Statistica Sinica Preprint No: SS-2016-0426	
Title	Multiple Quantile Modelling via Reduced Rank Regression
Manuscript ID	SS-2016-0426
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0426
Complete List of Authors	Heng Lian Weihua Zhao and Yanyuan Ma
Corresponding Author	Heng Lian
E-mail	henglian@cityu.edu.hk

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

Multiple Quantile Modeling via Reduced-Rank Regression

Heng Lian, Weihua Zhao and Yanyuan Ma

City University of Hong Kong, Nantong University and Pennsylvania State University

Abstract: Quantile regression estimators at a fixed quantile level rely mainly on a small subset of the observed data. As a result, efforts have been made to construct simultaneous estimations at multiple quantile levels in order to take full advantage of all observations and to improve the estimation efficiency. We propose a novel approach that links multiple linear quantile models by imposing a condition on the rank of the matrix formed by all of the regression parameters. This approach resembles a reduced-rank regression, but also shares similarities with the dimension-reduction modeling. We develop estimation and inference tools for such models and examine their optimality in terms of the asymptotic estimation variance. We use simulation experiments to examine the numerical performance of the proposed procedure, and a data example to further illustrate the method.

Key words and phrases: Check function; Composite quantile regression; Generalized method of moment; Linear quantile regression; Optimal estimating equations; Quantile regression; Reduced-rank regression.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

1. Introduction

Quantile regression models (Koenker and Bassett Jr; 1978) are important alternatives to mean regression models. When a single quantile relation is modeled, the parameter estimation is naturally more robust than the estimation in the mean regression model. When multiple quantiles are modeled, a quantile model provides a more complete description of the relation between the covariates and the response variable than a mean model does. Quantile models have numerous applications in the fields of economics and medicine, among others (Cade and Noon; 2003; Yu et al.; 2003; Wang and He; 2007). In addition to linear quantile models, many studies have developed various nonlinear, semiparametric, and even nonparametric extensions to these models, including He and Shi (1994); De Gooijer and Zerom (2003); Horowitz and Lee (2005); Kim (2007); Wang et al. (2009); Lian (2012); He et al. (2013), among many others. Bayesian quantile regressions have also received significant attention in recent years (Yu and Moyeed; 2001; Kozumi and Kobayashi; 2011).

In a linear quantile regression, we assume that the conditional τ-th quantile of the response Y is given by $a_\tau + \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_\tau$, that is, $\mathrm{pr}(Y \leq a_\tau + \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_\tau|\mathbf{X}) = \tau$ for $\tau \in (0, 1)$. The standard quantile regression estimator at a fixed quantile level depends only on a small subset of the data, and thus,

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

is known to be unstable in practice. Theoretically, the quantile estimator has a large asymptotic variance when the error density at this particular quantile level is small. Therefore, it is natural to hope that by borrowing information from other quantile levels, we can potentially improve the estimation efficiency. A method that follows this rationale is the composite quantile regression (CQR) proposed by Zou and Yuan (2008a), where the quantile regression errors are assumed to be independent of the covariates, and thus, all slope parameters $\boldsymbol{\beta}_\tau$ are identical for different values of τ. A similar assumption was used in Zhao and Xiao (2014) with a different estimation approach. In this case, the authors applied the method to the prediction of stock returns to demonstrate the improved efficiency of estimation. CQR estimators are more efficient than quantile regression estimators at a single quantile level. However, the independent regression error assumption more or less defies the purpose of the original quantile regression model and is somewhat stringent. The CQR assumption is relaxed in Jiang et al. (2013, 2014), who assumed that the quantile slope coefficients are identical in certain regions of quantiles, instead of for all quantiles. Penalization was then used to identify such regions. Their results showed improved estimation accuracy over that of single quantile regression model. They also use simulations and an application to the Barro growth data

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

to show the model's improved prediction performance. Taking a different approach, Zou and Yuan (2008b); He et al. (2016) borrowed information by assuming that the sparsity patterns in a linear quantile regression for neighboring quantile levels are similar. Furthermore, a penalized regression is used for variable selection. The advantage of borrowing information from neighboring quantiles in a linear quantile regression is also recognized in the Bayesian implementation. Using a random-walk prior distribution instead of an independent prior, Yang and He (2012) imposed a smoothness condition where the slope parameter $\boldsymbol{\beta}_\tau$ is a function of the quantile level τ. In addition, the authors demonstrated the improvement in performance over the case without the smoothness condition of the slopes.

In all of these models, the information at different quantile levels is interrelated. Hence, the information over a level can be used to benefit that on other levels because of the assumption that the quantile regression slope parameters are identical or similar at different quantile levels. However, we adopt a different approach borrowing information from multiple quantiles. Instead of assuming identical slope parameters, we assume that the slope parameter vectors $\boldsymbol{\beta}_\tau$ at different quantile levels τ have a certain linear dependency. In other words, for K quantile levels $0 < \tau_1 < \cdots < \tau_K < 1$, we assume the matrix $\mathbf{B} \equiv (\boldsymbol{\beta}_{\tau_1}, \ldots, \boldsymbol{\beta}_{\tau_K})$ satisfies a certain low-rank

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

constraint.

