Submitted to the Annals of Statistics
2014 Vol. 0, No. 0, 1–8

AN ADAPTIVE COMPOSITE QUANTILE APPROACH TO DIMENSION REDUCTION

By Efang kong[†] and Yingcun xia[*,‡]

University of Kent at Canterbury[†] *and National University of Singapore*[‡]

Sufficient dimension reduction [Li (1991)] has long been a prominent issue in multivariate nonparametric regression analysis. To uncover the central dimension reduction space, we propose in this paper an adaptive composite quantile approach. Compared to existing methods, (1) it requires minimal assumptions and is capable of revealing all dimension reduction directions; (2) it is robust against outliers; and (3) it is structure-adaptive, thus more efficient. Asymptotic results are proved and numerical examples are provided, including a real data analysis.

1. Introduction. Dimension reduction is a rather amorphous concept in statistics, changing its characteristics and taking different forms depending on the context. In regression, the paradigm of sufficient dimension reduction [Li (1991), Cook (1994), Cook (1998)] which combines the idea of dimension reduction with the concept of sufficiency, aims to generate low-dimensional summary plot without appreciable loss of information. In most cases, reductions are typically constrained to be linear and the goal then is to estimate the central dimension reduction subspace, or simply the central subspace.

Cook (2007) gave a formal definition and overviews of the sufficient dimension reduction in regression, which we adopt in this paper for the definition of the central subspace. Suppose Y is a scalar dependent variable and $\mathbf{X}$ is the corresponding $p \times 1$ vector of predictors. Let $\mathbf{B}$ be a $p \times q (q \le p)$ (constant) orthonormal matrix and $\mathbf{B}^\top$, its transpose. The space $\mathcal{S}(\mathbf{B})$ spanned by the columns of $\mathbf{B}$, is said to be the (sufficient) dimension reduction subspace (DRS), if the conditional distribution $F(.|\mathbf{B}^\top\mathbf{X})$ of Y given $\mathbf{B}^\top\mathbf{X}$ is identical to $F(.|\mathbf{X})$, i.e.

$$F(Y|\mathbf{X}) = F(Y|\mathbf{B}^\top\mathbf{X}) \quad \text{almost surely.} \tag{1.1}$$

Consequently, a subspace is called a central subspace (CS), if it is not only itself a DRS, but also a subset of any other DRS'. It thus represents the

[*]Supported by National Natural Science Foundation of China 71371095 and NUS grant R-155-000-121-112

MSC 2010 subject classifications: Primary 62J07

Keywords and phrases: Bahadur approximation, sufficient dimension reduction, local polynomial smoothing, quantile regression, semi-parametric models, U-processes

imsart-aos ver. 2014/02/20 file: AOS1242.tex date: May 27, 2014

minimal subspace that captured all the information relevant to regressing Y on $\mathbf{X}$. Under quite general conditions, the CS exists and is given by

$$\mathcal{S}_0 = \cap\{\mathcal{S}(\mathbf{B}) : \text{model (1.1) holds for } \mathbf{B}\};$$

see, Yin, Li and Cook (2008) for the latest results on sufficient conditions for the existence of CS. Its dimension $dim(\mathcal{S}_0) = q(\leq p)$ is referred to as the structural dimension, while its orthogonal basis $\beta_{01}, \cdots, \beta_{0q}$ is called the dimension reduction directions or simply the CS directions. Let $\mathbf{B}_0 = (\beta_{01}, \cdots, \beta_{0q})$ and thus equivalent to (1.1), we have

$$F(Y|\mathbf{X}) = F(Y|\mathbf{B}_0^\top \mathbf{X}) \quad \text{almost surely.} \tag{1.2}$$

Research in dimension reduction methodologies, namely the search of CS (directions), has garnered tremendous interest [Hristache et al (2001), Yin and Cook (2002), Xia et al (2002), Li et al (2003, 2005), Lue (2004), Zhu and Zeng (2006) and Ma and Zhu (2012)] since the seminal work of Li (1991). Some earlier research in this area such as Li (1991), was often based on either restrictive or hard-to-verify assumptions, which limited their applications; while others being model (moment)-based, targeted not at $S(\mathbf{B}_0)$, but instead the reduction subspace $S(\mathbf{B})$ associated with certain functional of $F(Y|\mathbf{X})$, e. g., the conditional mean [Cook and Li (2002)] or the conditional variance [Zhu and Zhu (2009)]. As we are going to demonstrate through the following example, such subspace quite often is strictly a subset of CS. Consider the following model where

$$Y = \beta_1^\top \mathbf{X} + \beta_2^\top \mathbf{X}\ \varepsilon, \quad \text{and} \quad E(\varepsilon|\mathbf{X}) = 0. \tag{1.3}$$

As $E(Y|\mathbf{X}) = \beta_1^\top \mathbf{X}$, the central mean subspace $S(\beta_1)$ is thus strictly contained in $S(\beta_1, \beta_2)$, the full CS.

Seeing the restrictions with the aforementioned moment-based methods, some consider the possibility of recovering all CS directions by taking transformation of the response variable Y. See, for example, Zhu and Zeng (2006), which practically requires assuming a parametric model for $\mathbf{X}$; or Fukumizu et al (2009), where no theoretical results are available; and Yin and Li (2011). Others [Xia (2007), Zhu et al (2010), Wang and Xia (2008)] tried to extract information on CS directly from the conditional density or distribution function. A major drawback of the methodologies in the preceding four references is that the embedded estimation procedure is not structure-adaptive, rendering the subsequent estimators of CS (directions) less efficient. To see this, take model (1.3) for example. As the conditional density (distribution) function is nonlinear, the smoothing parameter used in constructing their kernel estimators must therefore be small, i.e. only a small proportion of data is being used for local estimation. In contrary, the conditional quantile function is in this case at least piecewise linear and consequently its estimation can be made more efficient through the use of a larger (data-driven) bandwidth. Another reason for us to consider a conditional-quantile based approach is the theoretical equivalence between conditional distribution functions and conditional quantiles.

As in the case of conditional mean-based approach, we do not expect the CS (directions) to be fully revealed via quantile regression at any individual level. The solution we shall propose in this paper is a combination of dimension reduction methods of Xia et al (2002) and the composite quantile approach for regression [Zou and Yuan (2008), Kai et al (2010), He et al (2013)], together with a adaptive-weighting strategy. The advantages of this new approach include: (1) it requires minimal assumptions and can identify the CS directions exhaustively; (2) it is robust against outliers, a property inherited from quantile regression; and (3) the embedded estimation procedure is structure-adaptive, i.e. the use of a data-driven bandwidth means more efficient use of data.

The paper is organized as follows. In Section 2 we show how the CS characterizes the composite outer product of gradients matrix. Based on this characterization, Section 3 describes how an adaptive composite quantile approach is integrated with the outer-product of gradients (qOPG) method, and for comparison purposes, the composite quantile minimum average variance method (qMAVE). In Section 4, we present regularity conditions and theoretical results on the asymptotic normality of the qOPG estimator. Section 5 and 6 examine some practical issues, such as bandwidth selection and determination of the structural dimension. Section 7 contains some numerical results, including an example of real data analysis. Section 8 provides concluding remarks. All proofs are given in the Appendix.

2. A composite quantile approach. Under model (1.2), for any $0 < \tau < 1$, the τth conditional quantile of Y given $\mathbf{X}$,

$$Q_\tau(\mathbf{X}) = \min\{y : F(y|\mathbf{X}) \geq \tau\}$$

admits the following alternative expression

$$Q_\tau(\mathbf{X}) = \min\{y : F(y|\mathbf{B}_0^\top \mathbf{X}) \geq \tau\} = \tilde{Q}_\tau(\mathbf{B}_0^\top \mathbf{X}). \tag{2.1}$$

Its gradient vector

$$\nabla Q_\tau(\mathbf{x}) = \Big[\frac{\partial Q_\tau(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial Q_\tau(\mathbf{x})}{\partial x_p}\Big]^\top$$

defined for any $\mathbf{x} = (x_1, \cdots, x_p)^\top \in R^p$, is thus related to $\nabla \tilde{Q}_\tau(.)$, the gradient vector of $\tilde{Q}_\tau(.)$, via the following identity

$$\nabla Q_\tau(\mathbf{x}) = \mathbf{B}_0 \nabla \tilde{Q}_\tau(\mathbf{B}_0^\top \mathbf{x}). \tag{2.2}$$

Consequently, we have the following fact for the corresponding outer-product of gradients (OPG) matrix specific to level τ:

$$\begin{aligned}\Sigma(\tau) &= E\{\nabla Q_\tau(\mathbf{X})[\nabla Q_\tau(\mathbf{X})]^\top\} \\ &= \mathbf{B}_0 E\{\nabla\tilde{Q}_\tau(\mathbf{B}_0^\top\mathbf{X})[\nabla\tilde{Q}_\tau(\mathbf{B}_0^\top\mathbf{X})]^\top\}\mathbf{B}_0^\top. \qquad (2.3)\end{aligned}$$

It is obvious that for any $\tau \in (0,1)$,

$$\mathcal{S}(\Sigma(\tau)) \subseteq \mathcal{S}(\mathbf{B}_0).$$

Indeed, plenty of examples exist where the above inequality holds strictly for at least one $\tau \in (0,1)$. Consider, for example, model (1.3) with $\tau = 0.5$ and the median of ε equal to zero. In other words, the CS may not be fully recovered by OPG matrices specific to any finite number of quantile levels. The solution instead lies with the composite OPG matrix defined as

$$\Sigma = \int_0^1 \Sigma(\tau)d\tau, \qquad (2.4)$$

as stated in the following lemma.

LEMMA 1. *Suppose $\nabla Q_\tau(.)$ exists for almost all $\tau \in (0,1)$ and $\mathbf{X}$. We have $\mathcal{S}(\Sigma) = \mathcal{S}(\mathbf{B}_0)$.*

By definition, the composite OPG matrix Σ is simply an equally weighted average of the level-specific OPG matrices $\Sigma(\tau)$, $0 < \tau < 1$. As previously demonstrated, $\Sigma(\tau)$ for a given τ might contain little or no information at all about the CS. Consider another example where $Y = \mathbf{x}_1\varepsilon$, $X = (\mathbf{x}_1, ..., \mathbf{x}_p)^\top$ and ε has median zero. It is easy to see that $\Sigma(0.5) = \mathbf{0}$, a $p \times p$ zero matrix. We call such $\Sigma(\tau)$ uninformative, to which less weight should be assigned for the purpose of a more revealing composite OPG matrix. Since whether or not any level-specific $\Sigma(.)$ is uninformative is not given a priori, we suggest the following procedure to obtain an adaptively weighted composite OPG matrix. Suppose we have decided on the structural dimension q. For any given $\tau \in (0,1)$, denote by $\lambda_1(\tau) \geq ... \geq \lambda_p(\tau) \geq 0$, the p eigenvalues of $\Sigma(\tau)$. The 'adaptively weighted' composite OPG matrix is consequently defined as

$$\Sigma_w = \int_0^1 w(\tau)\Sigma(\tau)d\tau,$$

where the weight function

$$w(\tau) = \frac{\lambda_1(\tau) + ... + \lambda_q(\tau)}{\lambda_1(\tau) + ... + \lambda_p(\tau)}, \qquad (2.5)$$

reflects the percentage of information contained in the first q eigenvectors of $\Sigma(\tau)$. If $\Sigma(\tau) = \mathbf{0}$, we define $w(\tau) = 0$. Note that as $\mathcal{S}(\Sigma(\tau)) \subseteq \mathcal{S}(\mathbf{B}_0)$ for any τ, we have $w(\tau) = 1$ for any τ such that $\Sigma(\tau) > \mathbf{0}$. In practice, weights $w(.)$ are derived from eigenvalues of estimates of $\Sigma(\tau)$.

3. Estimation of the dimension reduction directions. Based on Lemma 1, the key to recovering the CS directions lies with the estimation of the composite OPG matrix Σ, which in turn depends on the availability of a proper estimate of the gradient vector $\nabla Q_\tau(\mathbf{x})$ for any given $\tau \in (0,1)$ and $\mathbf{x} \in R^p$. Let $\hat{\nabla} Q_\tau(\mathbf{x})$ denote such an estimate. We can then construct estimate of the level-specific OPG matrix (2.3) and consequently estimate of the composite OPG matrix (2.4), as follows

$$\hat{\Sigma}(\tau) = \frac{1}{n}\sum_{j=1}^{n} \hat{\nabla} Q_\tau(\mathbf{X}_j), \quad \hat{\Sigma} = \int_0^1 \hat{\Sigma}(\tau) d\tau. \tag{3.1}$$

Various nonparametric estimators of $\nabla Q_\tau(.)$ could be used in (3.1), including kernel smoothing, nearest neighbor and spline estimators; see, e. g., Truong (1989), Bhattacharya and Gangopadhyay (1991), and Koenker, Ng and Portnoy (1992, 1994). In this paper, we opt for the local polynomial estimators of Chaudhuri (1991) and Kong et al (2010). This is because, to show that $\hat{\Sigma}$ is root-n consistent and asymptotically 'normal', we need the following two pre-requisites: (i) $\hat{\nabla} Q_\tau(\mathbf{x})$ has a bias of order $o_p(n^{-1/2})$ uniformly in $\mathbf{x}$ and in τ; (ii) a Bahadur-type expansion of $\hat{\nabla} Q_\tau(\mathbf{x})$, again uniformly in $\mathbf{x}$ as well as in τ. Condition (i) can be met by approximating $Q_\tau(.)$ locally with polynomials in p variables with high enough degrees. Condition (ii), to be proved in the Appendix using results on empirical processes and $U-$processes, extends what was obtained in Kong et al (2010), where the uniformity is with respect to $\mathbf{x}$ only.

Suppose there exists some positive integer k such that, for all $\tau \in (0,1)$, $Q_\tau(.)$ has partial derivatives of order up to k on $\mathcal{D}$, the compact support of $\mathbf{X}$ in R^p. Consequently, for any given $\mathbf{x} = (x_1, \cdots, x_p)^\top \in \mathcal{D}$ and $\mathbf{X}$ near $\mathbf{x}$, $Q_\tau(\mathbf{X})$ can be approximated by its kth order Taylor expansion, i.e.

$$Q_\tau(\mathbf{X}) \approx Q_\tau(\mathbf{x}) + \sum_{1\le[u]\le k} \frac{D^{\mathbf{u}} Q_\tau(\mathbf{x})}{\mathbf{u}!}(\mathbf{X}-\mathbf{x})^{\mathbf{u}}, \tag{3.2}$$

where $\mathbf{u} = (u_1, \cdots, u_p)$ denotes a generic $p-$dimensional vector of nonnegative integers, $[\mathbf{u}] = \sum_{i=1}^p u_i$, $\mathbf{u}! = \prod_{i=1}^p u_i!$, $\mathbf{x}^{\mathbf{u}} = \prod_{i=1}^p x_i^{u_i}$ with the convention that $0^0 = 1$, and $D^{\mathbf{u}}$ denotes the differential operator $\partial^{[\mathbf{u}]}/\partial x_1^{u_1} \cdots, x_p^{u_p}$. For ease of reference, write $A = \{\mathbf{u} : [\mathbf{u}] \le k\}$ and $s(A) = \sharp(A)$, the cardinality of A.

