Central quantile subspace

Eliana Christou

Abstract
Quantile regression (QR) is becoming increasingly popular due to its relevance in many scientific investigations. There is a great amount of work about linear and nonlinear QR models. Specifically, nonparametric estimation of the conditional quantiles received particular attention, due to its model flexibility. However, nonparametric QR techniques are limited in the number of covariates. Dimension reduction offers a solution to this problem by considering low-dimensional smoothing without specifying any parametric or nonparametric regression relation. The existing dimension reduction techniques focus on the entire conditional distribution. We, on the other hand, turn our attention to dimension reduction techniques for conditional quantiles and introduce a new method for reducing the dimension of the predictor \( X \). The novelty of this paper is threefold. We start by considering a single index quantile regression model, which assumes that the conditional quantile depends on \( X \) through a single linear combination of the predictors, then extend to a multi-index quantile regression model, and finally, generalize the proposed methodology to any statistical functional of the conditional distribution. The performance of the methodology is demonstrated through simulation examples and real data applications. Our results suggest that this method has a good finite sample performance and often outperforms the existing methods.

Keywords
Dimension reduction · Multi-index · Quantile regression · Single index · Statistical functional

1 Introduction

Quantile regression (QR) was first introduced by Koenker and Bassett (1978), and since then, it has received a lot of attention. For a univariate response \( Y \) and a \( p \times 1 \) predictor vector \( X \), let, for \( \tau \in (0, 1) \),
\[
Q_\tau(Y|X) = \inf \{ y : \Pr(Y \leq y|X = x) \geq \tau \}
\] denote the \( \tau \)th conditional quantile of \( Y \) given \( X = x \). Then, \( Q_\tau(Y|X) \) satisfies
\[
Q_\tau(Y|X) = \arg\min_q E(\rho_\tau(Y - q)|X = x),
\]
where \( \rho_\tau(\cdot) \) is the loss function, also known as check function, defined as \( \rho_\tau(u) = (\tau - I(u < 0))u \). There is a considerably large body of work about QR models. Specifically, Koenker and Bassett (1978) considered the linear QR model
\[
Q_\tau(Y|X) = \alpha_\tau + \beta_\tau^\top x,
\]
where \( \alpha_\tau \in \mathbb{R}, \beta_\tau \in \mathbb{R}^p \), and used the representation (1) to define the estimator \((\hat{\alpha}_\tau, \hat{\beta}_\tau)\) as
\[
(\hat{\alpha}_\tau, \hat{\beta}_\tau) = \arg\min_{(\alpha_\tau, \beta_\tau)} \sum_{i=1}^n \rho_\tau(Y_i - \alpha_\tau - \beta_\tau^\top X_i).
\]
for \( \{Y_i, X_i\}_{i=1}^n \) independent and identically distributed (i.i.d.) observations. Then, \( \hat{\alpha}_\tau + \hat{\beta}_\tau^\top x \) gives the estimator of the \( \tau \)th conditional quantile under the linear QR model.

Because the linearity assumption is quite strict, several authors considered the completely flexible nonparametric QR model; see, for example, Chaudhuri (1991), Yu and Jones (1998), Kong et al. (2010), and Guerre and Sabbah (2012). However, when the number of the predictors is large, nonparametric methods require smoothing over a high-dimensional space, where the data becomes sparse. Dimension reduction offers a solution to this problem by considering low-dimensional smoothing without specifying any parametric or nonparametric regression relation.

The existing dimension reduction techniques focus on the entire conditional distribution and include, among others, sliced inverse regression (SIR, Li 1991), principal Hessian directions (pHd, Li 1992), sliced average variance estimation (SAVE, Cook and Weisberg 1991), parametric inverse regression (PIR, Bura and Cook 2001), minimum average variance estimation (MAVE, Xia et al. 2002), partial SIR
When specific aspects of the conditional distribution are of interest, such as the conditional mean, conditional variance, and conditional quantile of the response given the covariates, the above methods can be inefficient because they might provide more directions than necessary. This is because the methods focus on the complete conditional distribution of \( Y|X \) and thereby give a full picture of the dependence of \( Y \) on \( X \), while more specific characteristics of \( Y|X \) might be of interest (Cook and Li 2002). Recent research focuses on these statistical functionals. Specifically, Cook and Li (2002) focused on the conditional mean and introduced the central mean subspace (CMS), Yin and Cook (2002) focused on the conditional kth moment and introduced the central kth moment subspace (CKMS), and Zhu and Zhu (2009) focused on the conditional variance and introduced the central variance subspace (CVS).

In this work, we focus on dimension reduction techniques for conditional quantiles and propose a new method for finding the fewest linear combinations of \( X \) that contain all the information on that function. The existing literature considers a single-index quantile regression (SIQR) model and proposes an iterative algorithm for estimating the vector of the coefficients of the linear combination of \( X \) (Wu et al. 2010; Kong and Xia 2012). To avoid iterations and convergence issues, Christou and Akritas (2016) proposed a non-iterative algorithm for estimating that vector of coefficients. While their methodology has been extended to allow for simultaneous variable selection and parameter estimation (Christou and Akritas 2018), it is still limited to a small number of covariates. Kong and Xia (2014) proposed an adaptive composite QR approach, which can be used for estimating multiple linear combinations of \( X \) that contain all the information about the conditional quantile, while Luo et al. (2014) introduced a sufficient dimension reduction method that targets any statistical functional of interest, including the conditional quantile. The authors proposed an efficient estimator for the semiparametric estimation problem of this type. However, while their work covers a wide range of applications, the finite sample performance of the efficient estimator is not necessarily the best. For the above reasons, a different approach is needed.

In this paper, we introduce the \( \tau \)th central quantile subspace (\( \tau \)-CQS; see Sect. 2) and propose an algorithm for estimating it. Specifically, in Sect. 3, we focus on a SIQR model and propose a method for estimating the one-dimensional \( \tau \)-CQS, in Sect. 4 we extend the methodology to a multi-index quantile regression (MIQR) model, while in Sect. 5 we generalize the proposed methodology to any statistical functional of interest. In Sect. 6, we present a brief discussion on the estimation of the dimension of the \( \tau \)-CQS, and in Sect. 7, we present results from several simulation examples and real data applications. A discussion is given in Sect. 8. The assumptions, some lemmas, and the proofs of Theorems 4 and 8 are given in the “Appendix.”

2 The \( \tau \)th central quantile subspace

We start by recalling some basic definitions from Li (1991). For a univariate response \( Y \) and a \( p \times 1 \) vector of predictors \( X \), let \( A = (\alpha_1, \ldots, \alpha_d) \) denote a \( p \times d \) matrix, where \( \alpha_1, \ldots, \alpha_d \) are column vectors and \( d \leq p \). Assume that \( Y \perp X|A^\top X \), i.e., that \( Y \) and \( X \) are independent given \( A^\top X \). This means that, the \( d \times 1 \) predictor vector \( A^\top X \) captures all we need to know about \( Y \), implying that we can replace the \( p \times 1 \) predictor vector \( X \) with the \( d \times 1 \) predictor vector \( A^\top X \) without loss of information. The space spanned by the column vectors \( \alpha_1, \ldots, \alpha_d \), denoted by \( S(A) \), is called the dimension reduction subspace for the regression of \( Y \) on \( X \). The greatest dimension reduction in the predictor vector is achieved using the smallest dimension reduction subspace, called the central subspace (CS), and denoted by \( S_Y|X \).

A straightforward extension of the CS to conditional quantiles, and a special case of Definition 1 of Luo et al. (2014), states the following. If \( Y \perp \tau(Y|X)|B_\tau^\top X \), where \( B_\tau \) is a \( p \times d_\tau \) matrix, \( d_\tau \leq p \), then the space spanned by \( B_\tau \), denoted by \( S(B_\tau) \), is a \( \tau \)th quantile dimension reduction subspace for the regression of \( Y \) on \( X \). This implies that the \( d_\tau \times 1 \) predictor vector \( B_\tau^\top X \) contains all the information about \( Y \) that is available from \( \tau(Y|X) \). The \( \tau \)th central quantile subspace (\( \tau \)-CQS), denoted by \( S_{\tau|Y|X} \), is defined to be the intersection of all \( \tau \)th quantile dimension reduction subspaces. For the remainder of this paper, we assume that the \( \tau \)-CQS exists.

The following notation will be used throughout the rest of the paper. The CS is spanned by the \( p \times d \) matrix \( A \), i.e., \( S_Y|X = S(A) \), and for a given \( \tau \), the \( \tau \)-CQS is spanned by the \( p \times d_\tau \) matrix \( B_\tau \), i.e., \( S_{\tau|Y|X} = S(B_\tau) \). The matrices \( A \) and \( B_\tau \) are called the basis matrices. It is easy to see that \( S_{\tau|Y|X} \subseteq S_Y|X \), for any \( \tau \). Therefore, \( B_\tau^\top X \) provides a refined structure for the CS, i.e., \( B_\tau^\top X = C_\tau^\top A^\top X \), where \( C_\tau \) is a \( d \times d_\tau \) matrix.

3 Estimation of the \( \tau \)-CQS for a SIQR model

3.1 Population level

A SIQR model assumes that \( Q_\tau(Y|X) = g_\tau(B_\tau^\top x) \), where \( g_\tau(\cdot) : \mathbb{R} \rightarrow \mathbb{R} \) is an unknown univariate link function,
called the nonparametric component, and \( B_\tau \in \mathbb{R}^p \) is a fixed, but unknown, vector of parameters, called the parametric component. Since the quantile is assumed to depend on \( x \) only through \( B_\tau^\top x \), we sometimes write \( Q_\tau(Y|B_\tau^\top x) \) instead of \( Q_\tau(Y|x) \). The SIQR model received particular attention as its nonparametric component is univariate and thus tractable. For example, Fan et al. (2018) and Christou and Grabchak (2019) used the SIQR model for value-at-risk estimation.

The SIQR model implies that \( Y \perp Q_\tau(Y|X)|B_\tau^\top X \) and therefore assumes a one-dimensional \( \tau \)-CQS. The goal is to estimate the parametric component \( B_\tau \), which corresponds to the vector of coefficients for the linear combination \( B_\tau^\top X \). Let

\[
R(\alpha_\tau, b_\tau) = E[L(\alpha_\tau + b_\tau^\top X, Q_\tau(Y|X))],
\]

where \( \alpha_\tau \in \mathbb{R}, \ b_\tau \in \mathbb{R}^p \), and \( L(\theta_\tau, Q_\tau(Y|X)) \) is a function strictly convex in \( \theta_\tau \) with a unique minimum at \( \theta_\tau = Q_\tau(Y|X) \). Under the SIQR model, and if the conditional expectation \( E(\beta_\tau^\top X|B_\tau^\top X) \) is linear in \( B_\tau^\top X \) for every \( B_\tau \in \mathbb{R}^p \), then \( \beta_\tau \in S_{Q_\tau(Y|X)} \), where

\[
(a_\tau, b_\tau) = \arg \min_{(a_\tau, b_\tau)} R(\alpha_\tau, b_\tau).
\]

This implies that \( \beta_\tau \) is equal to the coefficients of the linear combination of the predictors up to a multiplier, i.e., \( \beta_\tau = cB_\tau \) for some \( c \in \mathbb{R} \setminus \{0\} \). This idea comes from the work of Brillinger (1983) and Li and Duan (1989), who considered a similar task but for estimating the one-dimensional CS.

Moreover, since \( S_{Y|X} = S(A) \), then minimizing \( R(\alpha_\tau, b_\tau) \) with respect to \( \alpha_\tau \) and \( b_\tau \), is the same as minimizing \( R^*(\alpha_\tau, b_\tau) \), where

\[
R^*(\alpha_\tau, b_\tau) = E\left[L(\alpha_\tau + b_\tau^\top X, Q_\tau(Y|A^\top X))\right].
\]

Then, \( \beta_\tau^* \in S_{Q_\tau(Y|X)} \), where

\[
(a_\tau^*, b_\tau^*) = \arg \min_{(a_\tau, b_\tau)} R^*(\alpha_\tau, b_\tau).
\]

This allows for an initial dimension reduction using \( A \) for all choices of \( \tau \), which is then converted into an estimate of \( B_\tau \) for a specific \( \tau \).