Thus, our approach is related to the reduced-rank regression proposed in the 1950s (Anderson; 1951), and our goal is to introduce a more parsimonious model in cases with multiple responses. With reduced degrees of freedom, the reduced-rank regression has the potential to produce a more efficient estimator of $\mathbf{B}$, as shown in Anderson (1999). More recent works on reduced-rank regression include those of Geweke (1996); Bunea et al. (2011, 2012); Chen and Huang (2012); Chen et al. (2012, 2013); Lian and Ma (2013).

Next we provide several examples to motivate the low-rank assumption of $\mathbf{B}$, based on common data-generation procedures. In all of these situations, we fit a linear quantile regression model. First, we generate data from the relation $Y_i = a + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\alpha} + \epsilon_i$, where ϵ_i is independent of the covariates, as in the CQR model. When we fit the linear quantile regression model

$$Y_i = a_\tau + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_\tau + \epsilon_{\tau,i}, \tag{1.1}$$

where $\epsilon_{\tau,i} = Y_i - a_\tau - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_\tau$ has the τth conditional quantile zero, we obtain $\boldsymbol{\beta}_\tau = \boldsymbol{\alpha}$. Hence, $\mathbf{B}$ has rank 1. Note that the rank 1 constraint on $\mathbf{B}$ allows slightly more flexibility than the CQR model in that it does not require all columns of $\mathbf{B}$ to be identical. This additional flexibility allows other data-generation procedures to be included. For example, assume

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

that the true relation between $\mathbf{X}_i$ and Y_i is $Y_i = (b + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\gamma})\epsilon_i$, where, as in CQR, ϵ_i is independent of $\mathbf{X}_i$. When we fit the same linear quantile regression model given in (1.1), we obtain $\boldsymbol{\beta}_\tau = \boldsymbol{\gamma} F_\epsilon^{-1}(\tau)$ if $b+\mathbf{X}^{\mathrm{T}}\boldsymbol{\gamma} \geq 0$, and $\boldsymbol{\beta}_\tau = \boldsymbol{\gamma} F_\epsilon^{-1}(1-\tau)$ if $b + \mathbf{X}^{\mathrm{T}}\boldsymbol{\gamma} < 0$, where F_ϵ is the cumulative distribution function and F_ϵ^{-1} is the quantile function of ϵ. Thus, $\mathbf{B}$ also has rank 1. A more familiar example that combines these two situations is when the data come from a location-scale structure. Consider $(\mathbf{X}_i, Y_i)$ that satisfies $Y_i = a + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\alpha} + (b + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\gamma})\epsilon_i$, where ϵ_i is independent of $\mathbf{X}_i$. At the τth quantile level, when we fit model (1.1), we obtain $\boldsymbol{\beta}_\tau = \boldsymbol{\alpha} + F_\epsilon^{-1}(\tau)\boldsymbol{\gamma}$ if $b + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\gamma} \geq 0$, and $\boldsymbol{\beta}_\tau = \boldsymbol{\alpha} + F_\epsilon^{-1}(1-\tau)\boldsymbol{\gamma}$ if $b + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\gamma} < 0$. Because $\boldsymbol{\beta}_\tau$ can be obtained from the linear combination of $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ for any τ, the matrix $\mathbf{B}$ has rank 2 for data with the location-scale structure. Finally, consider the following very general data structure often used in the sufficient dimension-reduction literature:

$$Y_i = f(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{X}_i, \epsilon_i), \tag{1.2}$$

where the error ϵ_i is independent of the covariate $\mathbf{X}_i$, and $\boldsymbol{\alpha}$ is a $p \times d$ matrix that spans the dimension-reduction space. Assume the covariates $\mathbf{X}_i$ satisfy the linearity condition, that is, $E(\mathbf{X}|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{X})$ is a linear function of $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{X}$. Even though the true link function is unknown and may be nonlinear in (1.2), we can still fit the linear quantile regression model given in (1.1).

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

In the online Supplementary Material, we show that when we minimize the expected check function $\rho_\tau(Y - a - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta})$ with respect to a and $\boldsymbol{\beta}$, where $\rho_\tau(t) = t\{\tau - I(t \leq 0)\}$, the minimizer $\boldsymbol{\beta}_\tau \in \mathrm{span}(\boldsymbol{\alpha})$ for all $\tau \in (0, 1)$. This implies that $\mathbf{B}$ has rank at most d. These examples illustrate that the low-rank phenomenon of $\mathbf{B}$ often arises naturally in practice, and thus is a sensible constraint to impose. Under the low-rank assumption, the number of parameters in a multiple linear quantile regression model can be reduced, and hence, we can expect to achieve better estimation accuracy.