Suppose $(\mathbf{X}_i, Y_i)$, $i = 1, \cdots, n$, are i.i.d. copies of $(\mathbf{X}, Y)$, and h_n is a smoothing parameter such that $h_n \to 0$, as $n \to \infty$. For any given $\mathbf{x} \in R^p$ and $\tau \in (0,1)$, define two $s(A) \times 1$ vectors as follows

$$\mathbf{x}(h_n, A) = (\mathbf{x}(h_n, \mathbf{u}))_{\mathbf{u}\in A} \quad \text{with} \quad \mathbf{x}(h_n, \mathbf{u}) = h_n^{-[\mathbf{u}]}\mathbf{x}^{\mathbf{u}},$$
$$\mathbf{c}_n(\mathbf{x}; \tau) = (c_{n,\mathbf{u}}(\mathbf{x}; \tau))_{\mathbf{u}\in A} \quad \text{with} \quad c_{n,\mathbf{u}}(\mathbf{x}; \tau) = h_n^{[\mathbf{u}]} D^{\mathbf{u}} Q_\tau(\mathbf{x})/\mathbf{u}!.$$

The local polynomial estimate of $\mathbf{c}_n(\mathbf{x};\tau)$ is defined as a solution to the following problem

$$\min_{\mathbf{c}} \sum_{i=1}^{n} \rho_\tau\{Y_i - \mathbf{c}^\top \mathbf{X}_{i\mathbf{x}}(h_n, A)\} K_{h_n}(|\mathbf{X}_{i\mathbf{x}}|), \tag{3.3}$$

where $\mathbf{c} = (c_{\mathbf{u}})_{\mathbf{u}\in A} \in R^{s(A)}$, $\rho_\tau(s) = |s| + (2\tau - 1)s$, $\mathbf{X}_{i\mathbf{x}} = \mathbf{X}_i - \mathbf{x}$, $|.|$ stands for the supremum norm, $K(.)$ is a kernel function in R^p with finite support, and $K_{h_n}(.) = K(./h_n)/h_n$. Note that although in this paper we take $K(.)$ to be the uniform density function on $[-1,1]^p$, the p-dimensional cube in R^p, the results we obtain apply to other cases such as the Epanechnikov kernel as well.

Since $\rho_\tau(s) \to \infty$, as $|s| \to \infty$, solution to (3.3) always exists as long as $K_{h_n}(|\mathbf{X}_{i\mathbf{x}}|) > 0$ for at least one $\mathbf{X}_i$. Denote by $\hat{\mathbf{c}}_n(\mathbf{x};\tau) = (\hat{c}_{n,\mathbf{u}}(\mathbf{x};\tau))_{\mathbf{u}\in A}$, a solution to (3.3) and by $\hat{\nabla} Q_\tau(\mathbf{x})$, the local polynomial estimate of the gradient vector $\nabla Q_\tau(\mathbf{x})$:

$$\hat{\nabla} Q_\tau(\mathbf{x}) = h_n^{-1}(\hat{\mathbf{c}}_{n,\mathbf{u}}(\mathbf{x};\tau))_{\mathbf{u}\in A,[\mathbf{u}]=1}.$$

Consequently, we can construct estimates of the level-specific OPG matrix $\Sigma(\tau)$ and of the composite OPG matrix Σ as follows:

$$\hat{\Sigma}(\tau) = \frac{1}{n}\sum_{j=1}^{n} \hat{\nabla} Q_\tau(\mathbf{X}_j)\{\hat{\nabla} Q_\tau(\mathbf{X}_j)\}^\top; \quad \hat{\Sigma} = \int_0^1 \hat{\Sigma}(\tau) d\tau. \tag{3.4}$$

For the sake of technical convenience, we focus on rather than the $\hat{\Sigma}$ in (3.4) but instead the following truncated version

$$\hat{\Sigma}_{\mathrm{T}} = \int_{\delta^*}^{1-\delta^*} \hat{\Sigma}(\tau) d\tau, \tag{3.5}$$

for some small $\delta^* \in (0,1)$. This is due to the fact that the uniformity in τ of the strong Bahadur type representation of $\hat{\nabla} Q_\tau(\mathbf{x})$ requires the conditional density of Y given $\mathbf{X}$ at $Q_\tau(\mathbf{X})$ to be uniformly bounded away from zero, a condition apparently cannot be met by all $\tau \in (0,1)$. See Lemma 2 and its proof given in the Appendix for more details. Nevertheless, such truncation need not cause much concern. The reasons are twofold. On one hand, the integral in (3.4) is approximated as a summation over a sequence of discretised τ values. On the other hand, the CS which is derived from $\{Q_\tau(.|\mathbf{x}) : 0 < \tau < 1, \ \mathbf{x} \in \mathcal{D}\}$ or equivalently from Σ, is expected to closely resemble, if not completely identical to, that from $\{Q_\tau(.|\mathbf{x}) : \delta^* \le \tau \le 1 - \delta^*, \ \mathbf{x} \in \mathcal{D}\}$ or equivalently from

$$\Sigma_{\mathrm{T}} = \int_{\delta^*}^{1-\delta^*} \Sigma(\tau) d\tau,$$

provided that $\delta^* > 0$ is small enough. We assume this is indeed the case, i.e. $\Sigma_{\mathrm{T}} = \Sigma$.

As suggested at the end of Section 2, we could further construct an estimate of the adaptively-weighted truncated composite OPG matrix as

$$\hat{\Sigma}_{w\mathrm{T}} = \int_{\delta^*}^{1-\delta^*} \hat{\Sigma}(\tau)\hat{w}(\tau)d\tau, \tag{3.6}$$

with weight $\hat{w}(\tau)$ calculated according to formula (2.5) using the eigenvalues of $\hat{\Sigma}(\tau)$. However, to make sure less weights are assigned to those uninformative matrices $\hat{\Sigma}(\tau)$ which are close to but not exactly zero, we set $\hat{w}(\tau) = 0$ if the largest eigenvalue of $\hat{\Sigma}(\tau)$ is below certain threshold.

In the ideal case where the structural dimension q is known a priori, estimates of the CS directions are simply given by the first q eigenvectors of $\hat{\Sigma}_{\mathrm{T}}$: $\hat{\beta}_k,\ k = 1, \cdots, q$. Details on how to estimate q when it is unknown as well as bandwidth selection are given in Sections 5 and 6, respectively. Similar to Xia et al (2002), the above estimator can be further refined as follows. Re-label the above obtained estimate $\hat{\mathbf{B}} = (\hat{\beta}_1, \cdots, \hat{\beta}_q)$ as $\mathbf{B}^{(1)}$, and the smoothing parameter h_n used in obtaining it as $h_n^{(1)}$. Construct a refined estimate of $\nabla Q_\tau(\mathbf{x})$ as

$$\hat{\nabla} Q_\tau^{(2)}(\mathbf{x}) = (\hat{c}_{n,\mathbf{u}}^{(2)}(\mathbf{x};\tau))_{\mathbf{u}\in A,[\mathbf{u}]=1}/h_n^{(2)},$$

where

$$\hat{c}_{n,\mathbf{u}}^{(2)}(\mathbf{x};\tau) = \arg\min_{\mathbf{c}} \sum_{i=1}^{n} \rho_\tau\{Y_i - \mathbf{c}^\top \mathbf{X}_{i\mathbf{x}}(h_n^{(1)}, A)\} K_{h_n^{(2)}}(|\mathbf{X}_{i\mathbf{x}}^\top \mathbf{B}^{(1)}|), \tag{3.7}$$

and $K(.)$ is a kernel density in R^q. Accordingly, the estimates $\hat{\Sigma}(\tau)$ and $\hat{\Sigma}_{\mathrm{T}}$ in (3.4) and (3.5) could be refined respectively as

$$\hat{\Sigma}^{(2)}(\tau) = \frac{1}{n}\sum_{j=1}^{n} \hat{\nabla} Q_\tau^{(2)}(\mathbf{X}_j)\{\hat{\nabla} Q_\tau^{(2)}(\mathbf{X}_j)\}^\top$$

and

$$\hat{\Sigma}_{\mathrm{T}}^{(2)} = \int_{\delta^*}^{1-\delta^*} \hat{w}^{(2)}(\tau)\hat{\Sigma}^{(2)}(\tau)d\tau,$$

where $\hat{w}^{(2)}(\tau)$ is constructed in the same way as $\hat{w}(\tau)$, using eigne-values of $\hat{\Sigma}^{(2)}(\tau)$. Again, pick the first q eigenvectors of $\hat{\Sigma}_{\mathrm{T}}^{(2)}$ to construct a new matrix $\mathbf{B}^{(2)}$ which can then be substituted into (3.7) for $\mathbf{B}^{(1)}$. Repeat the above two steps until convergence is reached. Intuitively, this refined estimate of Σ is more efficient due to the use of a lower dimensional kernel when estimating

$\nabla Q_\tau(\mathbf{x})$, thus mitigating the so-called 'curse of dimensionality' problem. We call the above procedure the adaptive composite quantile outer product of gradients (qOPG).

We can also incorporate this 'composite-quantile' idea into the minimum average variance estimation (MAVE) procedure of Xia et al (2002) and propose a composite quantile MAVE (qMAVE) as follows. With structural dimension q, consider the following minimization problem

$$\int_{\delta^*}^{1-\delta^*} \sum_{j=1}^{n}\sum_{i=1}^{n} \rho_\tau\{Y_i - a_j - b_j^\top \mathbf{B}^\top \mathbf{X}_{ij}\} K_{h_n}(|\mathbf{X}_{ij}|) d\tau, \tag{3.8}$$

with respect to $p\times q$ matrix $\mathbf{B}$, where $\mathbf{X}_{ij} = \mathbf{X}_i - \mathbf{X}_j$. Again, just as in (3.7), a possibly lower dimensional kernel $K_{h_n}(|\mathbf{B}^\top \mathbf{X}_{ij}|)$ could be used to replace $K_{h_n}(|\mathbf{X}_{ij}|)$ in (3.8), in the hope of an improved efficiency of the resulted estimator, at least with finite-sample size. Estimates of the q CS directions are thus given by the orthonormal-ized columns of $\hat{\mathbf{B}}$, the solution to (3.8). Realization of (3.8) is similar to that of Xia (2007) and its theoretical properties can also be similarly investigated by combining the results obtained in the Appendix and the proofs in Xia (2007).

To find out whether a qMAVE procedure would benefit from some 'adaptive' weighting scheme, one could consider for example a level-specific qMAVE procedure, where

$$\hat{\mathbf{B}}(\tau) = \arg\min_{\mathbf{B}\in R^{p\times q}} \min_{a_j,b_j} \sum_{j=1}^{n}\sum_{i=1}^{n} \rho_\tau\{Y_i - a_j - b_j^\top \mathbf{B}^\top \mathbf{X}_{ij}\} K_{h_n}(|\mathbf{X}_{ij}|) d\tau,$$

and consequently define

$$\hat{\Sigma}_w^* = \int_{\delta^*}^{1-\delta^*} \hat{\mathbf{B}}(\tau)\hat{\mathbf{B}}(\tau)^\top \hat{w}(\tau) d\tau,$$

where $\hat{w}(\tau)$ is the same as in (3.6) derived from the level-specific OPG matrix. Our experience is such that this level-specific qMAVE is always outperformed by both the qMAVE procedure of (3.8) and qOPG. A possible explanation is that $\hat{\mathbf{B}}(\tau)$ being an orthonormal matrix means that all directions (columns of $\hat{\mathbf{B}}(\tau)$) are equally weighted, whereas in qOPG the corresponding directions (eigenvectors) are given different weights dictated by their respective eigenvalues.

4. Assumptions and theoretical results.

For any $s_0 = l + \gamma$, with non-negative integer l and $0 < \gamma \le 1$, we say a function $m(.) : R^p \to R$ has the order of smoothness s_0 on $\mathcal{D}$, denoted by $m(.) \in H_{s_0}(\mathcal{D})$ if, it is differentiable up to order l and there exists a constant $C > 0$, such that

$$|D^{\mathbf{u}} m(\mathbf{x}_1) - D^{\mathbf{u}} m(\mathbf{x}_2)| \le C|\mathbf{x}_1 - \mathbf{x}_2|^\gamma, \quad \text{for all } \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}, \text{ and } [\mathbf{u}] = l.$$

We assume the following conditions hold throughout the paper.

[A1] The support $\mathcal{D}$ of $\mathbf{X}$ is open, convex and the probability density function of $\mathbf{X}$ is such that $f_{\mathbf{X}}(.) \in H_{s_1}(\mathcal{D})$, for some $s_1 > 0$.

[A2] The conditional quantile function $Q_\tau(.) \in H_{s_2}(\mathcal{D})$ for some $s_2 > 0$ uniformly in $\tau \in (0,1)$.

[A3] There exist some positive values δ^*, b_1, b_2 and $s_3 > 0$, such that the conditional probability density $f_{Y|\mathbf{X}}(.|.)$ of Y given $\mathbf{X}$ belongs to $H_{s_3}(\mathcal{D})$ and is uniformly bounded away from zero in $(Q_\tau(\mathbf{x}) - b_1, Q_\tau(\mathbf{x}) + b_2)$ for all $\tau \in [\delta^*, 1-\delta^*]$ and $\mathbf{x} \in \mathcal{D}$.

The order of smoothness $s_1,\ s_2,\ s_3$ will be specified later. The above assumptions are standard in local polynomial smoothing for quantile regression; see, for example, Chaudhuri et al (1997) and Kong et al (2010). Among them, [A2] implies that for any $\mathbf{x} \in \mathcal{D}$ and $\mathbf{X}_i \in S_n(\mathbf{x}) = \{i : 1 \le i \le n,\ |\mathbf{X}_{i\mathbf{x}}| \le h_n\}$, the error from approximating $Q_\tau(\mathbf{X}_i)$ by the $k(=[s_2])th$ order Taylor expansion

$$Q_n(\mathbf{X}_i, \mathbf{x}; \tau) = [\mathbf{X}_{i\mathbf{x}}(h_n, A)]^\top \mathbf{c}_n(\mathbf{x}; \tau)$$

is of order $O(h_n^{s_2})$, uniformly in $\{(\mathbf{x}, \mathbf{X}_i) : \mathbf{x} \in \mathcal{D},\ \mathbf{X}_i \in S_n(\mathbf{x})\}$ and $\tau \in (0,1)$. [A3] strengthens CONDITION 3 of Chaudhuri et al (1997), where it is required that for a pre-specified τ, $g(\mathbf{x}|\tau) = f_{Y|\mathbf{X}}(Q_\tau(\mathbf{x})|\mathbf{x}) > 0$, for all $\mathbf{x} \in \mathcal{D}$.

The following lemma concerns the strong uniform Bahadur type representation of $\hat{\mathbf{c}}_n(.;\tau)$ derived from (3.3).

LEMMA 2. *Suppose [A1]-[A3] hold with $s_1 > 0,\ s_2 > 0,\ s_3 > 1/2$, and $k = [s_2]$. The bandwidth h_n is chosen such that*

$$h_n \propto n^{-\kappa} \quad \textit{with} \quad \frac{1}{2(s_2+p)} \le \kappa < \frac{1}{p}.$$

Then we have with probability one,

$$\hat{\mathbf{c}}_n(\mathbf{x};\tau) - \mathbf{c}_n(\mathbf{x};\tau) = -\frac{\Sigma_n^{-1}(\mathbf{x};\tau)}{N_n(\mathbf{x})} \sum_{i \in S_n(\mathbf{x})} \mathbf{X}_{i\mathbf{x}}(h_n, A)[I\{Y_i \le Q_n(\mathbf{X}_i, \mathbf{x}; \tau)\} - \tau] \qquad (4.1) \quad + O\Big\{\Big(\frac{\log n}{n h_n^p}\Big)^{3/4}\Big\}$$

uniformly in $\tau \in [\delta^*, 1-\delta^*]$ and $\mathbf{x} = \mathbf{X}_1, \cdots, \mathbf{X}_n$, where $N_n(\mathbf{x}) = \sharp S_n(\mathbf{x})$ and

$$\Sigma_n(\mathbf{x};\tau) = \mathbf{E}_i\Big[g(\mathbf{X}_i|\tau)\mathbf{X}_{i\mathbf{x}}(h_n, A)\mathbf{X}_{i\mathbf{x}}^\top(h_n, A)|\mathbf{X}_i \in S_n(\mathbf{x})\Big].$$

This strengthens the results obtained in Chaudhuri (1991) for nonparametric quantile regression and Kong et al (2010) for general nonparametric M-regression, both of which concerned the uniformity in $\mathbf{x}$ only. The uniformity in both $\mathbf{x}$ and τ plays a central role in examining the asymptotic properties of $\hat{\Sigma}_{\mathrm{T}}$, defined via averaging over $\mathbf{x} = \mathbf{X}_1, \cdots, \mathbf{X}_n$, and then integration with respect to τ over $[\delta^*, 1-\delta^*]$.