**Assumption 1** For a given \( \tau \), the conditional expectation \( E(\beta_\tau^\top X|B_\tau^\top X) \) is linear in \( B_\tau^\top X \) for every \( B_\tau \in \mathbb{R}^p \).

**Theorem 1** For a given \( \tau \in (0, 1) \), assume that \( Y \perp Q_\tau(Y|X)|B_\tau^\top X \), where \( B_\tau \) is a \( p \times 1 \) vector. Under Assumption 1 and if

\[
(a_\tau^*, b_\tau^*) = \arg \min_{(a_\tau, b_\tau)} R^*(\alpha_\tau, b_\tau),
\]

where \( R^*(\alpha_\tau, b_\tau) \) is defined in (2), then \( \beta_\tau^* \in S_{Q_\tau(Y|X)} \).

**Proof** Observe that

\[
R^*(\alpha_\tau, b_\tau) = E[L(\alpha_\tau + b_\tau^\top X, Q_\tau(Y|A^\top X))]
\]

\[
= E[E[L(\alpha_\tau + b_\tau^\top X, Q_\tau(Y|A^\top X)|B_\tau^\top X]]]
\]

\[
\geq E[L(\alpha_\tau + b_\tau^\top X|B_\tau^\top X, E(Q_\tau(Y|A^\top X)|B_\tau^\top X))]
\]

\[
= E[E[L(\alpha_\tau + b_\tau^\top X|B_\tau^\top X, Q_\tau(Y|X))]
\]

\[
= E[L(\alpha_\tau + b_\tau^\top X|B_\tau^\top X, Q_\tau(Y|X))] = R^*(\alpha_\tau, d^aB_\tau^\top).
\]

where \( P_{B_\tau}(\Sigma_{xx}) = B_\tau(B_\tau^\top \Sigma_{xx}B_\tau)^{-1}B_\tau^\top \Sigma_{xx} \), \( \Sigma_{xx} \) is the covariance matrix of \( X \), and \( d^a \) is a constant. The first inequality follows from Jensen’s inequality. Moreover, the fourth line follows from the fact that \( E(Q_\tau(Y|A^\top X)|B_\tau^\top X) = E(Q_\tau(Y|B_\tau^\top X)|B_\tau^\top X) = Q_\tau(Y|X) \), and the fifth line follows from the fact that under Assumption 1, \( E(\beta_\tau^\top X) = P_{B_\tau}(\Sigma_{xx})^\top X \). Finally, the last line follows from replacing the expression for \( P_{B_\tau}(\Sigma_{xx}) \). Specifically,

\[
b_\tau^\top P_{B_\tau}(\Sigma_{xx})^\top X = b_\tau^\top \Sigma_{xx}B_\tau\left(B_\tau^\top \Sigma_{xx}B_\tau\right)^{-1}B_\tau^\top X
\]

\[
= d^aB_\tau^\top X,
\]

where \( d^a = b_\tau^\top \Sigma_{xx}B_\tau(B_\tau^\top \Sigma_{xx}B_\tau)^{-1} \).

**Remark 2** It may be interesting to note that if the distribution of \( X \) is elliptically symmetric, then Assumption 1 is satisfied for every \( \tau \in (0, 1) \). Although the elliptical distribution assumption appears restrictive, among other existing results, Diaconis and Freedman (1984) showed that most low-dimensional projections of high-dimensional data are approximately normal. Moreover, deviations from the elliptical distribution imply predictor transformations or data reweighting to mitigate curvatures; see Cook (1998) and Cook and Nachtsheim (1994).

An important situation is when the objective function is

\[
L(\alpha_\tau + b_\tau^\top X, Q_\tau(Y|X)) = [Q_\tau(Y|X) - a_\tau - b_\tau^\top X]^2
\]

which implies the following minimization problem

\[
(a_\tau^*, b_\tau^*) = \arg \min_{(a_\tau, b_\tau)} E\left[Q_\tau(Y|A^\top X) - a_\tau - b_\tau^\top X\right]^2.
\]

(3)

Theorem 1 and relation (3) imply that the ordinary least squares (OLS) vector \( \beta_\tau^* \), resulting from regressing \( Q_\tau(Y|A^\top X) \) on \( X \), belongs to \( S_{Q_\tau(Y|X)} \), i.e., that \( \beta_\tau^* \) is equal to \( B_\tau \) up to a multiplier. However, the fact that we get \( \beta_\tau^* \) instead of \( B_\tau \) does not matter because the link function in the SIQR model is fully nonparametric.
3.2 Sample level: Algorithm 1

For computational simplicity, we will use the minimization problem (3) and suggest the following estimation procedure. First, use a standard dimension reduction technique to estimate \( \mathbf{A} \) by \( \hat{\mathbf{A}} \) and form the new \( d \times 1 \) predictor vector \( \hat{\mathbf{A}}^\top \mathbf{X} \). In this paper, we use SIR of Li (1991), but one can also use, for instance, the SAVE of Cook and Weisberg (1991) or another technique. Then, we use data \( \{Y_i, \mathbf{X}_i\}_{i=1}^n \) to estimate \( \beta_\tau^* \) by

\[
(\hat{\alpha}_\tau, \hat{\beta}_\tau) = \arg \min_{(\alpha_\tau, \beta_\tau)} \sum_{i=1}^n \left\{ \hat{Q}_\tau(Y|\hat{\mathbf{A}}^\top \mathbf{X}_i) - a_\tau - b_\tau^\top \mathbf{X}_i \right\}^2,
\]

where \( \hat{Q}_\tau(Y|\hat{\mathbf{A}}^\top \mathbf{X}_i) \) is a nonparametric estimate of \( Q_\tau(Y|\hat{\mathbf{A}}^\top \mathbf{X}_i) \). There are many ways to estimate \( Q_\tau(Y|\hat{\mathbf{A}}^\top \mathbf{X}_i) \); we choose the local linear conditional quantile estimation method introduced in Guerre and Sabbah (2012) as it is simple to implement and tends to work well in practice. The idea is to take \( \hat{Q}_\tau(Y|\hat{\mathbf{A}}^\top \mathbf{X}_i) = \hat{q}_\tau(X_i) \), where

\[
(\hat{Q}_\tau(X_i), \hat{s}_\tau(X_i)) = \arg \min_{(q_\tau, s_\tau)} \sum_{k=1}^n \rho_\tau \left\{ Y_k - q_\tau - s_\tau^\top \hat{\mathbf{A}}^\top (\mathbf{X}_k - X_i) \right\} \times K \left( \frac{\hat{\mathbf{A}}^\top (\mathbf{X}_k - X_i)}{h} \right). \tag{5}
\]

Here, \( K(\cdot) \) is a \( d \)-dimensional kernel function and \( h > 0 \) is a bandwidth. In this paper, we use a Gaussian kernel and choose the bandwidth using the rule-of-thumb given in Yu and Jones (1998). Specifically, we select

\[
h = h_m (1 - \tau)/\phi(\Phi^{-1}(\tau))^2 1/5, \tag{6}
\]

where \( \phi(\cdot) \) and \( \Phi(\cdot) \) denote the probability density and cumulative distribution functions of the standard normal distribution, respectively, and \( h_m \) denotes the optimal bandwidth used in mean regression local estimation. We now summarize the algorithm.

Sample Level Algorithm 1: Let \( \{Y_i, \mathbf{X}_i\}_{i=1}^n \) i.i.d. observations and fix \( \tau \in (0, 1) \).

1. Use SIR of Li (1991) or a similar dimension reduction technique to estimate the \( p \times d \) basis matrix \( \mathbf{A} \) of the CS, denoted by \( \hat{\mathbf{A}} \), and form the new sufficient predictors \( \hat{\mathbf{A}}^\top \mathbf{X}_i, i = 1, \ldots, n \).

2. For each \( i = 1, \ldots, n \), use the local linear conditional quantile estimation method of Guerre and Sabbah (2012) to estimate \( Q_\tau(Y|\hat{\mathbf{A}}^\top \mathbf{X}_i) \). Specifically, take \( \hat{Q}_\tau(Y|\hat{\mathbf{A}}^\top \mathbf{X}_i) = \hat{q}_\tau(X_i) \), where \( \hat{q}_\tau(X_i) \) satisfies (5).  

3. Take \( \hat{\beta}_\tau \) to be

\[
(\hat{\alpha}_\tau, \hat{\beta}_\tau) = \arg \min_{(\alpha_\tau, \beta_\tau)} \sum_{i=1}^n \left\{ \hat{Q}_\tau(Y|\hat{\mathbf{A}}^\top \mathbf{X}_i) - a_\tau - b_\tau^\top \mathbf{X}_i \right\}^2.
\]

Then, \( \hat{\beta}_\tau \) defines an estimated basis vector for \( S_{\tau}(Y|\mathbf{X}) \).

The main idea of Algorithm 1 is to choose a good starting point, after which we only need one step to obtain the final estimator. Note that the dimension reduction technique used in Step 1 gives the same \( \hat{\mathbf{A}} \) for all choices of \( \tau \). This is then converted, in Steps 2 and 3, into an estimate of \( \mathbf{B}_\tau \), which now depends on a given value of \( \tau \).

Remark 3 It is sometimes easy to transform \( \mathbf{X} \) linearly and study the relation between \( Y \) and the transformed predictors. In general, if \( \mathbf{Z} = \mathbf{W}^\top \mathbf{X} + \mathbf{b} \) for some invertible matrix \( \mathbf{W} \) and some vector \( \mathbf{b} \), then

\[
S_{\tau}(Y|\mathbf{Z}) = \mathbf{W}^{-1} S_{\tau}(Y|\mathbf{X}).
\]

This can be proved easily by noting that

\[
Q_\tau(Y|\mathbf{X}) = Q_\tau(Y|\mathbf{B}_\tau^\top \mathbf{X}) = Q_\tau(Y|\mathbf{B}_\tau^\top \mathbf{W}^\top \mathbf{Z} - \mathbf{b}) = Q_\tau(Y|\mathbf{W}^{-1} \mathbf{B}_\tau^\top \mathbf{Z}).
\]

Therefore, in practice, we can standardize \( \mathbf{X} \) to have zero mean and the identity covariance matrix. We apply the algorithm to \( \mathbf{Z} = \tilde{\mathbf{X}}_{xx}^{-1/2} (\mathbf{X} - E_\tau(\mathbf{X})) \), where \( \tilde{\mathbf{X}}_{xx} \) and \( E_\tau(\mathbf{X}) \) denote the sample covariance matrix and sample mean of \( \mathbf{X} \), respectively. If \( \tilde{\eta}_\tau \in S_{\tau}(Y|\mathbf{Z}) \), where \( \mathbf{Z} \) is the population version of \( \mathbf{Z} \), then \( \tilde{\mathbf{X}}_{xx}^{-1/2} \tilde{\eta}_\tau \in S_{\tau}(Y|\mathbf{X}) \).

Theorem 4 For a given \( \tau \in (0, 1) \), assume that \( Y \perp Q_\tau(Y|\mathbf{X})|\mathbf{B}_\tau^\top \mathbf{X} \), where \( \mathbf{B}_\tau \) is a \( p \times 1 \) vector. Under Assumption 1, Assumptions A1–A5 given in “Appendix A,” and the assumption that \( \hat{\mathbf{A}} \) is \( \sqrt{n} \)-consistent estimate of the directions of the CS, \( \hat{\beta}_\tau \) is \( \sqrt{n} \)-consistent estimate of the direction of \( S_{\tau}(Y|\mathbf{X}) \), where \( \hat{\beta}_\tau \) is defined in (4).  