The rest of the paper is organized as follows. In Section 2, we propose several multiple linear quantile regression estimators that take advantage of the low-rank property of $\mathbf{B}$. Computational procedures are also described. Section 3 contains the asymptotic properties of these estimators. Section 4 is devoted to simulation studies that illustrate the improved accuracy of the proposed estimators compared with that of the standard multiple linear quantile estimator and the estimators proposed in Jiang et al. (2013). Section 5 presents and analyzes a data example, and Section 6 concludes the paper. The technical proofs are contained in the Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

2. Methods

Consider independent and identically distributed (i.i.d.) observations $(\mathbf{X}_i, Y_i), i = 1, \ldots, n$, where $\mathbf{X}_i$ is a p-dimensional vector of covariates and Y_i is a scalar response, and $(\mathbf{X}_i, Y_i)$ are modeled using the τth linear quantile regression model (1.1) at a series of quantile levels $0 < \tau_1 < \cdots < \tau_K < 1$. The sequence of quantile levels can either be user specified, depending on the quantiles of interest of the problem, or follow some default choice, such as $\tau_k = k/(K+1), k = 1, \ldots, K$, with $K = 9$ or 19 (Zou and Yuan; 2008a). For notational simplicity, we write $a_{\tau_k}, \boldsymbol{\beta}_{\tau_k}$ as $a_k, \boldsymbol{\beta}_k$ respectively.

2.1 Multiple linear quantile regression

When no additional information is available other than the relation described in the K linear quantile regression models, the natural approach is to estimate each pair $(a_k, \boldsymbol{\beta}_k)$ from the standard procedure of minimizing the check function. More formally, let $\mathbf{a} = (a_1, \ldots, a_K)^{\mathrm{T}}$ and $\mathbf{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K)$, as defined before. Then, we can obtain an estimator for $\mathbf{a}$ and $\mathbf{B}$ by solving

$$\min_{\mathbf{a},\mathbf{B}} \sum_{k=1}^{K} \sum_{i=1}^{n} \rho_{\tau_k}(Y_i - a_k - \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}_k). \tag{2.1}$$

Note that although we express (2.1) as a single minimization problem, it is effectively identical to K separate minimization problems, each correspond-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

ing to a single linear quantile regression.

Of course, it is well known that minimizing the check function is not the best option in terms of estimating the regression parameters. To minimize the variability of the parameter estimation, we should instead solve the estimation equations

$$\sum_{i=1}^{n} f_{\epsilon_\tau|\mathbf{X}}(0, \mathbf{X}_i)\mathbf{Z}_i\psi_\tau(Y_i - a_\tau - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_\tau) = \mathbf{0},$$

where $f_{\epsilon_\tau|\mathbf{X}}$ is the probability density function (pdf) of ϵ_τ conditional on $\mathbf{X}$, $\psi_\tau(t) = \tau - I(t \leq 0)$, and $\mathbf{Z}_i = (1, \mathbf{X}_i^{\mathrm{T}})^{\mathrm{T}}$ (Newey and Powell; 1990; Lee; 2003). When we consider K different quantile levels simultaneously, it is tempting to simply concatenate the K equations. However, a more careful inspection reveals a better solution even though there is no apparent relation between the parameters corresponding to different quantile levels. Because ϵ_τ at different values of τ are correlated, a generalized estimating equation (GEE) principle can be used to combine the individual quantile level estimating functions in a more efficient way. Specifically, let $\boldsymbol{\psi}(Y_i - \mathbf{a} - \mathbf{B}^{\mathrm{T}}\mathbf{X}_i) \equiv \{\psi_{\tau_1}(Y_i - a_1 - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_1), \ldots, \psi_{\tau_K}(Y_i - a_K - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_K)\}^{\mathrm{T}}$, and let $\mathbf{V}$ be the variance covariance matrix of $\boldsymbol{\psi}$, that is, let $\mathbf{V}$ be a $K \times K$ matrix, where entry (k, k') is $\mathbf{V}_{k,k'} = \min(\tau_k, \tau_{k'}) - \tau_k\tau_{k'}$. Then, we can solve

$$\sum_{i=1}^{n}(\mathbf{f}_i \otimes \mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i - \mathbf{a} - \mathbf{B}^{\mathrm{T}}\mathbf{X}_i) = \mathbf{0} \tag{2.2}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

to obtain a more efficient estimator of $\mathbf{a}$ and $\mathbf{B}$ under a multiple linear quantile regression without any constraints, where $\otimes$ is the Kronecker product. Here, $\mathbf{f}_i \equiv \text{diag}\{f_{\epsilon_{\tau_k}|\mathbf{X}}(0, \mathbf{X}_i), k = 1, \ldots, K\}$. Although (2.2) yields a more efficient estimator, it is not as popular a method, even in the single linear quantile regression literature. This is because the estimation involves the conditional pdf $f_{\epsilon_\tau|\mathbf{X}}(0, \mathbf{X})$, the estimation of which usually involves a nonparametric device and is nearly impossible when the dimension of $\mathbf{X}$ is large.

One compromise, following the general idea of using a "working model," is to replace the quantity $\mathbf{f}_i$, which is difficult to estimate, with a guessed model $\mathbf{f}_i^*$. Then, we can calculate an estimator from the equation

$$\sum_{i=1}^{n} (\mathbf{f}_i^* \otimes \mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i - \mathbf{a} - \mathbf{B}^{\mathrm{T}}\mathbf{X}_i) = \mathbf{0}. \tag{2.3}$$

If the guessed model $\mathbf{f}_i^*$ happens to be correct, we obtain an estimator as efficient as that from (2.2). However, even if we guessed incorrectly, we still obtain a consistent estimator.