We now move on to present the asymptotic properties of $\hat{\Sigma}_{\mathrm{T}}$ and those of its eigenvalues and eigenvectors. Write $\nabla^2 Q_\tau(.)$ for the Hessian matrix of $Q_\tau(.)$ and $\nabla g(.|\tau)$, for the first order derivative vector of $g(.|\tau)$. For any $\tau \in (0,1)$ and $1 \le k, l \le p$, let $\nabla Q_\tau^{[k]}(\mathbf{X})$ stand for the kth element of $\nabla Q_\tau(\mathbf{X})$; $\nabla^{[l]} g(\mathbf{X}|\tau)$ for the lth element of $\nabla g(\mathbf{X}|\tau)$, $\nabla^2_{[k,l]} Q_\tau(\mathbf{X})$ for the (k,l) element of $\nabla^2 Q_\tau(.)$ and write

$$\rho(\mathbf{X}|\tau,k,l) = \Big[\frac{2\nabla^2_{[k,l]} Q_\tau(\mathbf{X})}{g(\mathbf{X}|\tau)} - \frac{\nabla Q_\tau^{[k]}(\mathbf{X})\nabla^{[l]} g(\mathbf{X}|\tau)}{g^2(\mathbf{X}|\tau)} - \frac{\nabla^{[l]} Q_\tau(\mathbf{X})\nabla^{[k]} g(\mathbf{X}|\tau)}{g^2(\mathbf{X}|\tau)}\Big].$$

For any $\tau_1, \tau_2 \in (0,1)$ and $1 \le k_1, l_1, k_2, l_2 \le p$, define

$$\begin{aligned} h(\tau_1,\tau_2|k_1,l_1,k_2,l_2) \quad &= \quad Cov\Big(\nabla Q_{\tau_1}^{(k_1)}(\mathbf{X})\nabla Q_{\tau_1}^{(l_1)}(\mathbf{X}), \nabla Q_{\tau_2}^{(k_2)}(\mathbf{X})\nabla Q_{\tau_2}^{(l_2)}(\mathbf{X})\Big) \\ &\quad +\{\min(\tau_1,\tau_2) - \tau_1\tau_2\}\mathrm{Cov}\Big(\rho(\mathbf{X}|\tau,k_1,l_1), \rho(\mathbf{X}|\tau,k_2,l_2)\Big). \end{aligned}$$

For any symmetric $p \times p$ matrix $\mathcal{S} = (s_{ij})$, form a $p(p+1)/2 \times 1$ vector using the elements of $\mathcal{S}$:

$$\mathrm{Vech}(\mathcal{S}) = (s_{11}, \cdots, s_{p1}, s_{22}, \cdots, s_{2p}, s_{22}, \cdots, s_{pp})^\top.$$

Denote by $v(.)$ the following 1-to-1 mapping from $\{1, 2, \cdots, p(p+1)/2\}$ onto $\{(i,j) : 1 \le i \le j \le p\}$:

$$v(k) = (v(k,1), v(k,2)) = (i,j) \text{ such that } \frac{(2p-i)(i-1)}{2} + j = k.$$

In other words, the kth element of $\mathrm{Vech}(\mathcal{S})$ is given by $s_{v(k)} = s_{v(k,1),v(k,2)}$.

Finally, for any symmetric $p \times p$ matrix $\mathcal{S}$, denote by $\lambda_k(\mathcal{S})$ and $\beta_k(\mathcal{S})$, $k = 1, \cdots, q$, the first q (nonzero) eigen-values and eigen-vectors of $\mathcal{S}$ respectively. Write $\tilde{\lambda}_{p-q}(\mathcal{S})$ for the average of the smallest $p-q$ eigenvalues of $\mathcal{S}$.

THEOREM 1. *Suppose [A1]-[A3] hold with $s_1 > 0$, $s_3 > 1/2$, $s_2 > 3/2p + 3$, and $k = [s_2]$. Furthermore, the smoothing parameter h_n is chosen such that*

$$h_n \propto n^{-\kappa} \quad \textit{with } \frac{1}{2(s_2-1)} \le \kappa < \frac{1}{3p+4}. \tag{4.2}$$

Then we have $\tilde{\lambda}_{p-q}(\hat{\Sigma}_{\mathrm{T}}) = o_p(n^{-1/2})$ *and*

$$\sqrt{n}(\hat{\Sigma}_{\mathrm{T}} - \Sigma_{\mathrm{T}}) \xrightarrow{d} \mathbb{N}, \tag{4.3}$$

where '$\xrightarrow{d}$*' stands for convergence in distribution and* $\mathbb{N}$ *stands for a symmetric* $p \times p$ *random matrix, such that* $Vech(\mathbb{N})$ *is multivariate normal with zero mean and covariance matrix* $\mathbf{H}$*, whose* (k,l)*th element is given by*

$$\int_0^1 \int_0^1 h(\tau_1, \tau_2 | v(k,1), v(k,2), v(l,1), v(l,2)) \; d\tau_1 d\tau_2.$$

Furthermore, if $\lambda_k(\Sigma_{\mathrm{T}})$*,* $k = 1, \cdots, q$*, are all distinct, then for each* $k = 1, \cdots, q$*,*

$$\sqrt{n}\{\lambda_k(\hat{\Sigma}_{\mathrm{T}}) - \lambda_k(\Sigma_{\mathrm{T}})\} \xrightarrow{d} \beta_k^\top(\Sigma_{\mathrm{T}})\mathbb{N}\beta_k(\Sigma_{\mathrm{T}}), \tag{4.4}$$

$$\sqrt{n}\{\beta_k(\hat{\Sigma}_{\mathrm{T}}) - \beta_k(\Sigma_{\mathrm{T}})\} \xrightarrow{d} \sum_{l=1, l\neq k}^{q} \frac{\beta_l(\Sigma_{\mathrm{T}})\beta_l^\top(\Sigma_{\mathrm{T}})\mathbb{N}\beta_k(\Sigma_{\mathrm{T}})}{\lambda_k(\Sigma_{\mathrm{T}}) - \lambda_l(\Sigma_{\mathrm{T}})}. \tag{4.5}$$

In theory, (4.4) could be applied to make inference on the structural dimension q. The proof of Theorem 1 is mainly based upon results on U-processes (Nolan and Pollard, 1987), namely a collection of U-statistics indexed by a family of symmetric kernels.

5. Bandwidth selection. As far as the point-wise estimation of $\nabla Q_\tau(.)$ is concerned, it followed from Lemma 2 that the 'optimal' bandwidth h_n which minimizes the point-wise mean square error (MSE) of $\hat{\nabla} Q_\tau(\mathbf{x})$, is of the order $O(n^{-1/(p+2k+2)})$. In this sense, the choice (4.2) of the bandwidth h_n under-smooths the estimator. Such undersmoothing is necessary for the estimator $\hat{\nabla} Q_\tau(\mathbf{x})$ to have a bias of order $o_p(n^{-1/2})$ thus negligible. The stochastic term of $\hat{\nabla} Q_\tau(\mathbf{x})$, once averaged over $\mathbf{x} = \mathbf{X}_1, \cdots, \mathbf{X}_n$, can achieve the rate of $O_p(n^{-1/2})$, independent of the speed at which h_n tends to zero. Similar observations have been made in Chaudhri et al (1997) and Kong et al (2013). In cases where the link function $Q_\tau(.)$ closely resembles a (local) polynomials, the bias thus becomes less of an issue as it either significantly reduces or completely vanishes; we can then afford to employ a larger bandwidth thus produce more efficient estimates of $\nabla Q_\tau(.)$, while results in Theorem 1 still hold. This also explains our assertion in Section 1 that qOPG is structure-adaptive. In practice, an empirical 'optimal' bandwidth can be obtained by plugging in estimates for the unknown quantities in the formula of the point-wise theoretical 'optimal' bandwidth.

We can also select bandwidth based on the cross-validation (CV) criterion for quantile regression; see, for example, Al-kenani and Yu (2010). This is carried out as follows. For any given $\tau \in (0,1)$ and fixed h_n, denote by $Q_\tau^{\setminus j}(\mathbf{x}|h_n)$, $j = 1, \cdots, n$, the leave-one-out estimate of $Q_\tau(\mathbf{X}_j)$ using $\{(X_i, Y_i) : i \neq j\}$ with bandwidth h_n. Let

$$CV(\tau, h_n) = n^{-1} \sum_{j=1}^{n} \rho_\tau \Big(Y_j - Q_\tau^{\setminus j}(X_j|h_n) \Big),$$

and denote by h_τ^{CV}, the level-specific cross-validated (CV) bandwidth, namely the h_n that minimizes $CV(\tau, h_n)$. However, based on our experience with simulated data, we found such level-specific CV bandwidth selection is not only rather time-consuming, but also terribly unstable, possibly due to the difficulty in assessing the goodness-of-fit in quantile regression; see Koenker and Machado (1999). Instead, We recommend the following modified level-specific CV bandwidth. First, consider an average of the level-specific CV bandwidth h_τ^{CV} with τ ranging over the set of $\{\tau_s = s/(T+1) : s = 1, ..., T\}$ for some positive integer T:

$$\bar{h}^{CV} = \sum_{s=1}^{T} h_{\tau_s}^{CV} / T.$$

Then in view of the relationship proposed in Yu and Jones (1998), we define the modified level-specific CV bandwidth as

$$\bar{h}_\tau^{CV} = \bar{h}^{CV} \{ \tau(1-\tau)/\phi(\Phi^{-1}(\tau)) \}^{1/5}, \tag{5.1}$$

where functions $\phi(.)$ and $\Phi(.)$ are respectively the probability and cumulative distribution functions of the standard normal distribution. Compared to h_τ^{CV}, $\bar{h}_\tau^{CV}$ is more stable and delivers much better results, but its computation is equally computationally intensive. We also tried out variations of $\bar{h}_\tau^{CV}$ defined as in (5.1) but with $\bar{h}^{CV}$ replaced by bandwidths chosen via other procedures. Our best experience lies with $\bar{h}_\tau^{CV}$ with $\bar{h}^{CV}$ set to be the CV bandwidth for conditional mean regression of $|Y - E(Y)|$ on $\mathbf{X}$.

6. Estimation of the structural dimension.. According to Theorem 1, the average of the smallest $p-q$ eigenvalues of $\hat{\Sigma}_{\mathrm{T}}$ defined in (3.5) is of order $o_p(n^{-1/2})$. For $k = 1, \cdots, p$, plot the average of the smallest k eigenvalues of $\hat{\Sigma}_{\mathrm{T}}$ against k and likely values for q could be then identified by noting the location of a noticeable increase. The asymptotic distribution of the eigenvalues of $\hat{\Sigma}_{\mathrm{T}}$ given in Theorem 1 could also be used for selecting q.

However, as the distribution depends on another unknown matrix $\mathbf{H}$ which is not easy to estimate, such approach might not be very practical.

Combining the CV method of Xia et al (2002) with the composite quantile regression provides an alternative way to select q. For illustration purposes, we here give details for the local constant quantile kernel smoothing. With working dimension q, suppose the $q-$columns of $\hat{B}_q$ are the corresponding estimates of the CS directions. For each observation $(X_j, Y_j), j = 1, ..., n$, calculate the delete-one-estimator of $\tilde{Q}_\tau(\hat{B}_q X_j)$ of (2.1) as

$$\hat{Q}_\tau^{\setminus j}(\hat{B}_q^\top X_j) = arg\min_c \sum_{i \neq j} \rho_\tau(Y_i - c) K_{h_n}(|\hat{B}_q X_{ij}|).$$

We then define the CV value specific to working dimension q as

$$CV(q) = \int_{\delta^*}^{1-\delta^*} \sum_{j=1}^{n} \rho_\tau(Y_i - \hat{Q}_\tau^{\setminus j}(\hat{B}_q^\top X_j)) d\tau,$$

and choose the dimension which minimizes $CV(q)$. Our simulation study suggests that this methodology works reasonably well, though it is also rather computationally intensive.

7. Numerical study. In this section, we first carry out comparison studies of the two newly proposed procedures, qOPG and qMAVE, with two existing methods using simulated data. The two new procedures are then applied to the analysis of a real data set for the purpose of discovering the dimension reduction space.

In the calculation below, the local linear quantile regression, i.e. $k = 1$, and the Epanechnikov kernel function are used. The integrations in (3.6) and (3.8) are evaluated by the weighted summation of $\hat{\Sigma}(\tau)$ over $\tau = 0.1, 0.2, ..., 0.9$.

Example 1 (Simulated data). We reconsider the following three models that are commonly tested out in the field of dimension reduction

$$\begin{aligned} \text{Model (A)}: \quad & Y = \mathbf{x}_1(\mathbf{x}_1 + \mathbf{x}_2 + 1) + 0.5\varepsilon, \\ \text{Model (B)}: \quad & Y = \mathbf{x}_1/(0.5 + (\mathbf{x}_2 + 1.5)^2) + 0.5\varepsilon, \\ \text{Model (C)}: \quad & Y = \mathbf{x}_1 + \exp(\mathbf{x}_2)\varepsilon, \end{aligned}$$

where $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_{10})^\top \sim N(0, (\sigma_{ij})_{1 \leq i,j \leq 10})$ with $\sigma_{ij} = 0.5^{|i-j|}$, and ε is the error term designed to have various distributions; see Table 1 below. The first two models were thoughtfully designed by Li (1991) for the study of Slice Inverse Regression (SIR). Model 3 was used in Xia (2007) in the context of conditional mean and conditional variance based dimension reduction.

Based on the conclusion of Ma and Zhu (2012) from their intensive comparison study using simulated data, we have chosen to compare our conditional quantile-based approaches, qOPG and qMAVE, with dOPG and dMAVE of (Xia, 2007), among the many existing dimension reduction procedures. Another reason for us to include dOPG and dMAVE in the study is the fact that these conditional probability-based approaches, are theoretical equivalences to qOPG and qMAVE. We hope through such comparison can manifest the structure-adaptive nature of our new methods. We also include in the comparison study the SIR of Li (1991), for which 8 slices are used when the sample size $n = 200$, and 10 when the sample size $n = 400$. For dOPG and dMAVE, following the rule-of-thumb as in Xia (2007), we use bandwidths of order $n^{-1/5}$ and $n^{-1/(p+4)}$ respectively for the two kernels in the estimation. For qOPG and dMAVE, the bandwidth is chosen as described in Section 5. For any estimator $\hat{\mathbf{B}}$ of $\mathbf{B}_0$, we define the estimation error as the largest among the absolute values of the elements of $\hat{\mathbf{B}}(\hat{\mathbf{B}}^\top\hat{\mathbf{B}})^{-1}\hat{\mathbf{B}} - \mathbf{B}_0(\mathbf{B}_0^\top\mathbf{B}_0)^{-1}\mathbf{B}_0$. Table 1 reports the mean and standard error (in brackets) of the estimation error from 100 replicates for various combinations of model, error distribution and sample size. The last column of Table 1 is the percentage of times that the structural dimension has been correctly identified by the CV method described in Section 6.

A a general observation is such that qOPG and qMAVE - either with data-driven bandwidth or with a bandwidth chosen according to the rule-of-thumb - outperform respectively dOPG and dMAVE as well as SIR for both models (A) and (B). The only exception lies with model (C), where qOPG using the rule-of-thumb bandwidth is beaten by dOPG, but the situation reverses with a data-driven bandwidth. This provides a line of empirical evidence for the assertion we made in Section 1 that if the conditional quantile function is well approximated locally by polynomials, then the data-driven bandwidth deduced from qOPG means more efficient estimators. Another noticeable pattern is that, contradictory to what happens with conditional density-based methods where dMAVE consistently outperforms dOPG, the expected superiority of qMAVE over qOPG is nowhere obvious. In fact, for models (A) and (B), qOPG outperforms qMAVE most of the time, especially so when data-driven bandwidths are used. Even for model (C) qMAVE seems to enjoy an obvious lead over qOPG, this again becomes less obvious when a data-driven bandwidth is used. A plausible explanation for this might be that an adaptive-weighting scheme has been incorporated into qOPG, while such procedure is hard to be combined with qMAVE.