Proof See “Appendix B.2.”

4 Estimation of the \( \tau \)-CQS for a MIQR model

4.1 Population level

A MIQR model is an extension of a SIQR model, which assumes that \( Q_\tau(Y|\mathbf{X}) = g_\tau(B_{\tau}^\top \mathbf{X}) \), where \( g_\tau(\cdot): \mathbb{R}^{d_{\tau}} \to \mathbb{R} \) is a \( d_{\tau} \)-dimensional link function, \( 1 \leq d_{\tau} \leq p \), and \( \mathbf{B}_\tau \) is a \( p \times d_{\tau} \) matrix of unknown parameters. The MIQR model implies that \( Y \perp Q_\tau(Y|\mathbf{X})|\mathbf{B}_\tau^\top \mathbf{X} \), and assumes a \( d_{\tau} \)-dimensional \( \tau \)-CQS. The goal is to estimate the space spanned by the column vectors of \( \mathbf{B}_\tau \), which correspond to the vectors of coefficients for the linear combinations \( B_{\tau}^\top \mathbf{X} \). If \( d_{\tau} \) is strictly greater than 1, then the OLS slope vector \( \beta_\tau^* \),
defined in (3), is inconsistent for estimating $S_{Q_{\tau}}(Y|X)$, and therefore, a different approach is necessary to produce more vectors in $S_{Q_{\tau}}(Y|X)$.

**Theorem 5** For a given $\tau \in (0, 1)$, assume that $Y \perp Q_{\tau}(Y|X)|B_{\tau}^T X$, where $B_{\tau}$ is a $p \times d_\tau$ matrix and $d_\tau \leq p$. Then, under Assumption 1, and the assumption that $U_\tau$ is a measurable function of $B_{\tau}^T X$,

$$E[Q_{\tau}(Y|U_\tau)X] \in S_{Q_{\tau}}(Y|X),$$

provided that $Q_{\tau}(Y|U_\tau)X$ is integrable.

**Proof** Observe that

$$E[Q_{\tau}(Y|U_\tau)X] = E[E[Q_{\tau}(Y|U_\tau)X|B_{\tau}^T X]] = E[E[Q_{\tau}(Y|U_\tau)E(X|B_{\tau}^T X)]$$

$$= E[Q_{\tau}(Y|U_\tau)P_{\beta_{\tau}}(\Sigma_{\beta_{\tau}}^{-1} X)] = P_{\beta_{\tau}}(\Sigma_{\beta_{\tau}}^{-1} X)E[Q_{\tau}(Y|U_\tau)X],$$

where the third line follows from Assumption 1.

**Corollary 6** Under the assumptions of Theorem 5, the vector $E[Q_{\tau}(Y|\beta_{\tau}^+ X)X]$ belongs to $S_{Q_{\tau}}(Y|X)$, where $\beta_{\tau}^+$ is defined in (3).

Corollary 6 provides a practical method of forming vectors in the $\tau$-CQS. Let $\beta_{\tau,0}^+$ as defined in (3) and set $\beta_{\tau,0} = \beta_{\tau,0}^+$. Then, for $j = 1, \ldots, n$, we can find more vectors in $S_{Q_{\tau}}(Y|X)$ using

$$\beta_{\tau,j} = E[Q_{\tau}(Y|U_j|\beta_{\tau,j-1} X)X] \in S_{Q_{\tau}}(Y|X),$$

for a function $u_j : \mathbb{R} \to \mathbb{R}$. The question now is how to find an initial vector. Theorem 1 states that $\beta_{\tau,1}$, defined in (3), belongs to $S_{Q_{\tau}}(Y|X)$. Therefore, we set $\beta_{\tau,0} = \beta_{\tau,1}$, and for simplicity, we take $u_j(t) = t$.

**Sample Level Algorithm 2** Let $\{Y_i, X_i\}_{i=1}^n$ i.i.d. observations and fix $\tau \in (0, 1)$.

1. Use Algorithm 1 to compute $\hat{\beta}_\tau$, defined in (4). Set $\hat{\beta}_{\tau,0} = \hat{\beta}_\tau$.
2. If $d_\tau = 1$ stop and report $\hat{\beta}_\tau$ as the estimated basis vector for $S_{Q_{\tau}}(Y|X)$. Otherwise, move to Step 3.
3. Given $j$, where $j = 1, \ldots, p - 1$,

   (a) form the predictors $\hat{\beta}_{\tau,j-1} X_i, i = 1, \ldots, n$, and use the local linear conditional quantile estimation method of Guerre and Sabbah (2012) to estimate $Q_{\tau}(Y|\hat{\beta}_{\tau,j-1} X_i)$.

   Specifically, take

   $$\hat{Q}_{\tau}(Y|\hat{\beta}_{\tau,j-1} X_i) = \hat{q}_\tau(X_i),$$

   where $\hat{q}_\tau(X_i)$ is given in (5), except that we replace $\hat{A}$ by $\hat{\beta}_{\tau,j-1}$. This leads to a univariate kernel function.

   (b) let $\hat{\beta}_{\tau,j} = n^{-1} \sum_{i=1}^n \hat{Q}_{\tau}(Y|\hat{\beta}_{\tau,j-1} X_i) X_i$.

4. Repeat Step 3 for $j = 1, \ldots, p - 1$.
5. Let $\hat{V}_\tau$ be the $p \times p$ matrix with column vectors $\hat{\beta}_{\tau,j}$, $j = 0, 1, \ldots, p - 1$, that is, $\hat{V}_\tau = (\hat{\beta}_{\tau,0}, \ldots, \hat{\beta}_{\tau,p-1})$, and choose the eigenvectors $\hat{v}_{\tau,k}, k = 1, \ldots, d_\tau$, corresponding to the $d_\tau$ largest eigenvalues of $\hat{V}_\tau^T \hat{V}_\tau$. Then,

   $$\hat{B}_\tau = (\hat{v}_{\tau,1}, \ldots, \hat{v}_{\tau,d_\tau})$$

   is an estimated basis matrix for $S_{Q_{\tau}}(Y|X)$.

**Remark 7** As in Algorithm 1, for convenience we can work with the standardized predictor $\hat{Z} = \hat{X}_{XX}^{-1/2} [X - E_X]$. 

**Theorem 8** For a given $\tau \in (0, 1)$, assume that $Y \perp Q_{\tau}(Y|X)|B_{\tau}^T X$, where $B_{\tau}$ is a $p \times d_\tau$ matrix and $d_\tau \leq p$. Under Assumption 1, Assumptions A1-A5 given in “Appendix A”, and the assumption that $\hat{A}$ is $\sqrt{n}$-consistent estimates of the directions of the CS, the column vectors of $\hat{B}_\tau$ are $\sqrt{n}$-consistent estimates of the directions of $S_{Q_{\tau}}(Y|X)$, where $\hat{B}_\tau$ is defined in (7).

**Proof** See “Appendix B.3”.
5 Central subspace for statistical functional $T$

Although the focus of this paper is on the conditional quantile function, the above methodology can be generalized to any statistical functional of interest. Luo et al. (2014) introduced the $T$-central subspace, denoted by $S_{T(Y|X)}$, as the smallest subspace spanned by the column vectors of the $p \times d_T$ matrix $\Gamma$ satisfying $Y \perp T(Y|X)\Gamma^T X$. Luo et al. (2014) proposed an efficient dimension reduction technique for estimating the $T$-central subspace, denoted by $\Gamma$, which now focuses on the statistical functional $T$.

Theorem 9 Assume that $Y \perp T(Y|X)\Gamma^T X$, where $\Gamma$ is a $p \times d_T$ matrix, $d_T \leq p$. Under Assumption 1 and if

$$ (\alpha^*, \gamma^*) = \arg\min_{(\alpha, \gamma)} E\{T(Y|\Gamma^T X) - a - \gamma^T X\}^2, \tag{8} $$

then $\gamma^* \in S_{T(Y|X)}$. Moreover, if $V$ is a measurable function of $\Gamma^T X$, then $E\{T(Y|V)X\} \in S_{T(Y|X)}$, provided that $T(Y|V)X$ is integrable.

Proof Straightforward extension of the proofs of Theorems 1 and 5. \qed

Corollary 10 Under the assumptions of Theorem 9, the vector $E\{T(Y|\gamma^*^T X)|X\}$ belongs to $S_{T(Y|X)}$, where $\gamma^*$ is defined in (8).

Sample Level Algorithm 3 Let $\{Y_i, X_i\}$ i.i.d. observations.

1. Use SIR of Li (1991) or a similar dimension reduction technique to estimate the $p \times d$ basis matrix $\hat{A}$ of the CS, denoted by $A$, and form the new sufficient predictors $\hat{A}^T X_i, i = 1, \ldots, n$.
2. For each $i = 1, \ldots, n$, estimate $T(Y|\hat{A}^T X_i)$ using nonparametric techniques. This step depends on what the function $T$ is.
3. Take $\hat{\gamma}$ to be

$$ (\hat{a}, \hat{\gamma}) = \arg\min_{(a, \gamma)} \sum_{i=1}^{n} \left| \hat{T}(Y|\hat{A}^T X_i) - a - \gamma^T X_i \right|^2. \tag{9} $$

4. If $d_T = 1$ stop and report $\hat{\gamma}$ as the estimated basis vector for $S_{T(Y|X)}$. Otherwise, move to Step 5.
5. Set $\hat{\gamma}_0 = \hat{\gamma}$, where $\hat{\gamma}$ is defined in (9).
6. Given $j$, for $j = 1, \ldots, p - 1$,

(a) form the predictors $\hat{\gamma}_{j-1}^T X_i, i = 1, \ldots, n$, and use nonparametric techniques to estimate $T(Y|\hat{\gamma}_{j-1}^T X_i)$.

7. Repeat Step 6 for $j = 1, \ldots, p - 1$.
8. Let $\hat{G}$ be the $p \times p$ matrix with column vectors $\hat{\gamma}_j, j = 0, 1, \ldots, p - 1$, that is, $\hat{G} = (\hat{\gamma}_0, \ldots, \hat{\gamma}_{p-1})$, and choose the eigenvectors $\hat{g}_k, k = 1, \ldots, d_T$, corresponding to the $d_T$ largest eigenvalues of $\hat{G} \hat{G}^T$. Then, $\hat{\Gamma} = (\hat{g}_1, \ldots, \hat{g}_{d_T})$ is an estimated basis matrix for $S_{T(Y|X)}$.

Remark 11 Note that the dimension reduction technique used in Step 1 focuses on the entire conditional distribution and performs and initial dimension reduction. This is then converted into an estimate of $\Gamma$, which now focuses on the statistical functional $T$.

6 Structural dimension

In all the above, we assume that the dimension of a subspace is known. However, in practice, the true dimension of a subspace is unknown and needs to be estimated. There are several methods proposed for estimating the dimension of a subspace, including a $\chi^2$-sequential test (Li 1991), a cross-validation (CV) criterion (Xia et al. 2002; Wang and Xia 2008), and a Bayesian information criterion (BIC; Zhu et al. 2010).

The construction of the $\chi^2$-sequential test can be challenging, while the CV criterion can be computationally expensive. Therefore, we suggest estimating the dimension of a subspace using the modified BIC-type criterion of Zhu et al. (2010). The major advantage with this method is that the consistency of the estimator of the relevant matrix is enough to guarantee the consistency of the estimator of the dimension.