In the special case of using a uniform model for $\mathbf{f}_i^*$, we actually obtain a GEE improved version of (2.1). To see this, we first recognize that minimizing the check function at a single quantile level can be re-expressed as

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

solving the estimating equation

$$\sum_{i=1}^{n} \psi_\tau(Y_i - a_k - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_k)\mathbf{Z}_i = \mathbf{0}.$$

Thus, we can follow the same "GEE principle" that we used to obtain (2.2) to take into account the correlation between the K sets of such estimating functions, yielding the following estimating equation:

$$\sum_{i=1}^{n} (\mathbf{I}_K \otimes \mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i - \mathbf{a} - \mathbf{B}^{\mathrm{T}}\mathbf{X}_i) = \mathbf{0}, \tag{2.4}$$

which is identical to (2.3) for $\mathbf{f}_i^* = \mathbf{I}$.

Because $\psi_\tau(\cdot)$ is discontinuous, instead of searching for zeros of the estimating equations, it is better to write estimators such as those described in (2.2), (2.3), and (2.4) as the solution to a minimization problem, such as

$$\min_{\mathbf{a},\mathbf{B}} \| \sum_{i=1}^{n} (\mathbf{f}_i^* \otimes \mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i - \mathbf{a} - \mathbf{B}^{\mathrm{T}}\mathbf{X}_i)\|.$$

To simplify the writing and to concentrate on the main idea, in the following text, we do not distinguish between the two ways of expressing the estimation methods.

2.2 Rank-constrained multiple linear quantile regression

When $\mathbf{B} \in R^{p\times K}$ has rank at most r, where $r < \min(p, K)$, we no longer have $(p+1)K$ free parameters. An immediate strategy, following the idea

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

of minimizing the check function, is to modify (2.1) into a constrained minimization:

$$\min_{\substack{\mathbf{a},\mathbf{B}\\ \text{rank}(\mathbf{B})\le r}} \sum_{k=1}^{K}\sum_{i=1}^{n} \rho_{\tau_k}(Y_i - a_k - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_k). \tag{2.5}$$

A similar strategy can be adopted to modify the estimating equation in the family of (2.3). We aim to minimize the l_2 norm of the $(p+1)K$ equations, that is, we solve the constrained minimization problem

$$\min_{\substack{\mathbf{a},\mathbf{B}\\ \text{rank}(\mathbf{B})\le r}} \| \sum_{i=1}^{n} (\mathbf{f}_i^* \otimes \mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i - \mathbf{a} - \mathbf{B}^{\mathrm{T}}\mathbf{X}_i)\|^2$$

to obtain an estimator. The special choices of $\mathbf{f}_i^* = \mathbf{f}_i$ and $\mathbf{f}_i^* = \mathbf{I}_K$ correspond to the constrained versions of (2.2) and (2.4). However, in the former case, we need to estimate the conditional pdf at zero. Hence, the estimator is difficult to implement.

2.3 Better usage of the reduced-rank constraint

The analyses in Sections 2.1 and 2.2 indicate that the adaption from a constraint-free multiple linear quantile regression to a reduced-rank case mainly requires taking into account that we have fewer free parameters, owing to the restriction on the rank of $\mathbf{B}$. This changes the original optimization from being unconstrained to being constrained. Furthermore, it causes the original just-identifying estimating equations to become over-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

identifying. As a tool to handle an excessive number of estimating equations, the generalized method of moments (GMM) is well developed. Thus, we use it here to better handle the reduced-rank constraint.

To modify the constrained minimization estimator from (2.5), the GMM-based procedure based on the estimating equations in (2.4) for the reduced-rank model minimizes

$$\frac{1}{n}\left\{\sum_{i=1}^{n}(\mathbf{I}_K\otimes\mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i-\mathbf{a}-\mathbf{B}^{\mathrm{T}}\mathbf{X}_i)\right\}^{\mathrm{T}}\mathbf{W}\left\{\sum_{i=1}^{n}(\mathbf{I}_K\otimes\mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i-\mathbf{a}-\mathbf{B}^{\mathrm{T}}\mathbf{X}_i)\right\}, \tag{2.6}$$

subject to $\mathrm{rank}(\mathbf{B})\le r$. Here, $\mathbf{W}$ is the $(p+1)K\times(p+1)K$ weight matrix given by

$$\begin{aligned}\mathbf{W} &= \left[\mathrm{var}\left\{(\mathbf{I}_K\otimes\mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i-\mathbf{a}-\mathbf{B}^{\mathrm{T}}\mathbf{X}_i)\right\}\right]^{-1}\\ &= \left[E\left\{(\mathbf{I}_K\otimes\mathbf{Z}_i)\mathbf{V}^{-1}(\mathbf{I}_K\otimes\mathbf{Z}_i)^{\mathrm{T}}\right\}\right]^{-1}.\end{aligned}$$

We call the GMM estimator obtained from (2.6) the GMM reduced-rank (GMMRR) estimator. The GMMRR estimator is easy to implement, as shown in Section 2.4, and is the main result proposed in this work.

