$ denotes the set of eigenvalues of a matrix A. By using (7),

$$\left\| - \left\{ \sum_{h\neq 1} \frac{1}{\lambda_{h,i} - \lambda_{1,i}} \gamma_{h,i}\gamma_{h,i}^\top \right\} \left( \widehat{\Sigma}^{-1} \mathrm{Q}\widehat{\mathrm{MDDM}}_{\mathbf{X}|Y}^{(i)} - \Sigma^{-1}\mathrm{QMDDM}_{\mathbf{X}|Y}^{(i)} \right) \gamma_{1,i} \right\|^2$$

$$= \sum_{h\neq 1} \frac{\left\{ \gamma_{h,i}^\top \left( \widehat{\Sigma}^{-1} \mathrm{Q}\widehat{\mathrm{MDDM}}_{\mathbf{X}|Y}^{(i)} - \Sigma^{-1}\mathrm{QMDDM}_{\mathbf{X}|Y}^{(i)} \right) \gamma_{1,i} \right\}^2}{(\lambda_{h,i} - \lambda_{1,i})^2} = O_p(n^{-1}).$$

Hence, we have

$$\|\widehat{\gamma}_{1,i} - \gamma_{1,i}\| = O_p(n^{-1/2}) \text{ for all } i = 1, \cdots, n-1. \tag{9}$$

Finally, we show that

$$\|\widehat{\Gamma} - \Gamma\|_2 = O_p(n^{-1/2}). \tag{10}$$

Based on the definition of $\widehat{\Gamma}$, we have

$$\begin{aligned} \widehat{\Gamma} - \Gamma &= \frac{1}{n}\widehat{V}\widehat{V}^\top - \frac{1}{n}VV^\top + \frac{1}{n}VV^\top - \Gamma \\ &= J_{n,1} + J_{n,2}, \end{aligned} \tag{11}$$

where $V = (\gamma_{1,1}, \cdots, \gamma_{1,n-1}) \in \mathbb{R}^{p\times n-1}$ and $\widehat{V} = (\widehat{\gamma}_{1,1}, \cdots, \widehat{\gamma}_{1,n-1}) \in \mathbb{R}^{p\times n-1}$, $J_{n,1} = \frac{1}{n}\widehat{V}\widehat{V}^\top - \frac{1}{n}VV^\top$, and $J_{n,2} = \frac{1}{n}VV^\top - \Gamma$.
Using (9), we obtain

$$\|J_{n,1}\|_2^2 \leq \|J_{n,1}\|_F^2 \leq \frac{c}{n^2}\|\widehat{V}(\widehat{V} - V)^\top\|_F^2 + \frac{c}{n^2}\|(\widehat{V} - V)V^\top\|_F^2 \leq O_p(n^{-1}), \tag{12}$$

where $c > 0$ is a constant. Furthermore, since $\Gamma(\tau)$ satisfies (D2) in Condition 3, we apply Theorem 1 (c) in Chui (1971) to $J_{n,2}$ and obtain $\|J_{n,2}\|_2 \leq \|J_{n,2}\|_F = o(n^{-1})$

which implies (10). Then by using the same argument showing (9), we have

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

For the eigenvalues, we apply part (a) of Lemma A.1. in Kneip and Utikal (2001). Then we have

$$\widehat{\nu}_j - \nu_j = tr\left(\eta_j \eta_j^T \{\widehat{\Gamma} - \Gamma\}\right) + \mathcal{R}_1 \ \text{ for } j = 1, \cdots, d, \tag{13}$$

where

$$|\mathcal{R}_1| \leq \frac{6\|\widehat{\Gamma} - \Gamma\|_2^2}{min_{\nu \in E(\Gamma), \nu \neq \nu_j} |\nu - \nu_j|} = O_p(n^{-1}). \tag{14}$$

As we further have $\left|tr\left(\eta_j \eta_j^T \{\widehat{\Gamma} - \Gamma\}\right)\right| \leq \|\widehat{\Gamma} - \Gamma\|_2 = O_p(n^{-1/2})$, we obtain

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

This completes the proof. ♢

# 2 Additional Simulation

Following the suggestion of a referee, we consider models with infinite variance noise.