Table 1. Average estimation errors and their standard derivation (in parenthesis) and frequency of correct structural dimension identification

						qOPG		qMAVE		
Model	ε	n	SIR	dOPG	dMAVE	h_0	h_{CV}	h_0	h_{CV}	freq.
(A)	N(0,1)	200	0.82	0.55	0.53	0.42	0.44	0.48	0.48	56%
			(0.14)	(0.20)	(0.18)	(0.15)	(0.15)	(0.16)	(0.15)	
		400	0.68	0.37	0.35	0.27	0.26	0.31	0.30	90%
			(0.16)	(0.14)	(0.10)	(0.08)	(0.08)	(0.08)	(0.08)	
	$t(3)/\sqrt{3}$	200	0.79	0.50	0.46	0.42	0.38	0.38	0.40	72%
			(0.15)	(0.22)	(0.16)	(0.15)	(0.14)	(0.14)	(0.14)	
		400	0.63	0.31	0.29	0.22	0.21	0.23	0.24	97%
			(0.16)	(0.13)	(0.08)	(0.07)	(0.07)	(0.06)	(0.06)	
	$\chi^2(1)$	200	0.78	0.61	0.50	0.48	0.49	0.46	0.49	50%
			(0.13)	(0.22)	(0.17)	(0.20)	(0.19)	(0.17)	(0.17)	
		400	0.61	0.39	0.32	0.30	0.28	0.28	0.29	79%
			(0.14)	(0.16)	(0.10)	(0.12)	(0.09)	(0.09)	(0.10)	
(B)	N(0,1)	200	0.69	0.58	0.59	0.44	0.50	0.54	0.52	56%
			(0.17)	(0.17)	(0.18)	(0.18)	(0.19)	(0.19)	(0.19)	
		400	0.51	0.35	0.38	0.24	0.27	0.32	0.32	87%
			(0.15)	(0.10)	(0.13)	(0.10)	(0.10)	(0.11)	(0.12)	
	$t(3)/\sqrt{3}$	200	0.57	0.48	0.47	0.38	0.37	0.40	0.40	84%
			(0.16)	(0.16)	(0.15)	(0.16)	(0.12)	(0.15)	(0.13)	
		400	0.41	0.34	0.29	0.19	0.18	0.21	0.22	97%
			(0.12)	(0.10)	(0.09)	(0.09)	(0.06)	(0.06)	(0.07)	
	$\chi^2(1)$	200	0.64	0.57	0.53	0.55	0.46	0.51	0.48	64%
			(0.17)	(0.18)	(0.20)	(0.24)	(0.20)	(0.22)	(0.19)	
		400	0.42	0.35	0.31	0.24	0.22	0.24	0.25	94%
			(0.11)	(0.13)	(0.09)	(0.11)	(0.08)	(0.07)	(0.07)	
(C)	N(0,1)	200	0.53	0.55	0.51	0.77	0.42	0.48	0.36	29%
			(0.13)	(0.14)	(0.17)	(0.15)	(0.14)	(0.17)	(0.10)	
		400	0.37	0.36	0.33	0.77	0.29	0.30	0.24	31%
			(0.08)	(0.11)	(0.09)	(0.16)	(0.10)	(0.08)	(0.05)	
	$t(3)/\sqrt{3}$	200	0.61	0.62	0.59	0.81	0.47	0.55	0.38	32%
			(0.15)	(0.14)	(0.18)	(0.15)	(0.19)	(0.19)	(0.14)	
		400	0.44	0.41	0.38	0.77	0.39	0.35	0.25	39%
			(0.12)	(0.14)	(0.15)	(0.15)	(0.20)	(0.14)	(0.07)	
	$\chi^2(1)$	200	0.63	0.60	0.49	0.50	0.46	0.44	0.42	37%
			(0.14)	(0.15)	(0.16)	(0.17)	(0.16)	(0.16)	(0.13)	
		400	0.43	0.42	0.32	0.35	0.30	0.31	0.27	46%
			(0.11)	(0.14)	(0.09)	(0.10)	(0.16)	(0.09)	(0.08)	

Example 2 (Real data). In financial economics, the capital asset pricing model (CAPM) indicates that the return of a portfolio strongly depends on the market performance. However, little is known about the factors that affect the volatility of a portfolio. In the following, we consider the daily return Y of a portfolio listed at

$$http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\_library.html$$

with covariate $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{15})$, where $\mathbf{x}_1, ..., \mathbf{x}_5$ are the returns of the portfolio in the past five days, and $\mathbf{x}_6, ..., \mathbf{x}_{10}$ are the absolute values of the returns which are proxy of the past volatilities; $\mathbf{x}_{11}, ..., \mathbf{x}_{15}$ are the market returns on the same day as Y and those in the past four days, and $\mathbf{x}_{16}, ..., \mathbf{x}_{20}$ are the absolute values of the market returns.

Applying qOPG, the first several eigenvalues of $\hat{\Sigma}_{\mathrm{T}}$ are respectively 1.0620, 0.0164, 0.0017, 0.0007 and 0.0004. With the structural dimension set as 2, we obtain the estimated CS directions β_1 and β_2; see Table 2. The scatter plots of Y against $\beta_1^\top \mathbf{X}$ and $\beta_2^\top \mathbf{X}$ are given in Fig. 1. The fitted curve in the bottom two panels are created with bandwidths $h = h_0/(\hat{f}_k(x))^{0.2}$ with h_0 being selected by the CV method, and $\hat{f}_k(.),\ k = 1, 2$, being the kernel estimate of the density function of $\beta_k^\top \mathbf{X}$. The fitted regression function of the portfolio's return on $\beta_1^\top \mathbf{X}$ in the bottom-left panel of Fig. 1 suggests the first CS direction β_1 is mostly about the conditional mean, while the second CS direction β_2 is clearly about the conditional variance, evident from the bottom-right panel. The first direction β_1 is dominated by $\mathbf{x}_{11}$, the market return of the day, with a coefficient 0.9940; this is in line with the CAPM in that the expected return of any portfolio largely depends on the present-day market performance. It is also interesting to note that the volatility of the portfolio also depends the market's volatility, as suggested by the large coefficients of $\mathbf{x}_{16}, \mathbf{x}_{17}$ and $\mathbf{x}_{18}$ on the second CS direction β_2. Also, its own past volatilities $(\mathbf{x}_8, \mathbf{x}_9)$ also contribute to its present-day volatility, although to a less extent.

Table 2: Estimated CS directions for Example 2

$\mathbf{x}_i$	β_1	β_2	$\mathbf{x}_i$	β_1	β_2	$\mathbf{x}_i$	β_1	β_2	$\mathbf{x}_i$	β_1	β_2
$\mathbf{x}_1$	-0.014	0.001	$\mathbf{x}_6$	0.006	-0.089	$\mathbf{x}_{11}$	0.994	-0.032	$\mathbf{x}_{16}$	0.029	0.490
$\mathbf{x}_2$	-0.042	0.045	$\mathbf{x}_7$	0.005	0.093	$\mathbf{x}_{12}$	0.048	-0.027	$\mathbf{x}_{17}$	-0.017	0.506
$\mathbf{x}_3$	-0.029	-0.239	$\mathbf{x}_8$	0.020	0.271	$\mathbf{x}_{13}$	0.048	-0.076	$\mathbf{x}_{18}$	0.008	0.302
$\mathbf{x}_4$	-0.008	-0.100	$\mathbf{x}_9$	0.008	0.277	$\mathbf{x}_{14}$	0.035	0.347	$\mathbf{x}_{19}$	-0.035	-0.126
$\mathbf{x}_5$	0.008	0.067	$\mathbf{x}_{10}$	-0.014	0.111	$\mathbf{x}_{15}$	-0.005	0.010	$\mathbf{x}_{20}$	0.001	-0.120

8. Conclusions. In this paper, we have proposed and investigated two composite quantile approach to dimension reduction, namely qOPG and qMAVE. Compared with moment-based methods, these methods require less restrictive assumptions and can identify all dimension reduction directions. It does not involve 'slicing' of the response variable Y, as is the case with SIR or conditional density-based methods (Xia, 2007). It carries out regression analysis directly on Y instead of transformations of Y. As a result of these characteristics, qOPG and qMAVE are structure-adaptive and thus more efficient. However, because the amount of computation embedded in quantile regression is significantly heavier than in least square minimization, the implementation of qOPG and qMAVE is rather time consuming compared to most of the existing methods. Because of this, we recommend the use of dOPG or dMAVE to obtain an initial estimator of the central subspace and of the structural dimension, and the use of qOPG or qMAVE for more efficient refined estimator.

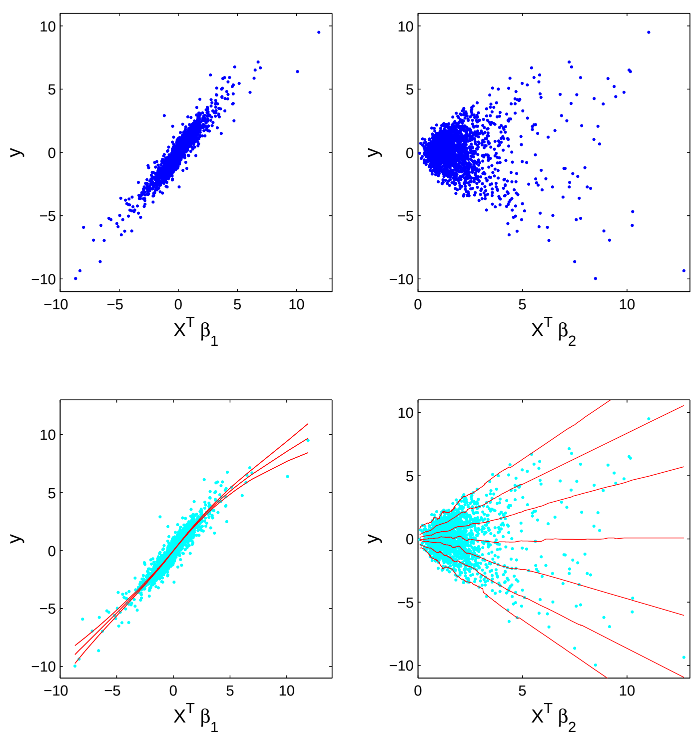

FIG 1. *Results for Example 2. The top two panels are the scatter plots of Y against the two estimated CS directions β_1 and β_2. The bottom-left panel is the fitted regression function of Y against the first CS direction and its 95% confidence interval. In the bottom-right panel, the curves are the regression quantiles of Y against the second directions at $\tau = 0.01, 0.1, 0.3, 0.5, 0.7, 0.9, 0.99$ respectively.*

Acknowledgements. We thank the Editor, the Associate Editor and two referees for their helpful comments that have improved earlier versions of this paper.

APPENDIX: PROOFS

Proof of Lemma 1. The assertion that $\mathcal{S}(\Sigma) \subseteq \mathcal{S}(\mathbf{B}_0)$ follows directly from (2.3). We show that the opposite holds too. Based on (2.3), we can see by definition

$$\Sigma = \mathbf{B}_0 \Big[\int_0^1 E\{\nabla \tilde{Q}_\tau(\mathbf{B}_0^\top \mathbf{X})[\nabla \tilde{Q}_\tau(\mathbf{B}_0^\top \mathbf{X})]^\top\} d\tau \Big] \mathbf{B}_0^\top.$$

It thus suffices if we can prove the matrix

$$M = \int_0^1 E\{\nabla \tilde{Q}_\tau(\mathbf{B}_0^\top \mathbf{X})[\nabla \tilde{Q}_\tau(\mathbf{B}_0^\top \mathbf{X})]^\top\} d\tau$$

is of full rank. For if otherwise, there must exist some vector $\mathbf{b}_1 \in R^q$, with Euclidean norm one such that $\mathbf{b}_1^\top M \mathbf{b}_1 = 0$. Seeing the definition of M, this implies that

$$\mathbf{b}_1^\top \nabla \tilde{Q}_\tau(\mathbf{B}_0^\top \mathbf{X}) = 0 \ a.s. \tag{a.1}$$

for all $\tau \in (0, 1)$ except on a set of Lebesgue measure zero.

Let $\mathbf{B} = (\mathbf{b}_1, \cdots, \mathbf{b}_q) \in R^{q\times q}$ denote an orthonormal basis for R^q, i.e. $\mathbf{B}^\top \mathbf{B} = \mathbf{I}_q$. For any given $\tau \in (0, 1)$, write

$$G_\tau(\mathbf{u}) = \tilde{Q}_\tau(\mathbf{u}), \quad \tilde{G}_\tau(\mathbf{u}) = \tilde{Q}_\tau(\mathbf{B}\mathbf{u}), \quad \tilde{\mathbf{B}}_0 = \mathbf{B}_0\mathbf{B}. \tag{a.2}$$

Thus

$$G_\tau(\mathbf{B}\mathbf{u}) = \tilde{G}_\tau(\mathbf{u}); \quad G_\tau(\mathbf{B}_0^\top \mathbf{X}) = G_\tau(\tilde{\mathbf{B}}_0^\top \mathbf{X}).$$

Consider the gradient vector of $\tilde{G}_\tau(\mathbf{u})$ and then evaluate it for $\mathbf{u} = \tilde{\mathbf{B}}_0^\top \mathbf{X}$:

$$\frac{\partial \tilde{G}_\tau(\mathbf{u})}{\partial \mathbf{u}} = \frac{\partial G_\tau(\mathbf{B}\mathbf{u})}{\partial \mathbf{u}} = \mathbf{B}^\top \frac{\partial G_\tau(\mathbf{B}\mathbf{u})}{\partial(\mathbf{B}\mathbf{u})} = \mathbf{B}^\top \nabla G_\tau(\mathbf{B}\mathbf{u}) \overset{\mathbf{u}=\tilde{\mathbf{B}}_0^\top \mathbf{X}}{=} \mathbf{B}^\top \nabla G_\tau(\mathbf{B}_0^\top \mathbf{X}),$$

the first element of which, according to (a.1), equals zero. This suggests the value of $\tilde{G}_\tau(\tilde{\mathbf{B}}_0^\top \mathbf{X})$, as a function of $\tilde{\mathbf{B}}_0^\top \mathbf{X} = (\mathbf{b}_1^\top \mathbf{B}_0^\top \mathbf{X}, \cdots, \mathbf{b}_q^\top \mathbf{B}_0^\top \mathbf{X})^\top$, does not change with $\mathbf{b}_1^\top \mathbf{B}_0^\top \mathbf{X}$. This together with the fact that

$$\tilde{G}_\tau(\tilde{\mathbf{B}}_0^\top \mathbf{X}) = G_\tau(\mathbf{B}_0^\top \mathbf{X}) = \tilde{Q}_\tau(\mathbf{B}_0^\top \mathbf{X}) = Q_\tau(\mathbf{X})$$

implies that $Q_\tau(\mathbf{X})$ is in fact a function of $q-1$ variables: $\mathbf{b}_2^\top \mathbf{B}_0^\top \mathbf{X}, \cdots, \mathbf{b}_q^\top \mathbf{B}_0^\top \mathbf{X}$. And this according to (a.1) holds for any $\tau \in (0, 1)$. As $\{Q_\tau(\mathbf{X}) : \tau \in (0, 1)\}$

collectively defines $F(.|\mathbf{X})$, we can conclude that $F(.|\mathbf{X})$ is in fact a function of $(\mathbf{b}_2^\top \mathbf{B}_0^\top \mathbf{X}, \cdots, \mathbf{b}_q^\top \mathbf{B}_0^\top \mathbf{X})^\top = [\mathbf{B}_0(\mathbf{b}_2, \cdots, \mathbf{b}_q)]^\top \mathbf{X}$, expressed as

$$F(Y|\mathbf{X}) = F(Y|\tilde{\mathbf{B}}^\top \mathbf{X}),\ a.s. \text{ where } \tilde{\mathbf{B}} = \mathbf{B}_0(\mathbf{b}_2, \cdots, \mathbf{b}_q).$$

This means $S(\tilde{\mathbf{B}})$ is SDR and as $S(\mathbf{B}_0)$ is the CS, we should have $S(\mathbf{B}_0) \subseteq S(\tilde{\mathbf{B}})$. This contradicts the fact that $dim(S(\mathbf{B}_0)) = q > q-1 = dim(S(\tilde{\mathbf{B}}))$. □

The proof of Lemma 2 is left until the end. To prove Theorem 1, we also need to introduce more notations. For any $\mathbf{t} = (t_1, \cdots, t_p)^\top \in [-1,1]^p$, let $\mathbf{t}(A)$ stand for the $s(A) \times 1$ vector $(\mathbf{t}^{\mathbf{u}})_{\mathbf{u} \in A}$. Define

$$\Gamma = \int_{[-1,1]^p} \mathbf{t}(A)\{\mathbf{t}(A)\}^\top d\mathbf{t}.$$

Standard result in kernel smoothing, e.g. Masry (1996), is such that with probability one,

$$\frac{N_n(\mathbf{x})}{nh_n^p} - f_{\mathbf{X}}(\mathbf{x}) = O(h_n^2 + (nh_n^p/\log n)^{-1/2}) \tag{a.3}$$

uniformly in $\mathbf{x} \in \mathcal{D}$, and

$$\Sigma_n(\mathbf{x};\tau) - g(\mathbf{x}|\tau)\Gamma = O\Big((nh_n^p/\log n)^{-1/2} + h_n\Big) \tag{a.4}$$

uniformly in $\tau \in (0,1)$ and $\mathbf{x} \in \mathcal{D}$.