To generalize for any subspace of interest, and without notational confusion, we write $A$, with a sample version $\hat{A}$, as a $p \times q$ candidate matrix that targets the subspace of interest. Let $q$ be the true dimension of the subspace of interest, and $\hat{q}$ the estimated dimension. The modified BIC-type criterion is defined as

$$ G_n(k) = n \left( \frac{\sum_{i=1}^{k} \hat{\lambda}_i^2}{\sum_{i=1}^{p} \hat{\lambda}_i^2} - C_n \left( \frac{k(k+1)}{2} \right) \right), $$

where $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p$ are the eigenvalues of the matrix $\hat{A}$, $C_n/n \to 0$ as $n \to \infty$ and $C_n \to \infty$. A usual choice for $C_n$ is $2n^{3/4}/p$. Then, $q$ can be estimated by

$$ \hat{q} = \arg\max_{1 \leq k \leq p} G_n(k). $$

In fact, $P(\hat{q} = q) \to 1$, under the assumption that $\hat{A}$ is consistent estimate of $A$. 

\( \odot \) Springer
7 Numerical studies

7.1 Computational remarks

Our proposed methodology consists of two steps. First, we use Algorithm 1 to obtain the initial value $\hat{\beta}_r$ as the OLS slope estimate for the regression of $\hat{Q}_r(Y|\tilde{A}^TX)$ on $X$. Then, if the dimension of the $\tau$-CQS is one, we can stop and report $\hat{\beta}_r$ as the estimated vector of coefficients for the linear combination $B_r^\top X$. Otherwise, if the dimension of the $\tau$-CQS is greater than one, we set the initial vector $\hat{\beta}_{r,0}$ as $\hat{\beta}_r$ and use Algorithm 2 to produce more vectors in $S_{\hat{Q}_r(Y|X)}$. For the first step, the estimation of the basis matrix $A$ of the CS is performed using the existing consistent dimension reduction procedures. For the simulations, we tried different methods (SIR, SAVE, DR, and SIMR), but the results were similar. Here, we report the results from estimating $A$ using SIR of Li (1991), where the number of slices is chosen to be $\max(10, 2p/n)$. For the computation of the conditional quantile estimators $\hat{Q}_r(Y|\tilde{A}^TX)$ and $\hat{Q}_r(Y|\hat{B}_r^\top_{r-1}X)$, used in Algorithms 1 and 2, respectively, we use local linear conditional quantile estimators, which are computed using the function lprq in the R package quantreg. We use a Gaussian kernel and choose the bandwidth according to (6).

The methods that we compare our results with are:

(a) Wu et al. (2010). The idea is to minimize the objective function

$$\sum_{j=1}^n \sum_{i=1}^n \rho_{\tau}(Y_i - a_{r,j} - c_{r,j}b_r^\top(X_i - X_j)) \times \frac{K\left\{ \frac{b_r^\top(X_i - X_j)}{h} \right\}}{\sum_{k=1}^n K\left\{ \frac{b_r^\top(X_i - X_j)}{h} \right\}}$$

in $a_{r,j}, c_{r,j}$, and $b_r$. This is done iteratively by starting with an initial estimate of $b_r$ and using it to minimize the objective function over $a_{r,j}$ and $c_{r,j}$. The corresponding values are then used to minimize the function in $b_r$ to get an updated estimate. The procedure stops when either the absolute difference in two consecutive estimates of $b_r$ is within some threshold value or the number of iterations reaches a fixed value. The initial value of $b_r$ is chosen from average derivative estimate of Chaudhuri et al. (1997). The code that was used was provided by the authors.

(b) Alkenani and Yu (2013). The authors considered variable selection in the context of the SIQR model by including the LASSO and adaptive-LASSO penalties in the objective function of Wu et al. (2010). Specifically, they considered minimizing the objective function

$$\sum_{j=1}^n \sum_{i=1}^n \rho_{\tau}(Y_i - a_{r,j} - c_{r,j}b_r^\top(X_i - X_j)) \times \frac{K\left\{ \frac{b_r^\top(X_i - X_j)}{h} \right\}}{\sum_{k=1}^n K\left\{ \frac{b_r^\top(X_i - X_j)}{h} \right\}}$$

iteratively with respect to $a_{r,j}, c_{r,j}$, and $b_r$. For the simulations, we used the LASSO penalty. The code that was used is a modification of that of Wu et al. (2010).

(c) qMAVE of Kong and Xia (2014). According to the MIQR model, i.e., $Q_r(Y|x) = g_r(B_r^\top x)$, the gradient vector of $Q_r(Y|x)$, denoted by $\nabla Q_r(Y|x)$, satisfies $\nabla Q_r(Y|x) = B_r^\top g_r(B_r^\top x)$. Kong and Xia (2014) defined

$$\Sigma(\tau) = E\left[ \nabla Q_r(Y|X) \nabla Q_r(Y|X)^\top \right]$$

and proved that $S(\Sigma(\tau)) \subseteq S_{\hat{Q}_r(Y|X)}$. In this case, we use

$$\tilde{\Sigma}(\tau) = n^{-1} \sum_{i=1}^n \nabla Q_r(Y|X_i) \nabla Q_r(Y|X_i)^\top,$$

where $\nabla Q_r(Y|X_i)$ denotes an estimate of the gradient vector $\nabla Q_r(Y|X_i)$. Kong and Xia (2014) used the local polynomial estimator of Chaudhuri (1991) and Kong et al. (2010). The code that was used is available on the authors’ websites.

(d) Luo et al. (2014). The estimation procedure is a four-step procedure. To avoid getting into details and interrupting the flow of the paper, we mention that the method relies on the estimation of two important quantities: the efficient score and the efficient information. Those have explicit formulas for different statistical functionals of interest. Moreover, an initial value of the parametric component is chosen using MAVE of Xia et al. (2002). The code that was used was provided by the authors.

For the estimation accuracy, we use the angle between the two subspaces $\tilde{B}_r$ and $B_r$, where $\tilde{B}_r$ denotes an estimate of the $\tau$-CQS with a basis matrix $B_r$. This is used as the measure of the distance between two subspaces so that smaller number implies stronger correlation. The angle is measured in radians, and so we report the value divided by $\pi/2$ and is calculated using the function subspace in the R package pracma. We call this the estimation error. We note that we have also tried measuring the distance between two subspaces using the measure proposed by Li et al. (2005), as well...
as the vector correlation coefficient and the trace correlation coefficient, but the results exhibit similar patterns.

All simulation results are based on \( N = 100 \) iterations. Unless otherwise stated, the sample size is chosen to be \( n = 600 \), and the quantiles under consideration are \( \tau = 0.1, 0.25, 0.5, 0.75, \) and \( 0.9 \).

### 7.2 Simulation results

**Example 1** We demonstrate the performance of Algorithm 1, where \( d_t = 1 \), and use \( \beta_0 \), defined in (4), as the estimated basis vector for \( S_{Q_t}(Y|X) \).

(a) We begin by considering the performance of \( \hat{\beta}_e \) for different choices of \( n \) and \( p \). The data are generated according to the following SIQR model

\[
Y = 3X_1 + X_2 + \varepsilon,
\]

where \( X = (X_1, \ldots, X_p)^\top \) and the error \( \varepsilon \) are generated according to a standard normal distribution. The sample size is given by \( n = 200, 400, \) or \( 600 \), and the number of predictors is \( p = 10, 20, \) or \( 40 \). The \( \tau \)-CQS is spanned by \((3, 1, 0, \ldots, 0)^\top\), for \( \tau = 0.1, 0.25, 0.5, 0.75, \) and \( 0.9 \). The results are given in Table 1. We can observe that the mean estimation error increases with \( p \) and decreases with \( n \). Moreover, we observe that the performance of \( \hat{\beta}_e \) is robust to the specific quantile. This contradicts the performance of the Luo et al. (2014) method, where the authors observed that the median CQS performed better than the upper 0.75-CQS (see example (g), Luo et al. 2014).

(b) We now compare the performance of the proposed methodology with that of the iterative methods of Wu et al. (2010) and Alkenani and Yu (2013). The data are generated according to a standard normal distribution. The sample size is chosen to be \( n = 600 \), and the quantiles under consideration are \( \tau = 0.1, 0.25, 0.5, 0.75, \) and \( 0.9 \).

(c) We investigate the performance of the proposed methodology and the existing methods using an \( X \) with dependent components. The data are generated according to Models I–V, where \( X = (X_1, \ldots, X_{10})^\top \sim \mathcal{N}(0, \sigma_{ij}1_{|i-j| \leq 10}) \) with \( \sigma_{ij} = 0.5^{|i-j| \}, \) and the error \( \varepsilon \) is generated according to a standard normal distribution. To save space, and since the results follow similar patterns, we only report the results for Model I. From Table 4, we observe that the errors are larger than those for \( X \) with independent components. Further investigation will be considered later—see Example 2 (c).

(d) We now evaluate the performance of the modified BIC-type criterion, defined in Sect. 6. The data are generated according to Models I–V, with the predictors and error having a standard normal distribution. We consider the structural dimension of the CS, \( d \), to be unknown, and we estimate it using the modified BIC-type criterion. We apply Algorithm 1 using the \( p \times d \) matrix \( \hat{A} \). From Table 5, we observe that the mean estimation errors are very similar to the ones in Yu (2013) are iterative and hence have convergence issues. Both algorithms stop when either convergence is achieved or when the number of iterations reaches a fixed value. For Model V, both methods have convergence issues which result in unlikely outcomes distorting the estimation errors.

Another important limitation of iterative methods is that they are computationally expensive. To see this in practice, we compare the running times of the proposed methodology, with those of Wu et al. (2010) and Alkenani and Yu (2013) methods. Table 3 gives the computation time (in seconds) for 100 iterations of Models I–V for each method and each quantile level. We see that the proposed methodology is significantly faster than the other two methods. The computation time of the proposed method for 100 iterations lies between 121 and 133 seconds, which gives an average time for one run between 2.02 and 2.22 seconds. However, the computation time of the iterative methods of Wu et al. (2010) and Alkenani and Yu (2013) varies for the different models and the different quantile levels. This is because sometimes the algorithm takes longer as convergence is not reached before the maximum number of iterations, while there are other instances where convergence is reached and thus the algorithm requires less time to complete. Note that the computation times of the iterative methods of Wu et al. (2010) and Alkenani and Yu (2013) for Models IV and V are significantly higher than for the other three models since, as was mentioned in the previous paragraph, the algorithms have convergence issues and therefore take longer to complete. The runs were carried out on a Dell Poweredge 820 with 256 GB of memory, Intel(R) Xeon(R) CPU E5-4620, and running Ubuntu 18.04 and R version 3.4.4. To exclude any effect from a particular machine, the runs were repeated on two other machines. However, the results were similar and are not reported here.

(e) We investigate the performance of the proposed methodology and the existing methods using an \( X \) with dependent components. The data are generated according to Models I–V, where \( X = (X_1, \ldots, X_{10})^\top \sim \mathcal{N}(0, \sigma_{ij}1_{|i-j| \leq 10}) \) with \( \sigma_{ij} = 0.5^{|i-j| \}, \) and the error \( \varepsilon \) is generated according to a standard normal distribution. To save space, and since the results follow similar patterns, we only report the results for Model I. From Table 4, we observe that the errors are larger than those for \( X \) with independent components. Further investigation will be considered later—see Example 2 (c).