We can also achieve optimal efficiency is to by starting from (2.2). Following the same GMM idea, the optimal resulting estimator can be

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

obtained from the minimization problem

$$\min_{\substack{\mathbf{a},\mathbf{B}\\ \mathrm{rank}(\mathbf{B})\le r}} \frac{1}{n}\left\{\sum_{i=1}^{n}(\mathbf{f}_i\otimes\mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i-\mathbf{a}-\mathbf{B}^{\mathrm{T}}\mathbf{X}_i)\right\}^{\mathrm{T}}\mathbf{W}_f \left\{\sum_{i=1}^{n}(\mathbf{f}_i\otimes\mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i-\mathbf{a}-\mathbf{B}^{\mathrm{T}}\mathbf{X}_i)\right\}, \tag{2.7}$$

where the weight matrix in this case is

$$\mathbf{W}_f=\left[E\left\{(\mathbf{f}_i\otimes\mathbf{Z}_i)\mathbf{V}^{-1}(\mathbf{f}_i\otimes\mathbf{Z}_i)^{\mathrm{T}}\right\}\right]^{-1}.$$

We call this the optimal reduced-rank estimator. Note that this estimator is optimal in terms of both a single quantile regression estimation and combining multiple quantile requirements under a rank constraint. In addition, it requires knowledge or the assessment of the conditional pdfs of the quantile regression errors at zero for each covariate value $\mathbf{X}_i$. To achieve efficiency, these quantities need to be estimated nonparametrically, which is difficult when the dimension of $\mathbf{X}$ becomes large. Thus, we study this estimator for its theoretical value only, and do not recommend using it in practice.

Of course, both (2.6) and (2.7) can be viewed as special members of the general family of GMM estimators

$$\min_{\substack{\mathbf{a},\mathbf{B}\\ \mathrm{rank}(\mathbf{B})\le r}} \left\{\sum_{i=1}^{n}(\mathbf{f}_i^*\otimes\mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i-\mathbf{a}-\mathbf{B}^{\mathrm{T}}\mathbf{X}_i)\right\}^{\mathrm{T}}\mathbf{W}_f^* \left\{\sum_{i=1}^{n}(\mathbf{f}_i^*\otimes\mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i-\mathbf{a}-\mathbf{B}^{\mathrm{T}}\mathbf{X}_i)\right\}, \tag{2.8}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

with

$$\mathbf{W}_f^* = \left[E\left\{(\mathbf{f}_i^* \otimes \mathbf{Z}_i)\mathbf{V}^{-1}(\mathbf{f}_i^* \otimes \mathbf{Z}_i)^{\mathrm{T}}\right\}\right]^{-1}.$$

Here, $\mathbf{f}_i^*$ is based on the working model of $f_{\epsilon_\tau|\mathbf{X}}$. If $\mathbf{f}_i^* = \mathbf{f}_i$, we obtain (2.7), and if $\mathbf{f}_i^* = \mathbf{I}_K$, we obtain (2.6).

2.4 Computation

The development in the previous section indicates that, in practice, feasible reduced-rank estimators are based on either (2.5) or (2.6), with preference given to (2.6). In fact, the only reason we retain (2.5) is its conceptual simplicity.

However, even the relatively simple constrained minimization problem in (2.5) is not that straightforward computationally and, thus, requires some care. Our approach to the computation is to convert (2.5) into an unconstrained minimization problem. Because any matrix $\mathbf{B}$ with rank bounded by r can be written as $\mathbf{B} = \mathbf{DA}$, with $\mathbf{D} \in R^{p\times r}$ and $\mathbf{A} \in R^{r\times K}$, (2.5) is equivalent to solving the following unconstrained minimization problem:

$$\min_{\mathbf{a},\mathbf{D},\mathbf{A}} \sum_{k=1}^{K}\sum_{i=1}^{n} \rho_{\tau_k}(Y_i - a_k - \mathbf{X}_i^{\mathrm{T}}\mathbf{DA}_k), \tag{2.9}$$

where $\mathbf{A}_k$ is the k-th column of $\mathbf{A}$. Although the matrices $\mathbf{D}$ and $\mathbf{A}$ are not unique, because if $(\mathbf{D}, \mathbf{A})$ minimizes (2.9), $(\mathbf{D\Omega}^{-1}, \mathbf{\Omega A})$ does so as well for

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

any nonsingular matrix $\mathbf{\Omega}$. However, we only need to find one such $(\mathbf{D}, \mathbf{A})$ pair, after which, the resulting $\mathbf{B}$ will be uniquely determined. Because $\mathbf{D}$ is the only component involved in all K summands, this naturally suggests an optimization strategy that alternates between updating $\mathbf{D}$ and updating each $(\mathbf{A}_k, a_k)$ for $k = 1, \ldots, K$. Specifically, fixing $\mathbf{a}$ and $\mathbf{A}$, and using the relation $\mathbf{X}_i^{\mathrm{T}}\mathbf{D}\mathbf{A}_k = (\mathbf{A}_k^{\mathrm{T}} \otimes \mathbf{X}_i^{\mathrm{T}})\mathrm{vec}(\mathbf{D})$, the optimization with respect to $\mathrm{vec}(\mathbf{D})$ is rewritten as