Example **2.1**.

$$Y = 3x_1 + x_2 + \varepsilon,$$

where $\mathbf{X} = (x_1, \cdots, x_p)$ is independently generated from standard normal distribution and $\varepsilon$ is generated by Cauchy$(0, 0.25)$. For a given $\tau$, $\mathbf{B}_\tau = (3, 1, 0, \cdots, 0)^T/\sqrt{10}$.

Table 1: Simulation results for the central $\tau-$th quantile subspace estimation for Example 2.1. Reported results are `mean(standard deviation)` of the trace correlation from 100 replications.

| Method | | $p$ | $\tau=0.25$ $n=200$ | $\tau=0.25$ $n=400$ | $\tau=0.5$ $n=200$ | $\tau=0.5$ $n=400$ | $\tau=0.75$ $n=200$ | $\tau=0.75$ $n=400$ |
|---|---|---|---|---|---|---|---|---|
| QMDD | | 5 | 0.97 (0.02) | 0.99 (0.01) | 0.99 (0.01) | 0.99 (0.01) | 0.97 (0.02) | 0.99 (0.01) |
| | | 10 | 0.94 (0.04) | 0.97 (0.01) | 0.96 (0.02) | 0.98 (0.01) | 0.94 (0.03) | 0.97 (0.01) |
| QOPG | $c_h=0.7$ | 5 | 0.99 (0.01) | 1.00 (0.00) | 0.99 (0.01) | 1.00 (0.00) | 0.99 (0.00) | 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 (0.00) |
| | $c_h=1.5$ | 5 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| | | 10 | 0.99 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 0.99 (0.00) | 1.00 (0.00) |
| | $c_h=2.34$ | 5 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| | | 10 | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| SIQR | $c_h=0.7$ | 5 | 0.45 (0.29) | 0.38 (0.26) | 0.45 (0.28) | 0.39 (0.26) | 0.45 (0.29) | 0.39 (0.26) |
| | | 10 | 0.40 (0.30) | 0.35 (0.29) | 0.41 (0.29) | 0.37 (0.29) | 0.40 (0.30) | 0.35 (0.29) |
| | $c_h=1.5$ | 5 | 0.41 (0.30) | 0.34 (0.28) | 0.41 (0.29) | 0.34 (0.27) | 0.42 (0.30) | 0.35 (0.28) |
| | | 10 | 0.36 (0.31) | 0.31 (0.30) | 0.37 (0.31) | 0.31 (0.30) | 0.36 (0.31) | 0.31 (0.30) |
| | $c_h=2.34$ | 5 | 0.39 (0.30) | 0.33 (0.28) | 0.40 (0.30) | 0.33 (0.28) | 0.40 (0.30) | 0.33 (0.28) |
| | | 10 | 0.34 (0.32) | 0.29 (0.31) | 0.34 (0.32) | 0.30 (0.31) | 0.34 (0.32) | 0.29 (0.31) |

EXAMPLE **2.2**.

$$Y=\sqrt{x_1+1}+\sqrt{x_2+1}+\epsilon,$$

where $\mathbf{X}=(x_1,\cdots,x_p)$ is generated by $\chi^2(2)$, and $\varepsilon$ is from Cauchy$(0,0.25)$. Here $\mathbf{B}_\tau=(\beta_1,\beta_2)$, where $\beta_1=(1,0,\cdots,0)^\top$ and $\beta_2=(0,1,0,\cdots,0)^\top$.