(f) We now evaluate the performance of the modified BIC-type criterion, defined in Sect. 6. The data are generated according to Models I–V, with the predictors and error having a standard normal distribution. We consider the structural dimension of the CS, \( d \), to be unknown, and we estimate it using the modified BIC-type criterion. We apply Algorithm 1 using the \( p \times d \) matrix \( \hat{A} \). From Table 5, we observe that the mean estimation errors are very similar to the ones in

\[ \text{ Springer} \]
Table 1 Mean (and SD) of the estimation errors for $\hat{\beta}_\tau$, $\tau = 0.1, 0.25, 0.5, 0.75,$ and 0.9, for Example 1 (a)

| Model | Methodology | 0.1    | 0.25   | 0.5    | 0.75   | 0.9    |
|-------|-------------|--------|--------|--------|--------|--------|
| I     | Proposed    | 0.050  | 0.051  | 0.051  | 0.051  | 0.052  |
|       | Wu et al. (2010) | 0.056 | 0.056 | 0.061 | 0.059 | 0.054 |
|       | Alkenani and Yu (2013) | 0.052 | 0.053 | 0.055 | 0.053 | 0.051 |
| II    | Proposed    | 0.051  | 0.050  | 0.059  | 0.055  | 0.050  |
|       | Wu et al. (2010) | 0.056 | 0.056 | 0.058 | 0.058 | 0.055 |
|       | Alkenani and Yu (2013) | 0.052 | 0.054 | 0.055 | 0.054 | 0.052 |
| III   | Proposed    | 0.045  | 0.049  | 0.041  | 0.044  | 0.050  |
|       | Wu et al. (2010) | 0.052 | 0.045 | 0.041 | 0.045 | 0.050 |
|       | Alkenani and Yu (2013) | 0.056 | 0.043 | 0.038 | 0.043 | 0.054 |
| IV    | Proposed    | 0.072  | 0.089  | 0.088  | 0.095  | 0.077  |
|       | Wu et al. (2010) | 0.134 | 0.113 | 0.092 | 0.093 | 0.085 |
|       | Alkenani and Yu (2013) | 0.072 | 0.085 | 0.081 | 0.092 | 0.074 |
| V     | Proposed    | 0.343  | 0.383  | 0.403  | 0.386  | 0.336  |
|       | Wu et al. (2010) | 0.532 | 0.512 | 0.648 | 0.541 | 0.534 |
|       | Alkenani and Yu (2013) | 0.558 | 0.526 | 0.654 | 0.544 | 0.551 |

For each $\tau$, the value with the smallest mean estimation error is bolded. In the case of ties, all relevant values are bolded.

Table 2 Mean (and SD) of the estimation errors for $\hat{\beta}_\tau$, $\tau = 0.1, 0.25, 0.5, 0.75,$ and 0.9, using the proposed methodology and the Wu et al. (2010) and Alkenani and Yu (2013) estimators, for Example 1 (b)

| Model | Methodology | 0.1    | 0.25   | 0.5    | 0.75   | 0.9    |
|-------|-------------|--------|--------|--------|--------|--------|
| I     | Proposed    | 0.051  | 0.051  | 0.051  | 0.051  | 0.051  |
|       | Wu et al. (2010) | 0.051 | 0.051 | 0.051 | 0.051 | 0.051 |
|       | Alkenani and Yu (2013) | 0.051 | 0.051 | 0.051 | 0.051 | 0.051 |
| II    | Proposed    | 0.050  | 0.050  | 0.050  | 0.050  | 0.050  |
|       | Wu et al. (2010) | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 |
|       | Alkenani and Yu (2013) | 0.050 | 0.050 | 0.050 | 0.050 | 0.050 |
| III   | Proposed    | 0.049  | 0.049  | 0.049  | 0.049  | 0.049  |
|       | Wu et al. (2010) | 0.049 | 0.049 | 0.049 | 0.049 | 0.049 |
|       | Alkenani and Yu (2013) | 0.049 | 0.049 | 0.049 | 0.049 | 0.049 |
| IV    | Proposed    | 0.088  | 0.088  | 0.088  | 0.088  | 0.088  |
|       | Wu et al. (2010) | 0.113 | 0.113 | 0.113 | 0.113 | 0.113 |
|       | Alkenani and Yu (2013) | 0.085 | 0.085 | 0.085 | 0.085 | 0.085 |
| V     | Proposed    | 0.403  | 0.403  | 0.403  | 0.403  | 0.403  |
|       | Wu et al. (2010) | 0.512 | 0.512 | 0.512 | 0.512 | 0.512 |
|       | Alkenani and Yu (2013) | 0.526 | 0.526 | 0.526 | 0.526 | 0.526 |

Table 3 Computation time (in s) for estimating $\beta_\tau$, $\tau = 0.1, 0.25, 0.5, 0.75,$ and 0.9, for 100 simulation runs, using the proposed methodology and the Wu et al. (2010) and Alkenani and Yu (2013) estimators, for Example 1 (b)

| Model | Methodology | 0.1    | 0.25   | 0.5    | 0.75   | 0.9    |
|-------|-------------|--------|--------|--------|--------|--------|
| I     | Proposed    | 130    | 130    | 131    | 130    | 129    |
|       | Wu et al. (2010) | 3144  | 3984  | 4283  | 3863  | 3121  |
|       | Alkenani and Yu (2013) | 5426 | 6856 | 5993 | 6671 | 5499 |
| II    | Proposed    | 129    | 130    | 132    | 133    | 131    |
|       | Wu et al. (2010) | 3527  | 4020  | 4657  | 4145  | 3302  |
|       | Alkenani and Yu (2013) | 5734 | 7696 | 7406 | 7387 | 5823 |
| III   | Proposed    | 130    | 131    | 132    | 131    | 129    |
|       | Wu et al. (2010) | 2480  | 2783  | 3113  | 2646  | 2278  |
|       | Alkenani and Yu (2013) | 4166 | 4266 | 4358 | 4578 | 4009 |
| IV    | Proposed    | 126    | 129    | 132    | 131    | 129    |
|       | Wu et al. (2010) | 6964  | 6191  | 5159  | 5291  | 4410  |
|       | Alkenani and Yu (2013) | 10699 | 9178 | 5223 | 9153 | 7762 |
| V     | Proposed    | 122    | 122    | 122    | 122    | 121    |
|       | Wu et al. (2010) | 9626  | 11603 | 11307 | 11353 | 9519  |
|       | Alkenani and Yu (2013) | 16172 | 16557 | 15827 | 19641 | 16753 |

For each $\tau$, the value with the smallest computation time is bolded.
is due to the sample size is given by \( n \) generated according to a standard normal distribution. The predictors and the error are generated according to a statement in Theorem 4. We reconsider Model II, where the relationship between the mean estimation error and 1/\( \sqrt{n} \), demonstrating the \( \sqrt{n} \)-consistency of the proposed estimator \( \widehat{\beta} \).

Example 2 We demonstrate the performance of Algorithm 2, where \( d_r > 1 \), and use \( \widehat{B}_r \), defined in (7), as the estimated basis matrix for \( S_{Q_r}(Y|X) \).

(a) We begin by considering the performance of \( \widehat{B}_r \) for different choices of \( n \) and \( p \). The data are generated according to the following MIQR model

\[
Y = X_1^3 + X_2 + \varepsilon,
\]

where \( X = (X_1, \ldots, X_p)^T \) and the residual \( \varepsilon \) are generated according to a standard normal distribution. The sample size is given by \( n = 200, 400, \ldots, 1000 \). Figure 1 shows the observed mean values for the estimation error for the different quantile levels. The plots indicate an approximate linear relationship between the mean estimation error and 1/\( \sqrt{n} \), demonstrating the \( \sqrt{n} \)-consistency of the proposed estimator \( \widehat{\beta} \).

Table 4 Mean (and SD) of the estimation errors for \( \widehat{\beta}_r \), \( \tau = 0.1, 0.25, 0.5, 0.75, \) and 0.9, using the proposed methodology and the Wu et al. (2010) and Alkenani and Yu (2013) estimators, for Example 1 (c)

| Model | Methodology | 0.1   | 0.25  | 0.5   | 0.75  | 0.9   |
|-------|-------------|-------|-------|-------|-------|-------|
| I     | Proposed    | **0.102** (0.041) | **0.113** (0.043) | 0.146 (0.038) | **0.109** (0.039) | 0.108 (0.041) |
| Wu et al. (2010) | 0.111 (0.025) | 0.118 (0.035) | 0.149 (0.044) | 0.118 (0.036) | 0.110 (0.036) |
| Alkenani and Yu (2013) | 0.105 (0.043) | 0.114 (0.047) | **0.136** (0.059) | 0.111 (0.043) | **0.102** (0.041) |

For each \( \tau \), the value with the smallest mean estimation error is bolded.

Table 2, suggesting that the dimension of the CS has been consistently estimated using the BIC-type criterion.

(e) As was already observed in Example 1 (a), the finite sample performance of \( \widehat{\beta}_r \) improves as \( n \) increases. This is due to the \( \sqrt{n} \)-consistency of the proposed estimator, stated in Theorem 4. We reconsider Model II, where the predictors and the error are generated according to a standard normal distribution. The sample size is taken to be \( n = 200, 400, \ldots, 1000 \). Figure 1 shows the observed mean values for the estimation error for the different quantile levels. The plots indicate an approximate linear relationship between the mean estimation error and 1/\( \sqrt{n} \), demonstrating the \( \sqrt{n} \)-consistency of the proposed estimator \( \widehat{\beta} \).

Example 2 We demonstrate the performance of Algorithm 2, where \( d_r > 1 \), and use \( \widehat{B}_r \), defined in (7), as the estimated basis matrix for \( S_{Q_r}(Y|X) \).

(a) We begin by considering the performance of \( \widehat{B}_r \) for different choices of \( n \) and \( p \). The data are generated according to the following MIQR model

\[
Y = X_1^3 + X_2 + \varepsilon,
\]

where \( X = (X_1, \ldots, X_p)^T \) and the residual \( \varepsilon \) are generated according to a standard normal distribution. The sample size is given by \( n = 200, 400, \ldots, 600 \), and the number of predictors is \( p = 10, 20, \) or 40. The \( \tau \)-CQS is spanned by \( \{(1, 0, \ldots, 0)^T, (0, 1, 0, \ldots, 0)^T\} \), for \( \tau = 0.1, 0.25, 0.5, 0.75, \) and 0.9. The results are given in Table 6. We observe that the mean estimation error increases with \( p \) and decreases with \( n \).

(b) We now compare the performance of the proposed estimator with the qMAVE procedure of Kong and Xia (2014) and the efficient estimator of Luo et al. (2014). The data are generated according to the following MIQR models

Model VI: \( Y = 0.5X_1 + 0.8X_2 + \sqrt{0.5 + 0.3X_1^2 + 0.6X_2^2} \varepsilon, \)

Model VII: \( Y = \frac{X_1}{0.5 + (X_2 + 1.5)^2} + X_3\varepsilon, \)

Model VIII: \( Y = X_1^3 + \exp(X_2) + X_3\varepsilon, \)

Model IX: \( Y = \frac{X_1}{0.5 + (X_2 + 1.5)^2} + 0.5\exp(X_1)\varepsilon, \)

where \( X = (X_1, \ldots, X_{10})^T \) and the residual \( \varepsilon \) are generated according to a standard normal distribution. Table 7 demonstrates the mean and standard deviation of the estimation errors of the proposed estimator and of the Kong and Xia (2014) and Luo et al. (2014) estimators. We observe that the performance of the proposed methodology is comparable with that of the qMAVE procedure of Kong and Xia (2014), while both methods outperform Luo et al. (2014) estimator.

(c) Here, we investigate the performance of the proposed methodology and the existing methods using an \( X \) with dependent components. The data are generated according to Model VI, where \( X = (X_1, \ldots, X_{10})^T \sim N((0, \sigma_{ij})_{1 \leq i,j \leq 10}) \) with \( \sigma_{ij} = 0.5^{|i-j|} \), and the residual \( \varepsilon \) is generated according to a standard normal distribution. From Table 7, we observe that the errors are larger than those for \( X \) with independent components, but the degree by which the mean estimation error of the proposed methodology and of the qMAVE procedure improves upon the Luo et al. (2014) method is similar to those for the independent component case. However, we observe that the proposed methodology performs better than the qMAVE procedure when \( X \) has dependent components.

(d) Finally, we demonstrate the \( \sqrt{n} \)-consistency of \( \widehat{B}_r \), stated in Theorem 8. We reconsider Model VI, where the predictors and the error are generated according to a standard normal distribution. Figure 2 shows the observed mean values for the estimation error for the different quantile levels. As before, the plots indicate an approximate linear relationship between the mean estimation error and 1/\( \sqrt{n} \), demonstrating the \( \sqrt{n} \)-consistency of the proposed estimator \( \widehat{B}_r \).