$$\min_{\mathrm{vec}(\mathbf{D})} \sum_{k=1}^{K} \sum_{i=1}^{n} \rho_{\tau_k}\{Y_i - a_k - (\mathbf{A}_k^{\mathrm{T}} \otimes \mathbf{X}_i^{\mathrm{T}})\mathrm{vec}(\mathbf{D})\},$$

which is essentially identical to the CQR estimation problem and can be solved using linear programming. On the other hand, after fixing $\mathbf{D}$, the optimization over $a_k, \mathbf{A}_k$ at each k involves only the kth summand in (2.9), and therefore is also a standard linear quantile regression problem. Despite the non-uniqueness of $(\mathbf{D}, \mathbf{A})$, in practice, the algorithm still converges to a minimizer that corresponds to the choice of the initial value. This alternating minimization algorithm is commonly used in reduced-rank regressions (Bunea et al.; 2012; Chen and Huang; 2012) and is found to be effective in our implementation.

The same idea of writing $\mathbf{B}$ as $\mathbf{DA}$ can be used in (2.6). In this case,

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

we rewrite (2.6) as

$$\left\{\sum_{i=1}^{n}(\mathbf{I}_K \otimes \mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i - \mathbf{a} - \mathbf{X}_i^{\mathrm{T}}\mathbf{D}\mathbf{A})\right\}^{\mathrm{T}} \mathbf{W} \left\{\sum_{i=1}^{n}(\mathbf{I}_K \otimes \mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i - \mathbf{a} - \mathbf{X}_i^{\mathrm{T}}\mathbf{D}\mathbf{A})\right\}, \quad (2.10)$$

and then minimize (2.10) with respect to $(\mathbf{a}, \mathbf{D}, \mathbf{A})$ without constraint. Note that, in practice, we replace $\mathbf{W}$ with its sample version

$$\widehat{\mathbf{W}} = \left\{n^{-1}\sum_{i=1}^{n}(\mathbf{I}_K \otimes \mathbf{Z}_i)\mathbf{V}^{-1}(\mathbf{I}_K \otimes \mathbf{Z}_i)^{\mathrm{T}}\right\}^{-1},$$

which needs to be computed only once in the minimization procedure. Because the target function (2.10) is not continuous, the optimization problem seems impossible to solve directly. Thus, we perform a smoothing approximation, which is often used in the quantile regression literature (Horowitz; 1998; Brown and Wang; 2005). Specifically, we replace the indicator function $I(t \leq 0)$ contained in $\boldsymbol{\psi}$ with a normal survival function $1 - \Phi(t/h)$, where Φ is the cumulative distribution function (cdf) of the standard normal distribution and h is a bandwidth. In the numerical studies, we set $h = n^{-1/2}$. After the smoothing operation, we can perform the optimization with respect to $\mathbf{D}$ and $\mathbf{A}$ iteratively using, for example, the Newton-Raphson procedure.

To minimize both (2.9) and (2.10), the iterative procedures require initial parameter values. We solve the K separate single linear quantile

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

regression problems as in Section 2.1 to obtain the initial estimators $\widetilde{\mathbf{a}}$ and $\widetilde{\mathbf{B}}$. Then, we perform a singular value decomposition on $\widetilde{\mathbf{B}}$ to obtain $\widetilde{\mathbf{B}} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}$. Next, we set $\widetilde{\mathbf{D}}$ as the first r columns of $\mathbf{U}$, and $\widetilde{\mathbf{A}}$ as the first r rows of $\mathbf{\Lambda}\mathbf{V}$.

Finally, although we do not recommend implementing the optimal estimator via solving (2.7) in practice, we experiment with it here to illustrate its performance and to provide it as a benchmark. If the conditional pdfs $\mathbf{f}_i$ are available, the same procedure used to minimize (2.6) can be used. To estimate $\mathbf{f}_i$, we adopt the method of Hendricks and Koenker (1992) and estimate $f_{\epsilon_{\tau_k}|\mathbf{X}}(0, \mathbf{X}_i)$ using the difference quotient $2h_n \left\{\mathbf{X}_i^{\mathrm{T}}(\boldsymbol{\beta}_{\tau_k+h_n} - \boldsymbol{\beta}_{\tau_k-h_n})\right\}^{-1}$, where $\boldsymbol{\beta}_{\tau_k+h_n}$ and $\boldsymbol{\beta}_{\tau_k-h_n}$ are estimated by the standard single linear quantile regression at quantile levels $\tau_k + h_n$ and $\tau_k - h_n$, respectively. Note that although this procedure avoids estimating $\mathbf{f}_i$ nonparametrically, it implicitly assumes that the quantile relation between $\mathbf{X}$ and Y is linear at quantile levels $\tau_k + h_n$ and $\tau_k - h_n$, which may or may not hold in practice. Here, h_n is a bandwidth that approaches zero as $n \to \infty$. In our numerical studies, we choose $h_n = 1.57n^{-1/3}(1.5\phi^2\{\Phi^{-1}(\tau)\}/[2\{\Phi^{-1}(\tau)\}^2 + 1])^{2/3}$, following Hall and Sheather (1988), where ϕ and Φ are the pdf and cdf, respectively, of the standard normal distribution.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

2.5 Determine the rank r

We have derived various estimators under a fixed-rank constraint r. In practice, when data are available to determine r, these methods can be applied directly. However, when no such data are available, we also need to determine r. To this end, we treat r as a tuning parameter and select it based on the data.