Table 2: Simulation results for the central $\tau-$th quantile subspace estimation for Example 2.2. 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 | 0.82 (0.10) | 0.90 (0.07) | 0.89 (0.07) | 0.94 (0.04) | 0.89 (0.08) | 0.95 (0.04) |
| | | 10 | 0.68 (0.13) | 0.79 (0.09) | 0.80 (0.09) | 0.89 (0.06) | 0.80 (0.10) | 0.88 (0.06) |
| QOPG | $c_h = 0.7$ | 5 | 0.58 (0.12) | 0.61 (0.12) | 0.57 (0.10) | 0.62 (0.14) | 0.59 (0.12) | 0.60 (0.11) |
| | | 10 | 0.43 (0.10) | 0.50 (0.07) | 0.45 (0.08) | 0.51 (0.07) | 0.45 (0.10) | 0.50 (0.07) |
| | $c_h = 1.5$ | 5 | 0.61 (0.13) | 0.62 (0.13) | 0.63 (0.14) | 0.66 (0.14) | 0.64 (0.15) | 0.64 (0.14) |
| | | 10 | 0.50 (0.07) | 0.52 (0.05) | 0.53 (0.07) | 0.56 (0.09) | 0.52 (0.08) | 0.56 (0.09) |
| | $c_h = 2.34$ | 5 | 0.62 (0.13) | 0.64 (0.15) | 0.69 (0.15) | 0.74 (0.16) | 0.71 (0.16) | 0.69 (0.15) |
| | | 10 | 0.51 (0.06) | 0.52 (0.04) | 0.56 (0.08) | 0.60 (0.10) | 0.57 (0.11) | 0.62 (0.12) |
| MIQR | $c_h = 0.7$ | 5 | 0.59 (0.18) | 0.60 (0.16) | 0.63 (0.17) | 0.62 (0.17) | 0.58 (0.16) | 0.59 (0.17) |
| | | 10 | 0.41 (0.16) | 0.43 (0.17) | 0.46 (0.14) | 0.49 (0.16) | 0.40 (0.15) | 0.42 (0.18) |
| | $c_h = 1.5$ | 5 | 0.57 (0.15) | 0.58 (0.16) | 0.58 (0.15) | 0.57 (0.16) | 0.57 (0.15) | 0.57 (0.18) |
| | | 10 | 0.42 (0.16) | 0.41 (0.16) | 0.42 (0.13) | 0.45 (0.15) | 0.39 (0.15) | 0.41 (0.15) |
| | $c_h = 2.34$ | 5 | 0.56 (0.17) | 0.56 (0.15) | 0.54 (0.14) | 0.56 (0.17) | 0.53 (0.13) | 0.53 (0.16) |
| | | 10 | 0.39 (0.15) | 0.40 (0.16) | 0.39 (0.14) | 0.40 (0.15) | 0.36 (0.14) | 0.38 (0.14) |

From Table 1, it shows that our method and the QOPG approach provide accurate estimates of the central quantile subspace and outperform the SIQR method. From Table 2, we observe that our approach produce more accurate results than the other existing methods in terms of a higher trace correlation.

# 3 Additional Figure

Below is the sufficient summary plots for the estimated directions of the central $\tau$th-quantile subspace applying QMDD-based approach for the riboflavin data set when $g = 10$.

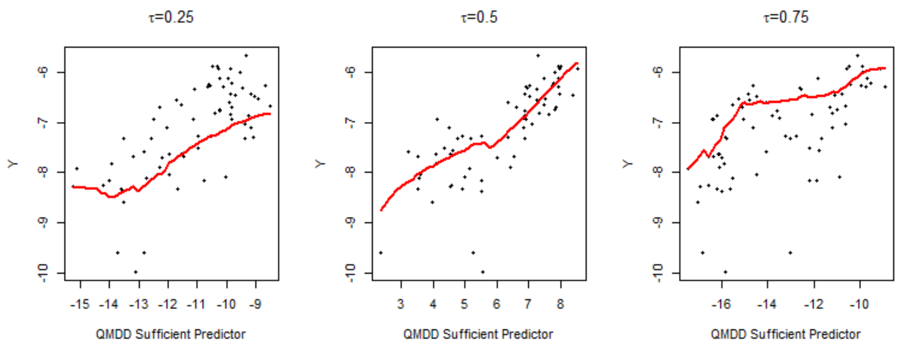


Figure 1: Sufficient summary plots of the central $\tau$th-quantile subspace for QMDD-based approach with $g = 10$. The solid lines refer to the local quantile regressions for each quantile.

## References

Chui, C. (1971). Concerning rates of convergence of Riemann sums. *Journal of Approximation Theory*, **4**, 279-287.

Kneip, A. and Utikal, K. J. (2001). Inference for density families using functional principal component analysis. *Journal of the American Statistical Association*, **96**, 519-542.

Lee, C. 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.

Pan, J. and Yao, Q. (2008). Modelling multiple time series via common factors. *Biometrika*, **95**, 365-379.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer.

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.