Example 3 Although the focus of this paper is on the estimation of the \( \tau \)-CQS, we present one example regarding another statistical functional of interest, the CMS. Further investigation can be considered in a future paper. The data are generated according to the model \( Y = X_1^3 + X_2\varepsilon \), where the predictors \( X = (X_1, \ldots, X_{10})^T \) and the residual \( \varepsilon \) are generated according to a standard normal distribution. Observe that the CMS is spanned by \( \{(1, 0, \ldots, 0)^T, (0, 1, 0, \ldots, 0)^T\} \). The sample size is taken to be \( n = 600 \). Since the CMS is one-

 Springer
Table 5  Mean (and SD) of the estimation errors for $\hat{\beta}_\tau$, $\tau = 0.1, 0.25, 0.5, 0.75, \text{ and } 0.9$, when $d$ is estimated using the BIC-type criterion, for Example 1 (d)

| Model | 01  | 0.25 | 0.5  | 0.75 | 0.9  |
|-------|-----|------|------|------|------|
| I     | 0.064 (0.020) | 0.069 (0.023) | 0.060 (0.018) | 0.066 (0.019) | 0.067 (0.016) |
| II    | 0.061 (0.019) | 0.060 (0.021) | 0.069 (0.016) | 0.066 (0.023) | 0.055 (0.018) |
| III   | 0.051 (0.014) | 0.055 (0.014) | 0.055 (0.012) | 0.054 (0.013) | 0.051 (0.015) |
| IV    | 0.093 (0.024) | 0.106 (0.028) | 0.093 (0.022) | 0.095 (0.024) | 0.082 (0.020) |
| V     | 0.363 (0.215) | 0.406 (0.244) | 0.377 (0.235) | 0.396 (0.241) | 0.350 (0.238) |

Fig. 1  The $\sqrt{n}$-consistency of the proposed estimator $\hat{\beta}_\tau$ for Example 1 (e)

Table 6  Mean (and SD) of the estimation errors for $\hat{B}_\tau$, $\tau = 0.1, 0.25, 0.5, 0.75, \text{ and } 0.9$, using the proposed methodology, for Example 2 (a)

| $n$  | $p$  | 0.1  | 0.25 | 0.5  | 0.75 | 0.9  |
|------|------|------|------|------|------|------|
| 200  | 10   | 0.112 (0.034) | 0.101 (0.031) | 0.098 (0.030) | 0.102 (0.030) | 0.112 (0.037) |
|      | 20   | 0.217 (0.054) | 0.189 (0.042) | 0.174 (0.040) | 0.188 (0.044) | 0.208 (0.049) |
|      | 40   | 0.345 (0.079) | 0.309 (0.067) | 0.293 (0.066) | 0.306 (0.064) | 0.327 (0.068) |
| 400  | 10   | 0.082 (0.024) | 0.076 (0.023) | 0.072 (0.019) | 0.074 (0.022) | 0.082 (0.025) |
|      | 20   | 0.170 (0.038) | 0.151 (0.035) | 0.137 (0.029) | 0.145 (0.032) | 0.170 (0.037) |
|      | 40   | 0.292 (0.055) | 0.244 (0.056) | 0.214 (0.048) | 0.233 (0.053) | 0.278 (0.057) |
| 600  | 10   | 0.072 (0.019) | 0.068 (0.018) | 0.064 (0.017) | 0.066 (0.019) | 0.070 (0.022) |
|      | 20   | 0.150 (0.032) | 0.129 (0.032) | 0.122 (0.030) | 0.129 (0.030) | 0.146 (0.033) |
|      | 40   | 0.260 (0.049) | 0.216 (0.048) | 0.196 (0.041) | 0.219 (0.045) | 0.270 (0.056) |
Table 7  Mean (and SD) of the estimation errors for $\hat{\mathbf{B}}_\tau$, $\tau = 0.1, 0.25, 0.5, 0.75,$ and 0.9, using the proposed methodology and the Kong and Xia (2014) and Luo et al. (2014) estimators, for Example 2 (b)

| Model | Methodology | $0.1$ | $0.25$ | $0.5$ | $0.75$ | $0.9$ |
|-------|-------------|-------|-------|-------|-------|-------|
| VI    | Proposed    | 0.104 (0.029) | 0.092 (0.027) | 0.089 (0.025) | 0.090 (0.025) | 0.098 (0.026) |
|       | Kong and Xia (2014) | 0.119 (0.032) | 0.097 (0.026) | 0.090 (0.023) | 0.092 (0.029) | 0.118 (0.040) |
|       | Luo et al. (2014) | 0.519 (0.156) | 0.534 (0.125) | 0.576 (0.144) | 0.599 (0.148) | 0.590 (0.153) |
| VII   | Proposed    | 0.055 (0.020) | 0.069 (0.026) | 0.105 (0.035) | 0.062 (0.017) | 0.053 (0.018) |
|       | Kong and Xia (2014) | 0.064 (0.019) | 0.060 (0.019) | 0.101 (0.031) | 0.060 (0.019) | 0.060 (0.020) |
|       | Luo et al. (2014) | 0.364 (0.139) | 0.378 (0.127) | 0.633 (0.157) | 0.412 (0.129) | 0.423 (0.142) |
| VIII  | Proposed    | 0.045 (0.016) | 0.041 (0.016) | 0.043 (0.015) | 0.041 (0.015) | 0.045 (0.016) |
|       | Kong and Xia (2014) | 0.038 (0.013) | 0.036 (0.012) | 0.047 (0.012) | 0.037 (0.013) | 0.043 (0.015) |
|       | Luo et al. (2014) | 0.158 (0.064) | 0.229 (0.095) | 0.202 (0.063) | 0.366 (0.149) | 0.408 (0.124) |
| IX    | Proposed    | 0.081 (0.028) | 0.101 (0.033) | 0.109 (0.027) | 0.107 (0.026) | 0.110 (0.029) |
|       | Kong and Xia (2014) | 0.104 (0.029) | 0.087 (0.025) | 0.080 (0.022) | 0.088 (0.028) | 0.121 (0.033) |
|       | Luo et al. (2014) | 0.564 (0.145) | 0.604 (0.138) | 0.612 (0.166) | 0.632 (0.125) | 0.639 (0.136) |
| X     | Proposed    | 0.056 (0.020) | 0.061 (0.021) | 0.091 (0.026) | 0.061 (0.019) | 0.058 (0.024) |
|       | Kong and Xia (2014) | 0.062 (0.018) | 0.065 (0.020) | 0.080 (0.023) | 0.049 (0.016) | 0.052 (0.018) |
|       | Luo et al. (2014) | 0.247 (0.109) | 0.374 (0.139) | 0.451 (0.155) | 0.398 (0.125) | 0.401 (0.124) |

For each $\tau$, the value with the smallest mean estimation error is bolded.

Table 8  Mean (and SD) of the estimation errors for $\hat{\mathbf{R}}_\tau$, $\tau = 0.1, 0.25, 0.5, 0.75,$ and 0.9, using the proposed methodology and the Kong and Xia (2014) and Luo et al. (2014) estimators, for Example 2 (c)

| Model | Methodology | $0.1$ | $0.25$ | $0.5$ | $0.75$ | $0.9$ |
|-------|-------------|-------|-------|-------|-------|-------|
| VI    | Proposed    | 0.251 (0.082) | 0.230 (0.071) | 0.221 (0.066) | 0.230 (0.063) | 0.264 (0.087) |
|       | Kong and Xia (2014) | 0.351 (0.106) | 0.293 (0.098) | 0.272 (0.084) | 0.284 (0.087) | 0.341 (0.101) |
|       | Luo et al. (2014) | 0.593 (0.161) | 0.588 (0.140) | 0.595 (0.139) | 0.595 (0.134) | 0.588 (0.145) |

For each $\tau$, the value with the smallest mean estimation error is bolded.

dimensional, we can estimate the coefficients of the linear combination of the predictors using the OLS slope estimate from the regression of $E(Y|\mathbf{A}^\top\mathbf{X})$ on $\mathbf{X}$; see (8). The conditional mean $E(Y|\mathbf{A}^\top\mathbf{X})$ is estimated nonparametrically using the Nadaraya–Watson estimator with a $d$-dimensional Gaussian kernel function. Specifically, for the data $\{Y_i, \mathbf{X}_i\}_{i=1}^n$, we use

$$\hat{E}(Y|\mathbf{A}^\top\mathbf{X}_i) = \sum_{j=1}^n \frac{Y_j K \left[ \frac{\mathbf{A}^\top(\mathbf{X}_j - \mathbf{X}_i)}{h} \right]}{\sum_{k=1}^n K \left[ \frac{\mathbf{A}^\top(\mathbf{X}_k - \mathbf{X}_i)}{h} \right]},$$

for a $d$-dimensional kernel function $K(\cdot)$ and a bandwidth $h > 0$. The mean and standard deviation of the estimation error are 0.0679 and 0.0182, respectively. For comparison purposes, we also estimate the CMS using the methods of Cook and Li (2002) and Luo et al. (2014) and report the mean and standard deviation of the estimation errors; 0.0707 and 0.0116 for the Cook and Li (2002) method and 0.0977 and 0.0216 for the Luo et al. (2014) method. Observe that the proposed methodology outperforms both methods.

### 7.3 Boston housing data

To illustrate the performance of the proposed methodology, we consider the Boston housing data. The data contain 506 observations on 14 variables. The dependent variable of interest is medv, the median value of owner-occupied homes in $1000s$, and the other thirteen variables are statistical measurements on the 506 census tracts in suburban Boston from the 1970 census. The data were originally published by Harrison and Rubinfeld (1978) and can be found in the MASS library in R.

Due to the collinearity in the data set, Breiman and Friedman (1985) applied their alternating conditional expectation (ACE) method for selecting the relevant variables and selected the four covariates: average number of rooms per house in the area (RM), full-value property tax (in dollar) per $10,000 (TAX), pupil–teacher ratio by town (PREDITO), and percentage of the population having lower economic status in the area (LSTAT). After that, many QR studies (Gooijer and Zerom 2003; Yu and Lu 2004; Wu et al. 2010; Jiang et al. 2013; Christou and Akritas 2016) used this data set, and by taking a logarithmic transformation on the covari-
Fig. 2 The $\sqrt{n}$-consistency of the proposed estimator $\hat{B}_\tau$ for Example 2 (d)

ates TAX and LSTAT, found potential relationship between the conditional quantile of the response medv and these four covariates.

The nature of the response variable makes QR suitable for the analysis of this data set. Specifically, the response variable is the median price of homes and the y-values larger than or equal to $50,000 have been recorded as $50,000. As was noted in Chaudhuri et al. (1997, p. 724), “such a truncation in the upper tail of the response makes quantile regression, which is not influenced very much by extreme values of the response, a very appropriate methodology.”

In this work, we apply the proposed methodology to estimate the $\tau$-CQS for five different quantile levels $\tau = 0.1, 0.25, 0.5, 0.75,$ and $0.9$. The scatterplot matrix of the four predictors does not indicate any serious departures from ellipticity. Also, the BIC-type criterion suggested that $\hat{d}_\tau = 1$, for all $\tau$ under consideration. The estimated vectors and their bootstrapped standard errors for the $\tau$-CQS are demonstrated in Table 9.

Several observations can be made. First, RM seems to have an opposite effect than that of the other predictor variables for all quantile levels. Second, while the effects of PTRATIO and log(LSTAT) seem to be stable across different quantiles, the effects of RM and log(TAX) vary quite much for different quantiles. Specifically, the effect of RM seems to increase from lower to upper quantile levels, while the effect of log(TAX) seems to decrease from lower to upper quantile levels (comparing the absolute values of the coefficients). This can be further observed from Fig. 3, where we plot the estimated vectors for $\tau$ ranging from 0.1 to 0.9 with 0.005 increments.