Also we will cite the following result, the proof of which will be given at the end of this section: with probability one,

$$\begin{aligned} \sum_i \quad & \mathbf{X}_{i\mathbf{x}}(h_n, A) I(|\mathbf{X}_{i\mathbf{x}}| \le h_n)\Big[I\{Y_i \le Q_n(\mathbf{X}_i, \mathbf{x}; \tau)\} \\ & -I\{Y_i \le Q_\tau(\mathbf{X}_i)\}\Big] = o(n^{-1/2}) \end{aligned} \tag{a.5}$$

uniformly in $\mathbf{x} \in \mathcal{D},\ \tau \in (0,1)$.

Proof of Theorem 1. Write as $\tilde{\Gamma}_n(\mathbf{X}_j;\tau)$, the $p \times s(A)$ matrix consisting the second up to the $(p+1)th$ row of $\Sigma_n^{-1}(\mathbf{X}_j;\tau)$. First note that under conditions in Theorem 1,

$$h_n^{-1}(nh_n^p/\log n)^{-3/4} = o(n^{-1/2}),\ h_n^{s_2-1} = o(n^{-1/2}),\ \text{and}\ \log n/(nh_n^p) = o(n^{-1/2}h_n).$$

This together with (3.4) and Lemma 2 leads to

$$\hat{\Sigma}(\tau) \;=\; \frac{1}{n}\sum_{j=1}^n \nabla Q_\tau(\mathbf{X}_j)\{\nabla Q_\tau(\mathbf{X}_j)\}^\top + h_n^{-1}[M_n(\tau) + M_n^\top(\tau)] + o(n^{-1/2}),$$

where

$$M_n(\tau) = \frac{1}{n}\sum_{i,j} \frac{\nabla Q_\tau(\mathbf{X}_j)}{N_n(\mathbf{X}_j)} I(|\mathbf{X}_{ij}| \le h_n)[I\{Y_i \le Q_n(\mathbf{X}_i, \mathbf{X}_j;\tau)\} - \tau]\mathbf{X}_{ij}^\top(h_n, A)\tilde{\Gamma}_n^\top(\mathbf{X}_j;\tau).$$

with $\mathbf{X}_{ij} = \mathbf{X}_i - \mathbf{X}_j$. Using results in (a.3), (a.4) and (a.5), we have

$$\hat{\Sigma}(\tau) = \frac{1}{n}\sum_{j=1}^{n} \nabla Q_\tau(\mathbf{X}_j)\{\nabla Q_\tau(\mathbf{X}_j)\}^\top + h_n^{-(p+1)}[\tilde{M}_n(\tau)\tilde{\Gamma}^\top + \tilde{\Gamma}\tilde{M}_n^\top(\tau)] + o(n^{-1/2}), \tag{a.6}$$

where $\tilde{\Gamma}$ is the $p \times s(A)$ matrix consisting of the second up to the $p+1$th rows of Γ^{-1} and

$$\tilde{M}_n(\tau) = \frac{1}{n^2}\sum_{i,j} \frac{\nabla Q_\tau(\mathbf{X}_j)\mathbf{X}_{ij}^\top(h_n, A)}{g(\mathbf{X}_j|\tau)f_{\mathbf{X}}(\mathbf{X}_j)}[I\{Y_i \le Q_\tau(\mathbf{X}_i)\} - \tau]I(|\mathbf{X}_{ij}| \le h_n).$$

The key to the study of the properties of $\hat{\Sigma}(\tau)$ is $\{\tilde{M}_n(\tau) : \tau \in (0,1)\}$, which is a typical example of $U-$processes (Nolan and Pollard, 1987).

To derive the Hoeffding's decomposition of $\tilde{M}_n(\tau)$, write $\mathbf{Z}_i = (Y_i, \mathbf{X}_i)$ and define

$$\begin{aligned}\xi_n(\mathbf{Z}_i, \mathbf{Z}_j; \tau) &= \Big\{\frac{\nabla Q_\tau(\mathbf{X}_j)\mathbf{X}_{ij}^\top(h_n, A)}{g(\mathbf{X}_j|\tau)f_{\mathbf{X}}(\mathbf{X}_j)}[I\{Y_i \le Q_\tau(\mathbf{X}_i)\} - \tau] \\ &\quad + \frac{\nabla Q_\tau(\mathbf{X}_i)\mathbf{X}_{ji}^\top(h_n, A)}{g(\mathbf{X}_i|\tau)f_{\mathbf{X}}(\mathbf{X}_i)}[I\{Y_j \le Q_\tau(\mathbf{X}_j)\} - \tau]\Big\}I(|\mathbf{X}_{ij}| \le h_n), \\ \zeta_n(\mathbf{Z}_i; \tau) &= E_j[\xi_n(\mathbf{Z}_i, \mathbf{Z}_j; \tau)] = h_n^p[I\{Y_i \le Q_\tau(\mathbf{X}_i)\} - \tau]\Big\{\frac{\nabla Q_\tau(\mathbf{X}_i)}{g(\mathbf{X}_i|\tau)}\gamma^\top + \\ &\quad h_n\Big[\frac{\nabla^2 Q_\tau(\mathbf{X}_i)}{g(\mathbf{X}_i|\tau)} - \frac{\nabla Q_\tau(\mathbf{X}_i)\nabla^\top g(\mathbf{X}_i|\tau)}{g^2(\mathbf{X}_i|\tau)}\Big]\Gamma_1 + O(h_n^2)\Big\},\end{aligned} \tag{a.7}$$

where

$$\gamma = \int_{[-1,1]^p} \mathbf{t}(A)d\mathbf{t}, \quad \Gamma_1 = \int \mathbf{t}\ \mathbf{t}^\top(A)d\mathbf{t}.$$

Note that $E[\xi_n(\mathbf{Z}_i, \mathbf{Z}_j; \tau)] = E[\zeta_n(\mathbf{Z}_i; \tau)] = 0$. Therefore, we have

$$\tilde{M}_n(\tau) = \frac{1}{n^2}\sum_{i<j} \xi_n(\mathbf{Z}_i, \mathbf{Z}_j; \tau) = U_n(\tau) + \frac{1}{n}\sum_i \zeta_n(\mathbf{Z}_i; \tau),$$

where $U_n(\tau)$ is its Hoeffding's decomposition

$$U_n(\tau) = \frac{1}{n^2}\sum_{i<j} \xi_n(\mathbf{Z}_i, \mathbf{Z}_j; \tau) - \frac{1}{n}\sum_i \zeta_n(\mathbf{Z}_i; \tau). \tag{a.8}$$

To decide the tail properties of $\sup\{|U_n(\tau)| : \tau \in [\delta^*, 1-\delta^*]\}$, first note that according to Lemma (2.13) of Pakes and Pollard (1989) [reproduced as [C1] at the end of this section] and Corollary A.3, $\{\xi_n(\mathbf{Z}_i, \mathbf{Z}_j; \tau) : \tau \in [\delta^*, 1-\delta^*]\}$

is Euclidean for a constant envelope, or in Arcones (1995) term, a uniformly bounded V-C subgraph class. Applying Proposition 4 in Arcones (1995) to $U_n(\tau)$, we conclude that there exists some finite $c_2 > 0$, such that for any $\epsilon > 0$,

$$P\{n^{1/2} \sup_{\tau\in[\delta^*,1-\delta^*]} |U_n(\tau)| \geq h_n^{p+1}\epsilon\} \leq 2\exp\{-c_2\epsilon n^{1/2}h_n^{-1}\}.$$

By an application of the Borel-Cantelli Lemma, we have

$$\sup_{\tau\in[\delta^*,1-\delta^*]} |U_n(\tau)| = o(n^{-1/2}h_n^{p+1}) \quad a.s.$$

This together with (a.6), (a.7), (a.8) and the facts that $\tilde{\Gamma}\gamma = \mathbf{0}$, $\tilde{\Gamma}\Gamma_1 = \mathbf{I}_p$ implies that with probability one,

$$\hat{\Sigma}(\tau) = \frac{1}{n}\sum_{i=1}^{n} \nabla Q_\tau(\mathbf{X}_i)\{\nabla Q_\tau(\mathbf{X}_i)\}^\top + \frac{1}{n}\sum_{i=1}^{n} \frac{[I\{Y_i \leq Q_\tau(\mathbf{X}_i)\} - \tau]}{g^2(\mathbf{X}_i|\tau)} \times$$
$$\Big[2g(\mathbf{X}_i|\tau)\nabla^2 Q_\tau(\mathbf{X}_i) - \nabla Q_\tau(\mathbf{X}_i)\nabla^\top g(\mathbf{X}_i|\tau) - \nabla g(\mathbf{X}_i|\tau)\nabla^\top Q_\tau(\mathbf{X}_i)\Big] + o(n^{-1/2})$$

where the term $o(n^{-1/2})$ is uniform in $\tau \in [\delta^*, 1-\delta^*]$. Consequently, we have

$$\begin{aligned} \hat{\Sigma}_{\mathrm{T}} = \int_{\delta^*}^{1-\delta^*} \hat{\Sigma}(\tau)d\tau &= \Sigma_{\mathrm{T}} + \frac{1}{n}\sum_{i=1}^{n} \Sigma^{(1)}(\mathbf{X}_i) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} \Sigma^{(2)}(\mathbf{X}_i, Y_i) + o(n^{-1/2}),\ a.s. \end{aligned} \tag{a.9}$$

where $\Sigma^{(1)}(.)$ and $\Sigma^{(2)}(.)$ are two symmetric random matrices with properties such that

$$E[\Sigma^{(1)}(\mathbf{X})] = \mathbf{0},\ E[\Sigma^{(2)}(\mathbf{X}, Y)] = \mathbf{0},\ \Sigma^{(1)}(\mathbf{X})\Pi = \mathbf{0},\ \Sigma^{(2)}(\mathbf{X}, Y)\Pi = \mathbf{0},$$

with $\Pi = \mathbf{I} - \mathbf{B}_0(\mathbf{B}_0^\top\mathbf{B}_0)^{-1}\mathbf{B}_0^\top$, the projection matrix such that $\Pi\mathbf{B}_0 = \mathbf{B}_0^\top\Pi = \mathbf{0}$. An application of Lemma A.1 in Li (1991) to the right-hand side of (a.9) with Σ_{T}, $n^{-1/2}$, $\hat{\Sigma}_{\mathrm{T}}$ and $n^{-1/2}\{\sum_i\{\Sigma^{(1)}(\mathbf{X}_i) + \Sigma^{(2)}(\mathbf{X}_i, Y_i)\}$ acting as T, w^2, $T(w)$ and $T^{(2)}$ therein respectively, we have with probability one,

$$\tilde{\lambda}_{p-q}(\hat{\Sigma}_{\mathrm{T}}) = \frac{n^{-1/2}}{p-q}\sum_i trace([\Sigma^{(1)}(\mathbf{X}_i) + \Sigma^{(2)}(\mathbf{X}_i, Y_i)]\Pi) + o(n^{-1/2}) = o(n^{-1/2})$$

We now move on to derive the asymptotic properties of the first q eigen-values and eigen-vectors of $\hat{\Sigma}$. First note that the three classes of functions, namely $\{\nabla Q_\tau(\mathbf{X}_i)\{\nabla Q_\tau(\mathbf{X}_i)\}^\top,\ \tau \in [\delta^*, 1-\delta^*]\}$, $\{g(\mathbf{X}_i|\tau)\}^{-2}[I\{Y_i \leq Q_\tau(\mathbf{X}_i)\}-\tau],\ \tau \in [\delta^*, 1-\delta^*]\}$, and $\{g(\mathbf{X}_i|\tau)\nabla^2 Q_\tau(\mathbf{X}_i) - \nabla Q_\tau(\mathbf{X}_i)\nabla^\top g(\mathbf{X}_i|\tau) -$

$\nabla g(\mathbf{X}_i|\tau)\nabla^\top Q_\tau(\mathbf{X}_i), \tau \in [\delta^*, 1-\delta^*]\}$ are, according to Corollary A.3, all Euclidean for a constant envelop. Therefore, the collection of random matrices $\{\hat{\Sigma}(\tau) : \tau \in [\delta^*, 1-\delta^*]\}$ are *Glivenko-Cantelli* as well as *Donsker* [van der Vaart and Wellner (2000)].

By *Glivenko-Cantelli*, we mean that

$$\sup_{\tau\in[\delta^*,1-\delta^*]} |\text{Vech}(\hat{\Sigma}(\tau)) - \text{Vech}(\Sigma(\tau))| \to 0 \ a.s.,$$

from which we can conclude that

$$\text{Vech}(\hat{\Sigma}_{\text{T}}) - \text{Vech}(\Sigma_{\text{T}}) \to 0 \ a.s.$$

which in turn implies that [Lemma 3.1, Bai et al (1991)],

$$\beta_k(\hat{\Sigma}_{\text{T}}) - \beta_k(\Sigma_{\text{T}}) \to 0 \ \ (k = 1, \cdots, q) \ a.s.$$

By *Donsker*, we mean that

$$\sqrt{n}\{\text{Vech}(\hat{\Sigma}(\tau)) - \text{Vech}(\Sigma(\tau))\} \xrightarrow{d} \mathbb{G}, \quad \text{in } \ell^\infty([\delta^*, 1-\delta^*]),$$

where $\ell^\infty([\delta^*, 1-\delta^*])$ stands for the space of all uniformly bounded multivariate real functions from $[\delta^*, 1-\delta^*]$ to $R^{p(p+1)/2}$ equipped with the supremum norm, and the limit $\mathbb{G}$ is a zero-mean $p(p+1)/2-$dimensional Gaussian process on $[\delta^*, 1-\delta^*]$, such that for any given $\tau_1, \tau_2 \in [\delta^*, 1-\delta^*]$, the covariance matrix $E[\mathbb{G}(\tau_1)\mathbb{G}(\tau_2)]$ has its (k,l)th element given by the covariance between

$$\nabla Q_{\tau_1}^{[v(k,1)]}(\mathbf{X})\nabla Q_{\tau_1}^{[v(k,2)]}(\mathbf{X}) + \tfrac{[I\{Y_i \le Q_{\tau_1}(\mathbf{X}_i)\} - \tau_1]}{g^2(\mathbf{X}_i|\tau_1)}\Big[2g(\mathbf{X}_i|\tau_1)]\nabla^2_{[v(k,1),v(k,2)]}Q_{\tau_1}(\mathbf{X}_i)$$
$$-\nabla Q_{\tau_1}^{[v(k,1)]}(\mathbf{X}_i)\nabla^{[v(k,2)]}g(\mathbf{X}_i|\tau_1) - \nabla^{[v(k,1)]}g(\mathbf{X}_i|\tau_1)\nabla^{[v(k,2)]}Q_{\tau_1}(\mathbf{X}_i)\Big]$$

and

$$\nabla Q_{\tau_2}^{[[v(l,1)]]}(\mathbf{X})\nabla Q_{\tau_2}^{[v(l,2)]}(\mathbf{X}) + \tfrac{[I\{Y_i \le Q_{\tau_2}(\mathbf{X}_i)\} - \tau_2]}{g^2(\mathbf{X}_i|\tau_2)}\Big[2g(\mathbf{X}_i|\tau_2)]\nabla^2_{[v(l,1),v(k,2)]}Q_{\tau_2}(\mathbf{X}_i)$$
$$-\nabla Q_{\tau_2}^{[v(l,1)]}(\mathbf{X}_i)\nabla^{[v(l,2)]}g(\mathbf{X}_i|\tau_2) - \nabla^{[v(l,1)]}g(\mathbf{X}_i|\tau_2)\nabla^{[v(l,2)]}Q_{\tau_2}(\mathbf{X}_i)\Big];$$

equation (4.3) thus follows by appealing to the Continuous-mapping Theorem.