Regardless of whether we estimate the model parameters using (2.5) or (2.6), a common feature is that each estimator is the result of minimizing a target function. In addition, the minimum value of the target function is a monotonically increasing function of the rank constraint r. Here, r directly determines the flexibility of the model, reflected in the number of free parameters of the model. More importantly, when the minimization is conducted under a working rank constraint r that is satisfied by the true data-generation procedure, we will not be able to shrink the target function value further by increasing r. On the other hand, when the minimization is conducted under a working rank constraint r that is too small, and hence is not satisfied by the true data-generation procedure, we will be able to shrink the target function value further by increasing r. To see this, note that the target function in (2.5) measures the goodness-of-fit of the model directly, whereas (2.6) is bounded in probability quantity when r is sufficiently large

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

and approaches infinity when r becomes too small.

Based on the above considerations, we propose determining the rank constraint value r by minimizing the information criterion

$$\begin{aligned} \mathrm{BIC}(r) &= (2n)\sum_{k=1}^{K}\log\left[\sum_{i=1}^{n}\rho_{\tau_k}(Y_i-\widehat{a}_k-\mathbf{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_k)\right] \\ &\quad +\log(n)(K+pr+Kr-r^2) \end{aligned} \tag{2.11}$$

if (2.5) is used for estimation, and

$$\begin{aligned} &\mathrm{BIC}(r) \\ &= \frac{1}{2n}\left\{\sum_{i=1}^{n}(\mathbf{I}_K\otimes\mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i-\widehat{\mathbf{a}}-\mathbf{X}_i^{\mathrm{T}}\widehat{\mathbf{B}})\right\}^{\mathrm{T}}\mathbf{W}\left\{\sum_{i=1}^{n}(\mathbf{I}_K\otimes\mathbf{Z}_i)\mathbf{V}^{-1}\boldsymbol{\psi}(Y_i-\widehat{\mathbf{a}}-\mathbf{X}_i^{\mathrm{T}}\widehat{\mathbf{B}})\right\} \\ &\quad +\log(n)(K+pr+Kr-r^2) \end{aligned} \tag{2.12}$$

if (2.6) is used. Here, $K+pr+Kr-r^2$ is the number of free parameters under the rank constraint r. We use the same target function to form the criterion as that used to form the estimator, to keep things simple, but this is not required. Similar target functions to those used in the above BIC construction are often used in quantile regressions and the literature on estimating equations; see, for example, Jiang et al. (2013) and Wang and Qu (2009). It is easy to see that the construction in (2.12) is applicable to all estimators of the family described in (2.8). In Section 3 we show that the rank determination based on our BIC is consistent.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

3. Theoretical Results

3.1 Regularity conditions

Recall that the conditional quantile at quantile level τ is $a_\tau + \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_\tau$ for $\tau = \tau_k, k = 1, \ldots, K$, where $0 < \tau_1 < \cdots < \tau_K$ is a fixed sequence of quantile levels. Furthermore, recall that $f_{\epsilon_\tau|\mathbf{X}}(\epsilon_\tau, \mathbf{x})$ and $F_{\epsilon_\tau|\mathbf{X}}(\epsilon_\tau, \mathbf{x})$ are the conditional pdf and cdf, respectively, of ϵ_τ.

We assume the following regularity conditions.

C1. $f_{\epsilon_\tau|\mathbf{X}}(\epsilon_\tau, \mathbf{x})$ is continuously differentiable in ϵ_τ. There exist constants $\overline{f}, \overline{f'}$, such that $f_{\epsilon_\tau|\mathbf{X}}(\epsilon_\tau, \mathbf{x}) < \overline{f}$ and $|\partial f_{\epsilon_\tau|\mathbf{X}}(\epsilon_\tau, \mathbf{x})/\partial\epsilon_\tau| < \overline{f'}$. Furthermore, there exists a constant $\underline{f} > 0$, such that $f_{\epsilon_\tau|\mathbf{X}}(0, \mathbf{x}) > \underline{f}$ for all $\mathbf{x}$ in its support and all $\tau \in \{\tau_1, \ldots, \tau_K\}$.

C2. $E\|\mathbf{X}\|^3 < \infty$. $\mathbf{C} \equiv E[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}]$ is positive definite.

C3. The parameter space for $\boldsymbol{\theta}$ is a compact set, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_K^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\theta}_k = (a_k, \boldsymbol{\beta}_k^{\mathrm{T}})^{\mathrm{T}}$ for $k = 1, \ldots, K$.

C4. The true slope parameter matrix $\mathbf{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K)$ has a rank bounded by r. Note that $\mathbf{B}$ has size $p \times K$; hence, $r \leq \min(p, K)$.