To compare the proposed methodology with that of Kong and Xia (2014) and Luo et al. (2014), we use the bootstrapped error measurement, introduced in Ye and Weiss (2003), which is reasonable because all estimators involved are consistent. According to the authors, it is not always straightforward to choose between dimension reduction methods by plotting the response against the corresponding estimated linear combinations. Instead, they proposed choosing the dimension reduction method that produces an estimated subspace with the smallest variability. To do that, for each quantile level, we generate 500 bootstrap samples of size 100, and for each sample we compute the proposed estimate $\hat{B}_\tau$, the Kong and Xia (2014) estimate $\hat{B}_\tau^{MAVE}$, and
Table 9 The estimated vectors (and SE) for the $\tau$-CQS, for $\tau = 0.1, 0.25, 0.5, 0.75,$ and $0.9$ for the Boston housing data

| Direction | RM        | log(TAX)   | PTRATIO   | log(LSTAT) |
|-----------|-----------|------------|-----------|------------|
| $\hat{\beta}_{0.1}$ | 0.304 (0.192) | −0.239 (0.227) | −0.099 (0.032) | −1.126 (0.225) |
| $\hat{\beta}_{0.25}$ | 0.338 (0.194) | −0.214 (0.220) | −0.100 (0.031) | −1.103 (0.230) |
| $\hat{\beta}_{0.5}$ | 0.381 (0.192) | −0.151 (0.220) | −0.102 (0.032) | −1.080 (0.231) |
| $\hat{\beta}_{0.75}$ | 0.406 (0.202) | −0.117 (0.219) | −0.102 (0.034) | −1.071 (0.241) |
| $\hat{\beta}_{0.9}$ | 0.407 (0.211) | −0.095 (0.220) | −0.105 (0.035) | −1.071 (0.246) |

Fig. 3 The coefficients of the estimated vector $\hat{\beta}_\tau$ for each predictor variable, for $\tau$ ranging from 0.1 to 0.9 with 0.005 increments for the Boston housing data

Table 10 Comparison between proposed methodology and that of Kong and Xia (2014) and Luo et al. (2014) estimators, using the bootstrapped error measurement for the Boston housing data

| Methodology | 0.1 | 0.25 | 0.5 | 0.75 | 0.9 |
|-------------|-----|------|-----|------|-----|
| Proposed    | 0.074 | 0.068 | 0.070 | 0.079 | 0.099 |
| Kong and Xia (2014) | 0.094 | 0.094 | 0.094 | 0.096 | 0.101 |
| Luo et al. (2014) | 0.445 | 0.454 | 0.445 | 0.452 | 0.529 |

For each $\tau$, the value with the smallest bootstrapped error measurement is bolded.

7.4 Bitcoin data

Bitcoin dates back to the 2008 financial crisis and was introduced by Nakamoto (2008). Since then, Bitcoin has become the most commonly known and used cryptocurrency and is characterized by great volatility (Grocer 2013). In this work, we apply the proposed methodology to the Bitcoin data set and compare its performance with the existing methods. The data were downloaded from https://finance.yahoo.com/ and consist of the daily returns from January 2016 to December 2018. Here, we write “returns” to refer to log returns, i.e., the difference of the log of the returns.

Kong and Xia (2014) considered the daily return $Y$ of a portfolio and used the returns and the absolute returns of the portfolio in the past five days, as well as the market returns and their absolute values of the past five days as the predictor variables. We followed this setup and considered $Y$ as the returns and $X = (X_1, \ldots, X_{10})^T$, such that $X_1, \ldots, X_5$ are the returns in the past five days, and $X_6, \ldots, X_{10}$ are the absolute values of those returns. These are a proxy of the past volatilities.

We apply the proposed methodology to estimate the $\tau$-CQS for five different quantile levels $\tau = 0.1, 0.25, 0.5, 0.75,$ and $0.9$. The BIC-type criterion suggested that $\hat{d}_\tau = 1$, for all $\tau$ under consideration. The estimated vectors and their bootstrapped standard errors for the $\tau$-CQS are demonstrated in Table 11. Note that the effect of each predictor variable varies significantly for different quantile levels. Moreover, the absolute values of the past five returns, $X_6, \ldots, X_{10}$, which are proxies for the past volatilities, seem to be the most important predictor variables for all quantile levels. This confirms the general observation that “simple nonlinear functions of returns, such as absolute or squared returns, exhibit significant persistence” (Cont 2001, p. 230).
Table 11 The estimated vectors (and SE) for the $\tau$-CQS, for $\tau = 0.1, 0.25, 0.5, 0.75, \text{and} 0.9$ for the Bitcoin data

| Direction | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---------|
| $\hat{\beta}_{0.1}$ | 0.070 | 0.025 | -0.083 | 0.031 | -0.049 | -0.553 | -0.359 | -0.455 | -0.413 | -0.416 |
| (0.244) | (0.224) | (0.236) | (0.253) | (0.246) | (0.347) | (0.333) | (0.333) | (0.339) | (0.331) |
| $\hat{\beta}_{0.25}$ | 0.057 | 0.027 | -0.087 | 0.099 | 0.009 | -0.628 | -0.348 | -0.493 | -0.374 | -0.284 |
| (0.252) | (0.226) | (0.236) | (0.252) | (0.246) | (0.348) | (0.344) | (0.336) | (0.348) | (0.335) |
| $\hat{\beta}_{0.5}$ | -0.101 | -0.117 | -0.044 | 0.195 | 0.325 | 0.036 | 0.219 | -0.031 | 0.455 | 0.757 |
| (0.249) | (0.232) | (0.235) | (0.254) | (0.250) | (0.396) | (0.378) | (0.337) | (0.365) | (0.369) |
| $\hat{\beta}_{0.75}$ | -0.112 | -0.044 | 0.057 | 0.017 | 0.192 | 0.471 | 0.332 | 0.350 | 0.416 | 0.563 |
| (0.249) | (0.227) | (0.236) | (0.254) | (0.238) | (0.352) | (0.346) | (0.312) | (0.315) | (0.336) |
| $\hat{\beta}_{0.9}$ | -0.103 | -0.044 | 0.038 | -0.012 | 0.145 | 0.517 | 0.359 | 0.366 | 0.426 | 0.504 |
| (0.253) | (0.228) | (0.236) | (0.254) | (0.240) | (0.354) | (0.361) | (0.317) | (0.323) | (0.337) |

Table 12 Comparison between proposed methodology and that of Kong and Xia (2014) and Luo et al. (2014) estimators, using the bootstrapped error measurement for the Bitcoin data

| Methodology | 0.1 | 0.25 | 0.5 | 0.75 | 0.9 |
|-------------|-----|------|-----|------|-----|
| Proposed    | 0.454 | 0.461 | 0.511 | 0.451 | 0.405 |
| Kong and Xia (2014) | 0.385 | 0.477 | 0.695 | 0.451 | 0.352 |
| Luo et al. (2014) | 0.780 | 0.788 | 0.819 | 0.752 | 0.776 |

For each $\tau$, the value with the smallest bootstrapped error measurement is bolded. In the case of ties, all relevant values are bolded.

Note that in a financial risk setup, $Q_{\tau}(Y|X)$ is an important quantity, which is closely related to what is called the value-at-risk (VaR). Specifically, for $\{Y_t\}_{t=1}^{\tau}$ a time series of returns, the $\tau$th VaR at time $\tau$, denoted by $VaR_{\tau}(t)$, is the smallest number for which $P\{Y_t < -VaR_{\tau}(t)|\mathcal{F}_{t-1}\} = \tau$, where $\mathcal{F}_{t-1}$ denotes the information set at time $t-1$. Therefore, $-VaR_{\tau}(t)$ is the $\tau$th conditional quantile of $Y_t$, and we can write $Q_{\tau}(Y_t|\mathcal{F}_{t-1}) = -VaR_{\tau}(t)$. The parameter $\tau$ is typically chosen to be a small number such as 0.01, 0.025, or 0.05.

8 Discussion

In this work, we proposed a new dimension reduction technique with respect to the conditional quantile and suggested an easy to implement algorithm for estimating the $\tau$-CQS, for a given $\tau$. This method can be further generalized to any statistical functional of interest. Simulation results and real data applications demonstrated the theory developed here and suggested that the proposed methodology has a good finite sample performance and often outperforms the existing methods. Specifically, for a SIQR model, the proposed methodology has a comparable performance with that of Wu et al. (2010) and Alkenani and Yu (2013) methods. However, the last two methods are iterative. While iterations add to the computational complexity of the procedure, a more serious issue is that of convergence. For a MIQR model, the proposed methodology has a comparable performance with the method of Kong and Xia (2014), while outperforms the method of Luo et al. (2014). However, the proposed methodology can be extended to any statistical functional of interest, while the existing methods only focus on the conditional quantiles.

The presented paper focuses on extracting linear subspaces. For future work, we will consider nonlinear dimension reduction. Specifically, assuming that $Y \perp Q_{\tau}(Y|X)|\psi_t(X)$, where $\psi_t$ is an arbitrary function, then $\psi_t(X)$...
defines a *nonlinear* sufficient predictor. The goal is to estimate the nonlinear function $\psi_\tau$.

**Acknowledgements** We would like to thank Professors Michael Akritas and Bing Li from the Pennsylvania State University for useful discussions regarding the presented paper. We would also like to thank Mr. Mark Hamrick for help in running some of the simulations, and the two anonymous referees, whose comments lead to improvements in the presentation of this paper.

**Appendix A: notation and assumptions**

**Notation** We say that a function $m(\cdot) : \mathbb{R}^p \rightarrow \mathbb{R}$ has the order of smoothness $s$ on the support $X_0$, denoted by $m(\cdot) \in H_\tau(X_0)$, if (a) it is differentiable up to order $\lfloor s \rfloor$, where $\lfloor s \rfloor$ denotes the lowest integer part of $s$, and (b) there exists a constant $L > 0$, such that for all $u = (u_1, \ldots, u_p)^T$ with $|u| = u_1 + \cdots + u_p = \lfloor s \rfloor$, all $\tau$ in an interval $[\tau, \bar{\tau}]$, where $0 < \tau < \bar{\tau} < 1$, and all $x, x'$ in $X_0$,

$$|D^u m(x) - D^u m(x')| \leq L \|x - x'\|^{\lfloor s \rfloor},$$

where $D^u m(x)$ denotes the partial derivative $\partial^{|u|} m(x)/\partial x_1^u \cdots x_p^u$ and $\|\|_\cdot$ denotes the Euclidean norm.

**Assumptions**

A1 The following moment conditions are satisfied

$$E \left\|XX^\top\right\| < \infty, \quad E|Q_\tau(Y|A^\top X)|^2 < \infty, \quad E\left\{Q_\tau(Y|A^\top X)^2 \left\|XX^\top\right\|\right\} < \infty,$$

for a given $\tau \in (0, 1)$.

A2 The distribution of $A^\top X$ has a probability density function $f_A(\cdot)$ with respect to the Lebesgue measure, which is strictly positive and continuously differentiable over the support $X_0$ of $X$.

A3 The cumulative distribution function $F_{Y|A}(\cdot|\cdot)$ of $Y$ given $A^\top X$ has a continuous probability density function $f_{Y|A}(y|A^\top x)$ with respect to the Lebesgue measure, which is strictly positive for $y$ in $\mathbb{R}$ and $A^\top x$, for $x$ in $X_0$. The partial derivative $\partial F_{Y|A}(y|A^\top x)/\partial A^\top x$ is...
The first term follows from the Bahadur representation of $\tilde{A}$, and the assumption that $(x, y, y') \in \mathcal{X}_0 \times \mathbb{R}$. 