The proof of (4.4) and (4.5), i.e. the asymptotic normality of the eigenvalues and eigenvectors of $\hat{\Sigma}$, can be done in exactly the same manner as in Theorem 2.2 of Zhu and Fang (1996), which, by an application of the Perturbation Theory [Sun (1988), Kato (1995)], relates the asymptotic normality of a random matrix to that of its eigenvalues and eigenvectors. □

To prepare for the proof of Lemma 2, we need to introduce more notations and some related results. For any given $\mathbf{x} \in \mathcal{D}$ let $DX_n(\mathbf{x})$ be the $N_n(\mathbf{x}) \times$

$s(A)$ matrix with rows given by the transposition of $\mathbf{X}_{i\mathbf{x}}(h_n, A)$, $i \in S_n(\mathbf{x})$, and $VY_n(\mathbf{x})$ be the $N_n(\mathbf{x}) \times 1$ vector whose components are Y_i, $i \in S_n(\mathbf{x})$.

For any subset $\mathbf{h} \subset S_n(\mathbf{x})$, denote by $DX_n(\mathbf{x}, \mathbf{h})$ and $VY_n(\mathbf{x}, \mathbf{h})$, the submatrix (vector) of $DX_n(\mathbf{x})$ and $VY_n(\mathbf{x})$, respectively, with indices of rows given by by $\mathbf{h}$. Further define

$$\mathbf{H}_n(\mathbf{x}) = \{\mathbf{h} : \mathbf{h} \subset S_n(\mathbf{x}),\ \sharp(\mathbf{h}) = s(A), DX_n(\mathbf{x}, \mathbf{h}) \text{ is of full rank}\}.$$

Suppose $DX_n(\mathbf{x})$ of rank $= s(A)$, $\mathbf{H}_n(\mathbf{x})$ is thus nonempty. The following two facts concern the uniqueness of $\hat{\mathbf{c}}_n(\mathbf{x}; \tau)$ and its 'matrix form' of , for any given $\mathbf{x} \in \mathcal{D}$ and $\tau \in (0, 1)$. They are essentially restatements of Theorems 3.1 and 3.2 in Koenker and Bassett (1978); see, also FACT 6.3 and 6.4 in Chaudhuri (1991b).

[B1] There exists positive constants c_1 and c_2, such that

$$P(A_n) = 1, \text{ where } A_n = \{c_1 n h_n^d \le N_n(\mathbf{x}) \le c_2 n h_n^d \text{ for all } \mathbf{x} \in \mathcal{D}\}$$

This follows easily from (a.4).

[B2] There exists a $\mathbf{h} \in \mathbf{H}_n(\mathbf{x})$, such that (3.3) has at least one solution of the form

$$\hat{\mathbf{c}}_n(\mathbf{x}; \tau) = [DX_n(\mathbf{x}, \mathbf{h})]^{-1} VY_n(\mathbf{x}, \mathbf{h}).$$

[B3] For $\mathbf{h} \in \mathbf{H}_n(\mathbf{x})$, let $\hat{\mathbf{c}}_n(\mathbf{x}; \tau) = [DX_n(\mathbf{x}, \mathbf{h})]^{-1} VY_n(\mathbf{x}, \mathbf{h})$ and define

$$L_n(\mathbf{h}; \mathbf{x}, \tau) = [DX_n(\mathbf{x}, \mathbf{h})]^{-1} \sum_{i \in \bar{\mathbf{h}}} \Big[I\{Y_i < \mathbf{X}_{i\mathbf{x}}^\top (h_n, A) \hat{\mathbf{c}}_n(\mathbf{x}; \tau)\} - \tau \Big] \mathbf{X}_{i,\mathbf{x}}(h_n, A),$$

where $\bar{\mathbf{h}}$ is the relative complement of $\mathbf{h}$ with respect to $S_n(\mathbf{x})$. Then $\hat{\mathbf{c}}_n(\mathbf{x}; \tau)$ is a unique solution to (3.3) if and only if $L_n(\mathbf{h}; \mathbf{x}, \tau) \in (\tau - 1, \tau)^{s(A)}$. Further, if $\hat{\mathbf{c}}_n(\mathbf{x}; \tau)$ is a solution (not necessarily unique) to (3.3), we must have $L_n(\mathbf{h}; \mathbf{x}, \tau) \in [\tau - 1, \tau]^{s(A)}$.

To facilitate the use of the conditioning arguments at various places in the proofs, for any $\mathbf{X}_j,\ j = 1, \cdots, n$, we exclude $\mathbf{X}_j$ from the previously defined $S_n(\mathbf{X}_j)$; instead we define $S_n(\mathbf{X}_j) = \{i : 1 \le i \le n,\ i \ne j,\ |\mathbf{X}_{ij}| \le h_n\}$ and $N_n(\mathbf{X}_j) = \sharp(S_n(\mathbf{X}_j))$.

The proof of Lemma 2 will be built upon the following slightly weaker result.

LEMMA A.1. *Let $\delta_n = (nh_n^p / \log n)^{-1/2}$. Suppose conditions in Lemma 2 hold. Then*

$$\sup_{1 \le j \le n, \tau \in [\delta^*, 1-\delta^*]} |\hat{\mathbf{c}}_n(\mathbf{X}_j; \tau) - \mathbf{c}_n(\mathbf{X}_j; \tau)| = O(\delta_n) \ a.s.$$

Proof of Lemma A.1. For any given positive constant K_1 and a generic $\mathbf{x} \in \mathcal{D}$, let U_n be the event defined as

$$U_n = \Big\{ \sup_{\tau\in[\delta^*,1-\delta^*]} |\hat{\mathbf{c}}_n(\mathbf{x};\tau) - \mathbf{c}_n(\mathbf{x};\tau)| \geq K_1\delta_n \Big\}. \tag{a.10}$$

In view of the fact that $P(A_n) = 1$, the assertion in Lemma A.1 will follow from an application of the Borel-Cantelli lemma, if we can show that there exists some $K_1 > 0$, such that

$$\sum_n nP(U_n \cap A_n) < \infty. \tag{a.11}$$

We now try to get an upper bound for $P(U_n \cap A_n)$. To this end, for given $\tau \in [\delta^*, 1-\delta^*], \mathbf{x} \in \mathcal{D}$ and $\mathbf{c} \in R^{s(A)}$, define

$$Z_{ni}(\mathbf{c}|\mathbf{x},\tau) = \Big[I\{Y_i < \mathbf{c}^\top \mathbf{X}_{i\mathbf{x}}(h_n, A)\} - \tau \Big] \mathbf{X}_{i,\mathbf{x}}(h_n, A).$$

Based on [B2] and [B3], there exists some positive constant K_2, which depends only on $s(A)$ such that $U_n \cap A_n$ is contained in the event

$$\begin{aligned}&\{\text{there exists some } \tau \in [\delta^*, 1-\delta^*] \text{ and } \mathbf{h} \in \mathbf{H}_n(\mathbf{x}), \text{ such that for}\\ &\hat{\mathbf{c}}_n(\mathbf{x};\tau) = [DX_n(\mathbf{x},\mathbf{h})]^{-1} VY_n(\mathbf{x},\mathbf{h}), \text{ we have } |\sum_{i\in\bar{\mathbf{h}}} Z_{ni}(\hat{\mathbf{c}}_n(\mathbf{x};\tau)|\mathbf{x},\tau)| \leq K_2,\\ &\text{and } |\hat{\mathbf{c}}_n(\mathbf{x};\tau) - \mathbf{c}_n(\mathbf{x};\tau)| \geq K_1\delta_n\} \cap A_n.\end{aligned} \tag{a.12}$$

Choose large enough K_1 such that we can apply Proposition A.2 to conclude that there exist some $\epsilon_1 > 0$, and $K_3 > 0$, such that, for all $\tau \in [\delta^*, 1-\delta^*]$,

$$E[Z_{ni}(\hat{\mathbf{c}}_n(\mathbf{x};\tau)|\mathbf{x},\tau)] \geq \min\{\epsilon_1, K_3K_1\delta_n\},$$

and consequently as a result of A_n and the fact that $\sharp(\bar{\mathbf{h}}) = N_n(\mathbf{x}) - s(A)$, we have

$$\begin{aligned}&\Big\{ \Big| \sum_{i\in\bar{\mathbf{h}}} Z_{ni}(\hat{\mathbf{c}}_n(\mathbf{x};\tau)|\mathbf{x},\tau) \Big| \leq K_2 \Big\}\\ &\subseteq \Big\{ \Big| \sum_{i\in\bar{\mathbf{h}}} \{ Z_{ni}(\hat{\mathbf{c}}_n(\mathbf{x};\tau)|\mathbf{x},\tau) - E[Z_{ni}(\hat{\mathbf{c}}_n(\mathbf{x};\tau)|\mathbf{x},\tau)] \} \Big| \geq c_1^* K_1 n h_n^p \delta_n \Big\}\end{aligned} \tag{a.13}$$

for some $c_1^* > 0$.

Next, note that given the set $S_n(\mathbf{x})$, $\mathbf{h} \subset S_n(\mathbf{x})$, and $(\mathbf{X}_i, Y_i)$ for $i \in \mathbf{h}$, and thus $\hat{\mathbf{c}}_n(\mathbf{x};\tau) = [DX_n(\mathbf{x},\mathbf{h})]^{-1}VY_n(\mathbf{x},\mathbf{h})$ is also fixed, the random vectors $\{Z_{ni}(\hat{\mathbf{c}}_n(\mathbf{x};\tau)|\mathbf{x},\tau),\ i \in \bar{\mathbf{h}}\}$ are conditionally i.i.d. This together with (a.12), (a.13) and the fact that $\sharp(\mathbf{H}_n(\mathbf{x}))$ is of order $(nh_n^p)^{s(A)}$, implies there exists some $c_2^* > 0$, such that

$$\begin{aligned}P(U_n \cap A_n) \;&\leq\; c_2^*(nh_n^p)^{s(A)}\\ &\times P\Big\{ \sup_{\substack{\tau\in[\delta^*,1-\delta^*],\\ \mathbf{c}\in R^{s(A)}}} \Big| \sum_{i\in\bar{\mathbf{h}}} \{ Z_{ni}(\mathbf{c}|\mathbf{x},\tau) - E[Z_{ni}(\mathbf{c}|\mathbf{x},\tau)] \} \Big| \geq c_1^* K_1 n h_n^p \delta_n \Big\}.\end{aligned} \tag{a.14}$$

To find a bound for the probability on the right hand side above, first note that according to Lemma 22 (ii) in Nolan and Pollard (1987), $\{Z_{ni}(\mathbf{c}|\mathbf{x},\tau) : \tau \in [\delta^*, 1-\delta^*], \mathbf{c} \in R^{s(A)}\}$ is contained in an Euclidean class for a constant envelope, since $Y_i - \mathbf{c}^\top \mathbf{X}_{i\mathbf{x}}(h_n, A) = [\mathbf{X}_{i\mathbf{x}}^\top(h_n, A), Y_i] * (\mathbf{c}^\top, -1)^\top$ and the indicator function $I(. < 0)$ is of bounded variation. As $E|Z_{ni}(\mathbf{c}|\mathbf{x},\tau) Z_{ni}^\tau(\mathbf{c}|\mathbf{x},\tau)|^2 = O(1)$ uniformly in $\tau \in [\delta^*, 1-\delta^*], \mathbf{c} \in R^{s(A)}$, through similar arguments used in the proof of Theorem 2.37 in Pollard (1984, pp.34), we have that

$$P\Big\{ \sup_{\substack{\tau \in [\delta^*, 1-\delta^*], \\ \mathbf{c} \in R^{s(A)}}} \Big| \sum_{i \in \bar{\mathbf{h}}} \{Z_{ni}(\mathbf{c}|\mathbf{x},\tau) - E[Z_{ni}(\mathbf{c}|\mathbf{x},\tau)]\} \Big| \geq c_1^* K_1 n h_n^p \delta_n \Big\} = o(n^{-a}),$$

for any $a > 0$. This together with (a.14) leads to (a.11). □

For any $\mathbf{x} \in \mathcal{D}$, let $\omega_{h_n}(\mathbf{t}|\mathbf{x})$ be the conditional probability density function of $(\mathbf{X}_i - \mathbf{x})/h_n$ given $i \in S_n(\mathbf{x})$. Note that it converges to the uniform density on $[-1,1]^p$ uniformly in $\mathbf{t} \in [-1,1]^p$ and $\mathbf{x} \in \mathcal{D}$.

Proof of Lemma 2. For any given $\tau \in [\delta^*, 1-\delta^*]$, $\mathbf{x} \in \mathcal{D}$, and $\mathbf{X} \in S_n(\mathbf{x})$, write

$$\hat{Q}_n(\mathbf{X}, \mathbf{x}; \tau) = [(\mathbf{X} - \mathbf{x})(h_n, A)]^\top \hat{\mathbf{c}}_n(\mathbf{x}; \tau).$$

The proof consists of the following steps.

Step 1: For any given $\tau \in [0,1]$, $\mathbf{c} \in R^{s(A)}$ and $\mathbf{x} \in R^p$, define

$$\begin{aligned} \tilde{H}_n(\mathbf{c}; \mathbf{x}) &= E[I\{Y_i < c^\top \mathbf{X}_{i\mathbf{x}}(h_n, A)\} \mathbf{X}_{i\mathbf{x}}(h_n, A) | i \in S_n(\mathbf{x})] \\ &= \int_{[-1,1]^p} F(\mathbf{c}^\top \mathbf{t}(A) | \mathbf{x} + h_n \mathbf{t}) \mathbf{t}(A) \omega_{h_n}(\mathbf{t}|\mathbf{x}) d\mathbf{t}, \\ R_n^{(1)}(\tilde{\mathbf{c}}, \mathbf{c}|\mathbf{x}, \tau) &= \tilde{H}_n(\mathbf{x}, \tilde{\mathbf{c}}) - \tilde{H}_n(\mathbf{x}, \mathbf{c}) - \Sigma_n(\mathbf{x}; \tau)(\tilde{\mathbf{c}} - \mathbf{c}). \end{aligned}$$

Therefore, under assumptions [A2] and [A3],

$$\begin{aligned} & R_n^{(1)}(\hat{\mathbf{c}}_n(\mathbf{x};\tau), \mathbf{c}_n(\mathbf{x};\tau)|\mathbf{x},\tau) \\ \text{(a.15)} &= \tilde{H}_n(\mathbf{x}, \hat{\mathbf{c}}_n(\mathbf{x};\tau)) - \tilde{H}_n(\mathbf{x}, \mathbf{c}_n(\mathbf{x};\tau)) - \Sigma_n(\mathbf{x};\tau)[\hat{\mathbf{c}}_n(\mathbf{x};\tau) - \mathbf{c}_n(\mathbf{x};\tau)], \\ &= \int_{[-1,1]^p} [F(\hat{Q}_n(\mathbf{x} + h_n\mathbf{t}, \mathbf{x}; \tau)|\mathbf{x} + h_n\mathbf{t}) - F(Q_n(\mathbf{x} + h_n\mathbf{t}, \mathbf{x}; \tau)|\mathbf{x} + h_n\mathbf{t}) \\ & \quad - g(\mathbf{x} + h_n\mathbf{t}|\tau)\mathbf{t}(A)\mathbf{t}^\top(A)\{\hat{\mathbf{c}}_n(\mathbf{x};\tau) - \mathbf{c}_n(\mathbf{x};\tau)\}] w_{h_n}(\mathbf{t}|\mathbf{x}) d\mathbf{t} \\ \text{(a.16)} &= O(\delta_n^{1+s_3}) = O\{[n^{(1-\kappa p)}/\log n]^{-3/4}\}, \ (\text{if } s_3 \geq 1/2), \end{aligned}$$

uniformly in $\tau \in [\delta^*, 1-\delta^*]$, where (a.16) follows from Lemma A.1 and the facts that $\hat{Q}_n(\mathbf{x} + h_n\mathbf{t}, \mathbf{x}; \tau) - Q_n(\mathbf{x} + h_n\mathbf{t}, \mathbf{x}; \tau) = \{\mathbf{t}(A)\}^\top [\hat{\mathbf{c}}_n(\mathbf{x};\tau) - \mathbf{c}_n(\mathbf{x};\tau)]$ and $Q_n(\mathbf{x} + h_n\mathbf{t}, \mathbf{x}; \tau) - Q_\tau(\mathbf{x} + h_n\mathbf{t}) = O(h_n^{s_2}) = o(\delta_n)$.

Step 2: For any given $\tau \in (0,1)$, $\mathbf{x} \in R^p$ and $\mathbf{h} \in H_n(\mathbf{x})$, define

$$\chi_n(\mathbf{x};\tau) = \sum_{i\in S_n(\mathbf{x})} [\mathbf{X}_{i\mathbf{x}}(h_n,A) I\{Y_i \le \hat{Q}_n(\mathbf{X}_i,\mathbf{x};\tau)\} - \tilde{H}_n(\hat{\mathbf{c}}_n(\mathbf{x};\tau);\mathbf{x})] - \sum_{i\in S_n(\mathbf{x})} [\mathbf{X}_{i\mathbf{x}}(h_n,A) I\{Y_i \le Q_n(\mathbf{X}_i,\mathbf{x};\tau)\} - \tilde{H}_n(\mathbf{c}_n(\mathbf{x};\tau),\mathbf{x})],$$

$$\hat{\mathbf{c}}_n^{\mathbf{h}}(\mathbf{x};\tau) = [DX_n(\mathbf{x},\mathbf{h})]^{-1} VY_n(\mathbf{x},\mathbf{h}),\ \hat{Q}_n^{\mathbf{h}}(\mathbf{X}_i,\mathbf{x};\tau) = \{\hat{\mathbf{c}}_n^{\mathbf{h}}(\mathbf{x};\tau)\}^\top \mathbf{X}_{i\mathbf{x}}(h_n,A),$$

and for any $\mathbf{c}_1,\ \mathbf{c}_2 \in R^{s(A)}$, define

$$\chi_n^{\mathbf{h}}(\mathbf{c}_1,\mathbf{c}_2;\mathbf{x}) = \sum_{i\in\bar{\mathbf{h}}} [\mathbf{X}_{i\mathbf{x}}(h_n,A) I\{Y_i \le \mathbf{c}_1^\top \mathbf{X}_{i\mathbf{x}}(h_n,A)\} - \tilde{H}_n(\mathbf{c}_1;\mathbf{x})] - \sum_{i\in\bar{\mathbf{h}}} [\mathbf{X}_{i\mathbf{x}}(h_n,A) I\{Y_i \le \mathbf{c}_2^\top \mathbf{X}_{i\mathbf{x}}(h_n,A)\} - \tilde{H}_n(\mathbf{c}_2;\mathbf{x})].$$

For any given $K_3 > 0$, consider the corresponding event

$$W_n(\mathbf{x}) = \Big\{ \sup_{\tau\in[\delta^*,1-\delta^*]} |\chi_n(\mathbf{x};\tau)| \ge K_3 [\log n]^{3/4} n^{(1-\kappa p)/4} \Big\}.$$

Then in view of definition of the events A_n, $U_n(\mathbf{x})$ of (a.10) and [B2], the event $W_n(\mathbf{x}) \cap A_n \cap \overline{U_n(\mathbf{x})}$ [$\overline{U_n(\mathbf{x})}$ is the complement of $U_n(\mathbf{x})$] is contained in the event

$$\{\text{for some } \tau \in [\delta^*, 1-\delta^*] \text{ and } \mathbf{h} \in \mathbf{H}_n(\mathbf{x}),\ |\chi_n^{\mathbf{h}}(\hat{\mathbf{c}}_n^{\mathbf{h}}(\mathbf{x};\tau), \mathbf{c}_n(\mathbf{x};\tau);\mathbf{x})| \ge K_4 [\log n]^{3/4} n^{(1-\kappa p)/4} \text{and } |\hat{\mathbf{c}}_n^{\mathbf{h}}(\mathbf{x}) - \mathbf{c}_n(\mathbf{x};\tau)| \le K_1 \delta_n\} \cap A_n$$

for large enough n, where $K_4 = K_3/2$ and for which we have implicitly used the facts that $\sharp(\mathbf{h}) = p$ and $[\log n]^{3/4} n^{(1-\kappa p)/4} \to \infty$ as $n \to \infty$. Again, since $\sharp(H_n(\mathbf{x}))$ is of order $n^{(1-\kappa p)n(A)}$ uniformly in $\mathbf{x} \in \mathcal{D}$, there exists some constant $c_3 > 0$, such that $P(W_n(\mathbf{x}) \cap A_n \cap \overline{U_n(\mathbf{x})})$ is bounded by $c_3 n^{(1-\kappa p)n(A)}$ multiplied by the probability of the following event

$$\Big\{ \sup_{\substack{\mathbf{c}_1,\mathbf{c}_2 \in R^{s(A)};\\ |\mathbf{c}_1-\mathbf{c}_2| \le K_1\delta_n}} |\chi_n^{\mathbf{h}}(\mathbf{c}_1,\mathbf{c}_2;\mathbf{x})| \ge K_4 [\log n]^{3/4} n^{(1-\kappa p)/4} \Big\} \cap A_n. \tag{a.17}$$

To find a bound for the probability of even (a.17), first note that according to Lemma 22 (ii) in Nolan and Pollard (1987) and Lemma (2.14) (i) in Pakes and Pollard (1989), the class of all functions on $R^{s(A)+1}$ of the form

$$(Y_i, \mathbf{X}_{i\mathbf{x}}(h_n,A)) \to \mathbf{X}_{i\mathbf{x}}(h_n,A)[I\{Y_i \le \mathbf{c}_1^\top \mathbf{X}_{i\mathbf{x}}(h_n,A)\} - I\{Y_i \le \mathbf{c}_2^\top \mathbf{X}_{i\mathbf{x}}(h_n,A)\}]$$

with $\mathbf{c}_1, \mathbf{c}_2$ ranging over $R^{s(A)}$ is again an Euclidean class for a constant envelope. Secondly,conditioning on $S_n(\mathbf{x})$, $\mathbf{h} \in H_n(\mathbf{x})$, and observations $\{(\mathbf{X}_i, Y_i) : i \in \mathbf{h}\}$, the terms in the sum defining $\chi_n^{\mathbf{h}}(\mathbf{c}_1, \mathbf{c}_2; \mathbf{x})$ are i.i.d. with mean zero, and variance-covariance matrix with Euclidean norm of order $O(|\mathbf{c}_1 - \mathbf{c}_2|)$. Following the steps in the proof of Theorem 2.37 in Pollard (1984, pp.34), we can conclude that there exist constant $c_4 > 0, c_5 > 0$, such that the probability of (a.17) is bounded by

$$K_4^{c_4}(\log n)^{c_4/2}\exp(-c_5 K_4^2 \log n) = o(n^{-\alpha}),\ \text{for any } \alpha > 0,$$

if K_4, or equivalently K_3, is chosen to be sufficiently large. Equivalently we have there exists some K_3, such that

$$P\Big\{\sup_{\tau\in[\delta^*,1-\delta^*]} |\chi_n(\mathbf{x};\tau)| \geq K_3[\log n]^{3/4} n^{(1-\kappa p)/4}\Big\} = o(n^{-2}).$$

An application of Borel-Cantelli lemma leads to

$$\sup_{\tau\in[\delta^*,1-\delta^*],j=1,\cdots,n} |\chi_n(\mathbf{X}_j;\tau)| = O\{(\log n)^{3/4} n^{(1-\kappa p)/4}\}\ a.s. \tag{a.18}$$

Step 3: Combining (a.15), (a.16) and (a.18), we have with probability one,

$$\begin{aligned}
&\frac{1}{N_n(\mathbf{x})}\sum_{i\in S_n(\mathbf{x})} \mathbf{X}_{i\mathbf{x}}(h_n, A)[I\{Y_i \leq Q_n(\mathbf{X}_i,\mathbf{x};\tau)\} - \tau] \\
= & -\frac{1}{N_n(\mathbf{x})}\chi_n^{\mathbf{h}}(\mathbf{x}) - \tilde{H}_n(\hat{\mathbf{c}}_n(\mathbf{x};\tau);\mathbf{x}) + \tilde{H}_n(\mathbf{c}_n(\mathbf{x};\tau);\mathbf{x}) \\
& + \frac{1}{N_n(\mathbf{x})}\sum_{i\in S_n(\mathbf{x})} \mathbf{X}_{i\mathbf{x}}(h_n, A)[I\{Y_i \leq \hat{Q}_n(\mathbf{X}_i,\mathbf{x};\tau)\} - \tau] \\
= & -\Sigma_n(\mathbf{x};\tau)[\hat{\mathbf{c}}_n(\mathbf{x};\tau) - \mathbf{c}_n(\mathbf{x};\tau)] + O\{[n^{(1-\kappa p)}/\log n]^{-3/4}\} \\
& + \frac{1}{N_n(\mathbf{x})}\sum_{i\in S_n(\mathbf{x})} \mathbf{X}_{ij}(\delta_n, A)[I\{Y_i \leq \hat{Q}_n(\mathbf{X}_i,\mathbf{x};\tau)\} - \tau]
\end{aligned} \tag{a.19}$$

uniformly in $\tau \in [\delta^*, 1-\delta^*]$ and $\mathbf{x} = \mathbf{X}_j,\ j = 1, \cdots, n$. Note that according to [B3], the last term in (a.19) is of order $O(n^{\kappa p-1}) = o\{[n^{(1-\kappa p)}/\log n]^{-3/4}\}$. □

PROPOSITION A.2. *There exists some $K_2 > 0, K_3 > 0, K_4 > 0$ such that for all $\tau \in [\delta^*, 1-\delta^*]$,*

$$\Big|\int_{[-1,1]^p} \{F(\mathbf{c}^\top t(A)|\mathbf{x} + h_n\mathbf{t}) - \tau\} t(A)\omega_{h_n}(\mathbf{t}|\mathbf{x})d\mathbf{t}\Big| \geq \min\{K_2, K_3|\mathbf{c} - \mathbf{c}_n(\mathbf{x};\tau)|\},$$

whenever $|\mathbf{c} - \mathbf{c}_n(\mathbf{x};\tau)| \geq K_4 h_n^{s_2}$.

Proof of Proposition A.2. First note that as $\omega_{h_n}(\mathbf{t}|\mathbf{x})$ converges to the uniform density on $[-1,1]^p$ uniformly in $\mathbf{t}\in[-1,1]^p,\ \mathbf{x}\in\mathcal{D}$, we have

$$\int_{[-1,1]^p}\{F(\mathbf{c}^\top t(A)|\mathbf{x}+h_n\mathbf{t})-\tau\}t(A)\omega_{h_n}(\mathbf{t}|\mathbf{x})d\mathbf{t}=H_n(\mathbf{c}|\mathbf{x},\tau)(1+o(1))$$
$$\text{where } H_n(\mathbf{c}|\mathbf{x},\tau)=\int_{[-1,1]^p}\{F(\mathbf{c}^\top t(A)|\mathbf{x}+h_n\mathbf{t})-\tau\}t(A)d\mathbf{t}.$$

The proof is split into the following steps.

Steps 1: We show that there exist $M_1>0$ and $\epsilon_1>0$, such that for all $\tau\in[\delta^*,1-\delta^*]$, and $\mathbf{x}\in\mathcal{D}$, $|H_n(\mathbf{c}|\mathbf{x},\tau)|\geq\epsilon_1$, whenever $|\mathbf{c}-\mathbf{c}_n(\mathbf{x};\tau)|\geq M_1$.

If this is false, there must exist three sequences $\{\tau_{n^*}\}$ in $[\delta^*,1-\delta^*]$, $\{\mathbf{x}_{n^*}\}$ in $\mathcal{D}$ and $\{\mathbf{c}_{n^*}\}$ in $R^{s(A)}$, such that as $n^*\to\infty$, $|\mathbf{c}_{n^*}-\mathbf{c}_n(\mathbf{x}_{n^*};\tau_{n^*})|\to\infty$, but $|H_n(\mathbf{c}_{n^*}|\mathbf{x}_{n^*},\tau_{n^*})|\to 0$. Without loss of generality, suppose there exist some $\tau^*\in[\delta^*,1-\delta^*]$ and $\mathbf{x}^*\in\mathcal{D}$, such that as $n^*\to\infty$, $\tau_{n^*}\to\tau^*$, and $\mathbf{x}_{n^*}\to\mathbf{x}^*$. Further construct the sequence $\{\Delta_{n^*}\}$ with $\Delta_{n^*}=\mathbf{c}_{n^*}-\mathbf{c}_n(\mathbf{x}_{n^*};\tau_{n^*})$, and for which we have, as $n^*\to\infty$, $|\Delta_{n^*}|\to\infty$, and $\Delta_{n^*}/|\Delta_{n^*}|\to\Delta^*$, for some $\Delta^*\in R^{s(A)}$.

Note that for any given $\mathbf{t}\in[-1,1]^p$, $\mathbf{c}_{n^*}^\top\mathbf{t}(A)=\mathbf{c}_n(\mathbf{x}_{n^*};\tau_{n^*})^\top\mathbf{t}(A)+\Delta_{n^*}^\top\mathbf{t}(A)$, the first term being finite, must tend to either $+\infty$ or $-\infty$ depending on whether $\mathbf{t}^\top(A)\Delta^*$ is positive or negative. Consequently, due to $F(.|.)$ being continuous in both its arguments, we have

$$\begin{aligned}\lim_{n^*\to\infty}F(\mathbf{c}_{n^*}^\top\mathbf{t}(A)|\mathbf{x}_{n^*}+h_n\mathbf{t}) &= \lim_{n^*}F(\mathbf{c}_{n^*}^\top\mathbf{t}(A)|\mathbf{x}^*+h_n\mathbf{t})\\ &= F(+\infty\times\ \text{sign}\{\mathbf{t}^\top(A)\Delta^*\}|\mathbf{x}^*+h_n\mathbf{t}),\end{aligned}$$

which must tend to either 1 or 0 depending on whether $\mathbf{t}^\top(A)\Delta^*$ is positive or negative respectively. As it is trivial to argue that the region $[-1,1]^p\cap\{\mathbf{t}:\mathbf{t}^\top(A)\Delta^*=0\}$ must have Lebesgue measure zero, a simple application of the Dominated Convergence Theorem to $H_n(\mathbf{c}_{n^*}|\mathbf{x}_{n^*},\tau_{n^*})$ yields

$$\tau^*\int_{[-1,1]^p\cap\{\mathbf{t}:\mathbf{t}^\top(A)\Delta^*<0\}}\mathbf{t}(A)d\mathbf{t}=(1-\tau^*)\int_{[-1,1]^p\cap\{\mathbf{t}:\mathbf{t}^\top(A)\Delta^*>0\}}\mathbf{t}(A)d\mathbf{t}.$$

Multiplying either side by Δ^*, we get

$$\tau^*\int_{[-1,1]^p\cap\{\mathbf{t}:\mathbf{t}^\top(A)\Delta^*<0\}}\mathbf{t}^\top(A)\Delta^*d\mathbf{t}=(1-\tau^*)\int_{[-1,1]^p\cap\{\mathbf{t}:\mathbf{t}^\top(A)\Delta^*>0\}}\mathbf{t}^\top(A)\Delta^*d\mathbf{t}.$$

As $0<\tau^*<1$, the above implies that both regions $[-1,1]^p\cap\{\mathbf{t}:\mathbf{t}^\top(A)\Delta^*<0\}$ and $[-1,1]^p\cap\{\mathbf{t}:\mathbf{t}^\top(A)\Delta^*>0\}$ must both have Lebesgue measure zero, which can't be true.