A4 The nonnegative kernel function $K(\cdot)$, used in (5), is Lipschitz over $\mathbb{R}^d$, $d \geq 1$, and satisfies $\int K(z)dz = 1$. For some $K > 0$, $K(z) \geq K/I(z \in B(0, 1))$, where $B(0, 1)$ is the closed unit ball. The associated bandwidth $h$, used in the estimation procedure, is in $[h, \bar{h}]$ with $0 < h \leq \bar{h} < \infty$, $\lim_n \rightarrow \infty \bar{h} = 0$ and $\lim_n \rightarrow \infty (\ln n)/(nh^d) = 0$.

A5 $Q_t(Y|\tilde{A}^T x)$ is in $H_{s_t}(T_A)$ for some $s_t$ with $|s_t| \leq 1$, where $T_A = \{ z \in \mathbb{R}^d : z = A^T x, x \in \mathcal{X}_0 \}$, and $\mathcal{X}_0$ is the support of $X$.

### Appendix B: Proof of main results

#### Appendix B.1: Some lemmas

**Lemma 1** Under Assumptions A2–A5 given in “Appendix A”, and the assumption that $\tilde{A}$ is $\sqrt{n}$-consistent estimate of the directions of the CS, then

$$\sup_{x \in \mathcal{X}_0} |\tilde{Q}_t(Y|\tilde{A}^T x) - Q_t(Y|A^T x)| = O_p(1),$$

where $\tilde{Q}_t(Y|\tilde{A}^T x)$ denotes the local linear conditional quantile estimate of $Q_t(Y|A^T x)$, given in (5).

**Proof** Observe that

$$\sup_{x \in \mathcal{X}_0} |\tilde{Q}_t(Y|\tilde{A}^T x) - Q_t(Y|A^T x)| \leq \sup_{x \in \mathcal{X}_0} |\tilde{Q}_t(Y|\tilde{A}^T x) - \hat{Q}_t(Y|A^T x)| + \sup_{x \in \mathcal{X}_0} |\hat{Q}_t(Y|A^T x) - Q_t(Y|A^T x)| = O_p(1).$$

The first term follows from the Bahadur representation of $\tilde{Q}_t(Y|\tilde{A}^T x) - Q_t(Y|A^T x)$ (see Guerre and Sabbah 2012) and the $\sqrt{n}$-consistency of $\tilde{A}$. The second term follows from Corollary 1 (ii) of Guerre and Sabbah (2012). $\square$

**Note** For the study of the asymptotic properties of $\hat{b}_t$, defined in (4), we consider an equivalent objective function. Observe that minimizing $\sum_{i=1}^n |\tilde{Q}_t(\tilde{A}^T X_i) - a_t - b_t^T X_i|^2$ with respect to $(a_t, b_t)$, is equivalent with minimizing

$$\tilde{S}_n(a_t, b_t) = \frac{1}{2} \sum_{i=1}^n |\tilde{Q}_t(\tilde{A}^T X_i) - a_t - b_t^T X_i|^2$$

with respect to $(a_t, b_t)$. By expanding the square, (10) can be written as

$$\tilde{S}_n(a_t, b_t) = -(a_t, b_t^T)\sum_{i=1}^n \tilde{Q}_t(\tilde{A}^T X_i)(1, X_i) + \frac{1}{2} (a_t, b_t^T)\sum_{i=1}^n (1, X_i)(1, X_i^T)(a_t, b_t).$$

(11)

**Lemma 2** Let $\tilde{S}_n(y_t/\sqrt{n} + (a_t^*, b_t^*))$ be as defined in (11), where $y_t = \sqrt{n}(a_t, b_t) - (a_t^*, b_t^*)$ and $(a_t^*, b_t^*)$ is defined in (3). Then, under the assumptions of Lemma 1 and additionally Assumption A1 of “Appendix A”, we have the following quadratic approximation, uniformly in $y_t$ in a compact set,

$$\tilde{S}_n(y_t/\sqrt{n} + (a_t^*, b_t^*)) = \frac{1}{2} y_t^T \mathbb{V} y_t + W_{t,n} y_t + C_{t,n} + o_p(1),$$

where $\mathbb{V} = E[(1, X)(1, X^T)^T]$, 

$$W_{t,n} = -\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{Q}_t(\tilde{A}^T X_i)(1, X_i),$$

and

$$C_{t,n} = -\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{Q}_t(\tilde{A}^T X_i)(1, X_i^T)(a_t^*, b_t^*) + \frac{1}{2} (a_t^*, b_t^*)^T \sum_{i=1}^n (1, X_i)(1, X_i^T)(a_t^*, b_t^*).$$

(13)

**Proof** Observe that

$$\tilde{S}_n(y_t/\sqrt{n} + (a_t^*, b_t^*)) = \frac{1}{2} y_t^T \mathbb{V} y_t - \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{Q}_t(\tilde{A}^T X_i)(1, X_i)^T y_t$$

$$- \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{Q}_t(\tilde{A}^T X_i)(1, X_i^T)^T y_t$$

$$- \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{Q}_t(\tilde{A}^T X_i)(1, X_i^T)(a_t^*, b_t^*)$$

$$+ \frac{1}{2} (a_t^*, b_t^*)^T \sum_{i=1}^n (1, X_i)(1, X_i^T)(a_t^*, b_t^*)$$

$$= \frac{1}{2} y_t^T \mathbb{V} y_t + W_{t,n} y_t + C_{t,n},$$

$\square$
Will prove that \( S_n(\gamma / \sqrt{n} + (\alpha^*, \beta^*)) = \frac{1}{2} \gamma ^T V \gamma + W_{t,n}^T \gamma + C_{r,n} + o_p(1). \)

Provided that \( W_{t,n} \) is stochastically bounded, it follows from the convexity lemma (Pollard 1991) that the quadratic approximation to the convex function \( S_n(\gamma / \sqrt{n} + (\alpha^*, \beta^*)) \) holds uniformly for \( \gamma \) in a compact set. Remains to prove that \( W_{t,n} \) is stochastically bounded.

Since \( W_{t,n} \) involves the quantity \( \hat{Q}_t(Y|\hat{A}^T X_i) \), which is data dependent and not deterministic function, we define

\[
\hat{W}_{t,n}(\phi_t) = -\frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_t(Y|\hat{A}^T X_i)(1, X_i),
\]

where \( \phi_t: \mathbb{R}^{d+1} \rightarrow \mathbb{R} \) is a function in the class \( \Phi_t \), whose value at \((y, \hat{A}^T x) \in \mathbb{R}^{d+1}\) can be written as \( \phi_t(y|\hat{A}^T x) \), in the non-separable space \( L^2(\gamma, \hat{A}^T x) = \{(\gamma, \hat{A}^T x) : \mathbb{R}^{d+1} \rightarrow \mathbb{R} : \|\phi_t\|_{(y, \hat{A}^T x)} := \sup_{(y, \hat{A}^T x) \in \mathbb{R}^{d+1}} \|\phi_t(y|\hat{A}^T x)\| < \infty, \) and satisfying \( E|\phi_t(Y|\hat{A}^T X)^2 < \infty \) and

\[
E \left\| \phi_t(Y|\hat{A}^T X)^2 XX^T \right\| < \infty.
\]

Since \( \Phi_t \) includes \( Q_t(Y|\hat{A}^T x) \), and according to Lemma 1, includes \( \hat{Q}_t(Y|\hat{A}^T x) \) for \( n \) large enough, almost surely, we will prove that \( W_{t,n}(\phi_t) \) is stochastically bounded, uniformly on \( \phi_t \in \Phi_t \).

Observe that

\[
\sup_{\phi_t \in \Phi_t} \left\| E \left[ W_{t,n}(\phi_t) W_{t,n}^T(\phi_t) \right] \right\| \\
\leq \sup_{\phi_t \in \Phi_t} \frac{1}{n} \sum_{i=1}^n E \left\| \phi_t(Y|\hat{A}^T X_i)(1, X_i) \right\|^2 \left\| (1, X_i) (1, X_i) \right\| \\
= O \left( E \left\| \phi_t(Y|\hat{A}^T X)^2 \right\| (1, X_i)(1, X_i) \right) = O(1),
\]

which follows from the properties of the class \( \Phi_t \) defined above. Bounded second moment implies that \( W_{t,n}(\phi_t) \) is stochastically bounded. Since

1. The result was proven uniformly on \( \phi_t \), and
2. The class \( \Phi_t \) includes \( \hat{Q}_t(Y|\hat{A}^T x) \) for \( n \) large enough, almost surely,

the proof follows.

\[\square\]

### Appendix B.2: Proof of theorem 4

To prove the \( \sqrt{n} \)-consistency of \( \hat{\beta}_t \), enough to show that for any given \( \delta_t > 0 \), there exists a constant \( C_t \) such that

\[
\Pr \left\{ \inf_{\|\gamma\| \geq C_t} \hat{S}_n(\gamma / \sqrt{n} + (\alpha^*, \beta^*)) > \hat{S}_n(\alpha^*, \beta^*) \right\} \geq 1 - \delta_t,
\]

(14)

where \( \hat{S}_n(\gamma / \sqrt{n} + (\alpha^*, \beta^*)) \) defined in (10) and implies that with probability at least \( 1 - \delta_t \), there exists a local minimum in the ball \( \{\gamma / \sqrt{n} + (\alpha^*, \beta^*) : \|\gamma\| \leq C_t \} \). This in turn implies that there exists a local minimizer such that \( \|\hat{\alpha}_t - (\alpha^*, \beta^*)\| = O_p(n^{-1/2}) \). The quadratic approximation derived in Lemma 2 yields that

\[
\hat{S}_n(\gamma / \sqrt{n} + (\alpha^*, \beta^*)) - \hat{S}_n(\alpha^*, \beta^*) \\
= \frac{1}{2} \gamma ^T V \gamma + W_{t,n}^T \gamma + o_p(1),
\]

(15)

for any \( \gamma \) in a compact subset of \( \mathbb{R}^{p+1} \). Therefore, the difference (15) is dominated by the quadratic term \( 1/2 \gamma ^T V \gamma \) for \( \|\gamma\| \) greater than or equal to sufficiently large \( C_t \). Hence, (14) follows.

\[\square\]

### Appendix B.3: Proof of theorem 8

Let \( \hat{V}_t = (\hat{\beta}_{t,0}, \ldots, \hat{\beta}_{t,p-1}) \) be a \( p \times p \) matrix, where \( \hat{\beta}_{t,0} = \hat{\beta}_t \), defined in (4), and \( \hat{\beta}_{t,j} = E_n(\hat{Q}_t(Y|\hat{\beta}_{t,j-1}^T X)|X) \) for \( j = 1, \ldots, p - 1 \). Moreover, let \( \hat{V}_t \) be the population level of \( \hat{V}_t \). It is easy to see that \( \hat{V}_t \) converges to \( V_t \) at \( \sqrt{n} \)-rate. This follows from the central limit theorem and Lemma 1. Then, for \( \|\cdot\| \) the Frobenius norm,

\[
\left\| \hat{V}_t \hat{V}_t^T - V_t V_t^T \right\| \leq \left\| \hat{V}_t \hat{V}_t^T - \hat{V}_t V_t^T \right\| + \left\| \hat{V}_t V_t^T - V_t V_t^T \right\| = O_p(n^{-1/2}),
\]

and the eigenvectors of \( \hat{V}_t \hat{V}_t^T \) converge to the corresponding eigenvectors of \( V_t V_t^T \). Finally, the subspace spanned by the \( d_t \) eigenvectors of \( V_t V_t^T \) falls into \( S_{\mathcal{Q}_t}(Y|X) \) and the proof is complete.

\[\square\]

### References

Alkenani, A., Yu, K.: Penalized single-index quantile regression. Int. J. Stat. Probab. 2(3), 12–30 (2013)

Breiman, L., Friedman, J.H.: Estimating optimal transformations for multiple regression and correlation. J. Am. Stat. Assoc. 80(391), 580–598 (1985)