Step 2: For any $\mathbf{t}\in[-1,1]^p$, write $R_n(\mathbf{t};\tau,\mathbf{x})=\mathbf{t}^\top(A)\mathbf{c}_n(\mathbf{x};\tau)-Q_\tau(\mathbf{x}+h_n\mathbf{t})$. Note that $R_n(\mathbf{t},\mathbf{x})=O(h_n^{s_2})$ uniformly in $\mathbf{t}\in[-1,1]^p,\tau\in[\delta^*,1-\delta^*]$

and $\mathbf{x} \in \mathcal{D} \subset R^p$. For any $\mathbf{t} \in [-1,1]^p$ and $\mathbf{c} \in R^{s(A)}$, define a real valued function as

$$g_n(\mathbf{c}, \mathbf{t}|\mathbf{x}, \tau) = \frac{F(\mathbf{c}^\top \mathbf{t}(A)|\mathbf{x} + h_n\mathbf{t}) - F(\mathbf{c}_n(\mathbf{x};\tau)^\top \mathbf{t}(A)|\mathbf{x} + h_n\mathbf{t})}{(\mathbf{c} - \mathbf{c}_n(\mathbf{x};\tau))^\top \mathbf{t}(A)}.$$

In the case where $(\mathbf{c} - \mathbf{c}_n(\mathbf{x};\tau))^\top \mathbf{t}(A) = 0$, $g_n(\mathbf{c}, \mathbf{t}|\mathbf{x}, \tau)$ can be defined arbitrarily because the set $\{\mathbf{t} \in [-1,1]^p : \mathbf{c}^\top \mathbf{t}(A) = 0\}$ has Lebesque measure zero for any nonzero $\mathbf{c}$. Write

$$\begin{aligned} H_n(\mathbf{c}|\mathbf{x}, \tau) &= \int_{[-1,1]^p} \{F(\mathbf{c}^\top \mathbf{t}(A)|\mathbf{x} + h_n\mathbf{t}) - F(\mathbf{c}_n(\mathbf{x};\tau)^\top \mathbf{t}(A)|\mathbf{x} + h_n\mathbf{t})\}\mathbf{t}(A)d\mathbf{t} \\ &+ \int_{[-1,1]^p} \{F(\mathbf{c}_n(\mathbf{x};\tau)^\top \mathbf{t}(A)|\mathbf{x} + h_n\mathbf{t}) - F(Q_\tau(\mathbf{x} + h_n\mathbf{t})|\mathbf{x} + h_n\mathbf{t})\}\mathbf{t}(A)d\mathbf{t} \\ &= \Big[\int_{[-1,1]^p} g_n(\mathbf{c}, \mathbf{t}|\mathbf{x}, \tau)\mathbf{t}(A)\{\mathbf{t}(A)\}^\top dt\Big](\mathbf{c} - \mathbf{c}_n(\mathbf{x};\tau)) \\ &+ \int_{[-1,1]^p} f_{Y|\mathbf{X}}(Q_\tau(\mathbf{x} + h_n\mathbf{t}) + \xi_1 R_n(\mathbf{t};\tau,\mathbf{x})|\mathbf{x} + h_n\mathbf{t})R_n(\mathbf{t};\tau,\mathbf{x})\mathbf{t}(A)d\mathbf{t}, \end{aligned} \tag{a.20}$$

where ξ_1 lies between 0 and 1, depending on $\mathbf{t}$, τ and $\mathbf{x}$.

Step 3: By Cauchy inequality, we have regarding the second term on the right-hand side of (a.20),

$$\begin{aligned} &\Big|\int_{[-1,1]^p} \{f_{Y|\mathbf{X}}(Q_\tau(\mathbf{x} + h_n\mathbf{t}) + \xi_1 R_n(\mathbf{t};\tau,\mathbf{x})|\mathbf{x} + h_n\mathbf{t})\}R_n(\mathbf{t},\mathbf{x})\mathbf{t}(A)d\mathbf{t}\Big|^2 \\ &\le |\sup_{y,\mathbf{x}} f_{Y|\mathbf{X}}(y|\mathbf{x})|^2[s(A)]2^p \int_{[-1,1]^p} |R_n(\mathbf{t};\tau,\mathbf{x})|^2 d\mathbf{t} = O(h_n^{2s_2}). \end{aligned} \tag{a.21}$$

uniformly in $\tau \in [\delta^*, 1-\delta^*]$ and $\mathbf{x} \in \mathcal{D}$.

Step 4: Now in view of assumption [A3], there exists $\lambda_1 > 0$, such that $g_n(\mathbf{c}, \mathbf{t}|\tau, \mathbf{x}) \ge \lambda_1$ for all $\mathbf{c}$, $\mathbf{t}$ and $\mathbf{x} \in \mathcal{D}$ and $\tau \in [\delta^*, 1-\delta^*]$, such that $|\mathbf{c} - \mathbf{c}_n(\mathbf{x};\tau)| \le M_1$ and $(\mathbf{c} - \mathbf{c}_n(\mathbf{x};\tau))^\top \mathbf{t}(A) \ne 0$. Let λ_2 be the smallest e-value of the $s(A) \times s(A)$ matrix Γ. Then for the first term on the right hand side of (a.20), we have

$$\Big|\Big[\textstyle\int_{[-1,1]^p} g_n(\mathbf{c}, \mathbf{t}|\mathbf{x}, \tau)\mathbf{t}(A)\{\mathbf{t}(A)\}^\top dt\Big](\mathbf{c} - \mathbf{c}_n(\mathbf{x};\tau))\Big| \ge \lambda_1\lambda_2|\mathbf{c} - \mathbf{c}_n(\mathbf{x};\tau)|, \tag{a.22}$$

for all $\mathbf{c} \in R^{s(A)}$ such that $|\mathbf{c} - \mathbf{c}_n(\mathbf{x};\tau)| \le M_1$. The assertion in the proposition thus follows from (a.20), (a.21) ,(a.22) and the conclusion reached in *Step 1.* □

We collect here some useful results for the verification of Euclidean property of a class of functions.

[C1] Let $\mathfrak{F} = \{f(.,t) : t \in T\}$ be a class of functions indexed by a bounded subset T of R^d. If there exists an $\alpha > 0$ and a nonnegative function

$\phi(.)$ such that

$$|f(.,t) - f(.,t')| \le \phi(.)\|t - t'\|^\alpha \quad \text{for any } t,\ t' \in T,$$

then $\mathfrak{F}$ is Euclidean for the envelope $|f(.,t_0)| + M\phi(.)$, where t_0 is an arbitrary point of T and $M = (2\sqrt{d}\sup_T \|t - t_0\|)^\alpha$. [Lemma (2.13) of Pakes and Pollard (1989)]

[C2] If a class of functions $\mathfrak{F}$ is Euclidean for an envelop F and $\mathfrak{g}$ is Euclidean for an envelop G, then $\{f + g : f \in \mathfrak{F},\ g \in \mathfrak{g}\}$ is Euclidean for the envelop $F + G$ and $\{fg : f \in \mathfrak{F},\ g \in \mathfrak{g}\}$ is Euclidean for the envelop FG. [Lemma (2.14) of Pakes and Pollard (1989)]

[C3] Let $\lambda(.)$ be a real-valued function of bounded variation on R. The class of all functions on R^p of the form $\{\lambda(\mathbf{b}^\top \mathbf{x} + c) :\ \mathbf{b} \in R^p,\ c \in R$ is Euclidean for a constant envelop. [Lemma 22 (ii) of Nolan and Pollard (1987)]

[C4] Let $\lambda(.)$ be a real-valued function of bounded variation on R^+. The class of all functions on R^p of the form $\{\lambda(\|\mathbf{B}\mathbf{x}+\mathbf{b}\|) :\ \mathbf{B} \in R^{m\times p},\ \mathbf{b} \in R^m$ is Euclidean for a constant envelop. [Lemma 22 (i) of Nolan and Pollard (1987)]

COROLLARY A.3. *The following classes of functions are all Euclidean for an constant envelope:* $\{I\{Y_i \le Q_\tau(\mathbf{X}_i)\} = I\{F(Y_i|\mathbf{X}_i) \le \tau\},\ \tau \in (0,1)\}$, $\{\mathbf{X}_{i\mathbf{x}}(h_n, A) :\ \mathbf{x} \in \mathcal{D}\}$, $\{I(|\mathbf{X}_{i\mathbf{x}}| \le h_n) : \mathbf{x} \in \mathcal{D}\}$ *and* $\{I\{Y_i \le Q_n(\mathbf{X}_i, \mathbf{x}; \tau)\} :\ \mathbf{x} \in \mathcal{D}, \tau \in (0,1)\}$.

Proof of Corollary A.3. This follows easily from [C2], [C3] and [C4]. □

Proof of (a.5). By Corollary A.3, any algebraic operations involving these classes of functions are also Euclidean; e.g. $\{\mathbf{X}_{ij}(h_n, A)[I\{Y_i \le Q_n(\mathbf{X}_i, \mathbf{X}_j; \tau)\} - I\{Y_i \le Q_\tau(\mathbf{X}_i)\}]I(|\mathbf{X}_{ij}| \le h_n) : \mathbf{X}_j \in \mathcal{D}, \tau \in (0,1)\}$. This together with Theorem 37 in Pollard (1984, pp. 34) and the fact that $Q_n(\mathbf{X}_i, \mathbf{X}_j; \tau) - Q_\tau(\mathbf{X}_i) = O(h_n^{s_2})$ lead to (a.5), i.e. with probability one,

$$\frac{1}{nh_n^p} \sum_i \mathbf{X}_{ij}(h_n, A)[I\{Y_i \le Q_n(\mathbf{X}_i, \mathbf{X}_j; \tau)\} - I\{Y_i \le Q_\tau(\mathbf{X}_i)\}]I(|\mathbf{X}_{ij}| \le h_n) = o(n^{-1/2})$$

uniformly in $\mathbf{X}_j \in \mathcal{D},\ \tau \in (0,1)$. □

REFERENCES

[1] ARCONES, M. A. (1995). A Bernstein-type inequality for U-statistics and U-processes. *Statist. Probab. Lett.*, **22** 239-247.

[2] BHATTACHARYA, P. K. AND AND GANGOPADHYAY, A. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Ann. Statist.*, **18** 1400-1415.

[3] Bai, Z. D., Miao, B. Q. and Rao, C. R. (1991). Estimation of directions of arrival of signals: Asymptotic results. *Advances in Spectrum Analysis and Array Processing*, Vol. **1** (Simon Haykin, ed.)

[4] Chaudhuri, P. (1991). Global nonparametric estimation of conditional quantile functions and their derivatives. *J. Multi. Anal.*, **39** 246-269.

[5] Chaudhuri, P., Doksum, K. and Samarov, A. (1997). On average derivative quantile regression. *Ann. Statist.*, **25** 715-744.

[6] Cook, R. D. (1994). Using dimension-reduction subspaces to identify important inputs in models of physical systems. In *Proceedings of the Section on Physical and Engineering Sciences*, pp. 18-25. Alexandria, VA: American Statistical Association.

[7] Cook, R. D. (1998). *Regression Graphics*. New York: Wiley.

[8] Cook, R. D. (2007). Fisher Lecture: dimension reduction in regression (with discussion). *Statist. Sci.*, **22** 1-26.

[9] Cook, R. D. and Li, B. (2002). Dimension reduction for conditional mean in regression. *Ann. Statist.*, **30** 455?74.

[10] Fukumizu, K., Bach, F. R. and Jordan, M. (2009). Kernel dimension reduction in regression. *Ann. Statist.*, **37** 1871-1905

[11] He, X., Wang, L., and Hong, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Ann. Statist.*, **41** 342-369.

[12] Hristache, M., Juditsky, A., Polzehl, J. and Spokoiny, V. (2001). Structure adaptive approach for dimension reduction. *Ann. Statist.*, **29** 1537–1566.

[13] Kai, B., Li, R. and Zou, H. (2010). Local CQR smoothing: an efficient and safe alternative to local polynomial regression. *J. Roy. Statist. Soc. Ser. B*, **72** 49-69.

[14] Kato, T. (1995) *Perturbation theory for linear operators*. Springer-Verlag: New York.

[15] Koenker, R. and Machado, J. (1999). Goodness of fit and related inference processes for quantile regression. *J. Amer. Statist. Assoc.*, **94** 1296-1310.

[16] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, **46** 33-50.

[17] Koenker, R., Portnoy, S. and Ng, P. (1992). Nonparametric estimation of conditional quantile function. In *Proceedings of the Conference on L_1 Statistical Analysis and Related Methods* (Y. Dodge, ed.) 217-229. North-Holland, Amsterdam.

[18] Koenker, R., Ng, P. and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, **81** 673-680.

[19] Kong, E., Linton, O. and Xia, Y. (2010). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory*, **26** 1529-1564.

[20] Kong, E., Linton, O. and Xia, Y. (2013) Global Bahadur representation for nonparametric censored regression quantiles and its applications. *Econometric Theory*, **29** 941-968.

[21] Li, B. Cook, D. and Chiaromonte, F. (2003). Dimension reduction for conditional mean in regression with categorical predictors. *Ann. Statist.*, **31** 1636-1668

[22] Li, B., Zha, H. and Chiaromonte, F. (2005). Contour regression: a general approach to dimension reduction. *Ann. Statist.*, **33** 1580-1616.

[23] Li, K. (1991). Sliced Inverse Regression for Dimension Reduction. *J. Amer. Statist. Assoc.*, **86** 316-342.

[24] Lue, H. (2004). Principal Hessian Directions for regression with measurement error. *Biometrika*, **91** 409-423.

[25] Ma, Y. and Zhu, L. (2012). A semiparametric approach to dimension reduction. *J. Amer. Statist. Assoc.*, **107** 168-179.

[26] Masry, E. (1996) Multivariate local polynomial regression for time series: uniform

strong consistency and rates. *J. Time Ser. Anal.*, **17** 571-599.

[27] Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence. *Ann. Statist.*, **15** 780-799.

[28] Pakes, A. and Pollard, D. (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, **57** 1027-1057.

[29] Pollard, D. (1984). *Convergence of stochastic processes.* Springer-Verlag: New York.

[30] Sun, S. (1988). Analytic expressions for the derivatives of the eigenvalues and eigenvectors of a matrix. *Adva. Mathema.* (in Chinese), **17** 391-397.

[31] Truong, Y. (1989). Asymptotic properties of kernel estimates based on local medians. *Ann. Statist.*, **17** 606-617.

[32] van der Vaart and Wellner, J. (1996). *Weak convergence and empirical processes: with applications to statistics.* Springer-Verlag: New York.

[33] Wang, H. and Xia, Y. (2008). Sliced regression for dimension reduction. *J. Amer. Statist. Assoc.*, **103** 811-821.

[34] Xia, Y., Tong, H., Li, W. K. and Zhu, L. (2002) An adaptive estimation of dimension reduction space (with discussion). *J. Roy. Statist. Soc. Ser. B*, **64** 363-410.

[35] Xia, Y. (2007). A constructive approach to the estimation of dimension reduction directions. *Ann. Statist.*, **35** 2654-2690.

[36] Yin, X. and Cook, R. D. (2002). Dimension reduction for the conditional kth moment in regression. *J. Roy. Statist. Soc. Ser. B*, **64** 159-175.

[37] Yin, X. and Li, B. (2011). Sufficient dimension reduction based on an ensemble of minimum average variance estimators. *Ann. Statist.*, **39** 3392-3416.

[38] Yin, X., Li, B. and Cook, R. D. (2008). Successive direction extraction for estimating the central subspace in a multiple-index regression. *J. Multi. Anal.*, **99** 1733-1757.

[39] Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.*, **93** 228-238.

[40] Zhu, L. and Fang, K. (1996). Asymptotics for kernel estimate of sliced inverse regression. *Ann. Statist.*, **24** 1053-1068.

[41] Zhu, L., Zhu, L. and Feng, Z. (2010). Dimension reduction in regressions through cumulative slicing estimation. *J. Amer. Statist. Assoc.*, **105** 1455-1466.

[42] Zhu, Y. and Zeng, P. (2006). Fourier methods for estimating the central subspace and the central mean subspace in regression. *J. Amer. Statist. Assoc.*, **101** 1638-1651.

[43] Zhu, L. and Zhu, L. (2009) Dimension reduction for conditional variance in regressions. *Statist. Sinica*, **19** 869-883.

[44] Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Ann. Statist.*, **36** 1108-1126.

School of Mathematics, Statistics and Actuarial Science
University of Kent at Canterbury
CT2 7NF United Kingdom
E-mail: e.kong@kent.ac.uk

Department of statistics and applied probability
National University of Singapore
117546 Singapore
E-mail: staxyc@nus.edu.sg