The regularity conditions in (C1) are commonly assumed for quantile regressions (Wang et al.; 2009; Belloni and Chernozhukov; 2011; Wang et al.;

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

2012). (C2) imposes a mild moment condition on the covariates. In (C3), we require the parameter space to be bounded in order to apply the empirical process theory. The empirical process theory is needed because our optimization problem is noncontinuous and nonconvex. Thus, other classical techniques based on Taylor's expansion or convexity are not viable. Even in a multiple quantile regression without rank constraints, the same assumption is needed if we use (2.3) to perform the estimation, even though the estimator defined in (2.1) does not require this condition because of the convexity of the target function. (C4) assumes that the rank constraint is indeed satisfied.

We define the matrix $\mathbf{U} \equiv \text{diag}(\mathbf{U}_1, \ldots, \mathbf{U}_K)$, which is a $(p+1)K \times (p+1)K$ block-diagonal matrix, where the kth block is $\mathbf{U}_k = E\left[f_{\epsilon_{\tau_k}|\mathbf{X}}(0, \mathbf{X})\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\right]$. Thus, $\mathbf{U} = E\{\mathbf{f} \otimes (\mathbf{Z}\mathbf{Z}^{\mathrm{T}})\}$.

3.2 Estimation properties without rank constraints

We first consider the asymptotic properties of the quantile regression estimator obtained by solving (2.1), which are applicable when no rank constraint is imposed. The asymptotic property of this estimator is known in the statistics community, although we were not able to find a documented statement. Therefore, we state the results in the following theorem, for

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

completeness.

Theorem 1. *Let $\widehat{\boldsymbol{\theta}}$ be the estimator of $\boldsymbol{\theta}$ from (2.1). Then, under Conditions C1 and C2, when $n \to \infty$,*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to N(\mathbf{0}, \boldsymbol{\Sigma})$$

in distribution, where $\boldsymbol{\Sigma} = \mathbf{U}^{-1}(\mathbf{V} \otimes \mathbf{C})\mathbf{U}^{-1}$.

The proof of Theorem 1 is outlined in the online Supplementary Material.

Next, we consider the estimator from (2.3). Note that (2.2) and (2.4) are special cases of (2.3).

Theorem 2. *Let the estimator from (2.3) be $(\widehat{\mathbf{a}}, \widehat{\mathbf{B}})$, alternatively written as $\widehat{\boldsymbol{\theta}}$. Under conditions (C1)–(C3), when $n \to \infty$,*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to N(\mathbf{0}, \boldsymbol{\Sigma})$$

in distribution, where

$$\begin{aligned} \boldsymbol{\Sigma} &= [E\{(\mathbf{f}^* \otimes \mathbf{Z})\mathbf{V}^{-1}(\mathbf{f} \otimes \mathbf{Z}^{\mathrm{T}})\}]^{-1} E\{(\mathbf{f}^* \otimes \mathbf{Z})\mathbf{V}^{-1}(\mathbf{f}^* \otimes \mathbf{Z}^{\mathrm{T}})\} \\ &\quad [E\{(\mathbf{f} \otimes \mathbf{Z})\mathbf{V}^{-1}(\mathbf{f}^* \otimes \mathbf{Z}^{\mathrm{T}})\}]^{-1}. \end{aligned} \tag{3.1}$$

As special cases, the asymptotic variance for the estimator from (2.4) is

$$\begin{aligned} \boldsymbol{\Sigma} &= E[(\mathbf{I}_K \otimes \mathbf{Z})\mathbf{V}^{-1}(\mathbf{f} \otimes \mathbf{Z}^{\mathrm{T}})]^{-1} E[(\mathbf{I}_K \otimes \mathbf{Z})\mathbf{V}^{-1}(\mathbf{I}_K \otimes \mathbf{Z}^{\mathrm{T}})] \\ &\quad E[(\mathbf{f} \otimes \mathbf{Z})\mathbf{V}^{-1}(\mathbf{I}_K \otimes \mathbf{Z})^{\mathrm{T}})]^{-1}, \end{aligned} \tag{3.2}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

and the asymptotic variance for the estimator from (2.2) is

$$\mathbf{\Sigma} = E[(\mathbf{f} \otimes \mathbf{Z})\mathbf{V}^{-1}(\mathbf{f} \otimes \mathbf{Z}^{\mathrm{T}})]^{-1}. \tag{3.3}$$

Corollary 1. *Among the multiple quantile regression estimators given in (2.1) (2.2), (2.3), and (2.4), the best estimation variance of* $\widehat{\boldsymbol{\theta}}$ *is given in (3.3).*

The result in Corollary 1 provides a lower bound for the multiple quantile estimators discussed in Section 2.1, when no additional constraints are imposed on the relation between different quantile functions. Although, in theory, the variance can be as small as that given in (3.3), it is not necessarily easy to construct the corresponding estimator, owing to the need to estimate the conditional density $f_{\epsilon_\tau|\mathbf{X}}(0, \mathbf{x})$. In reality, the variance of the form given in (3.2) is more realistic and is usually achieved.

3.3 Estimation properties with rank constraints

Here, we further consider the rank-constrained multiple quantile regression estimators discussed in Sections 2.2 and 2.3. Assume the true rank of the matrix $\mathbf{B}$ is at most r. In this case, we can write $\mathbf{B} = \mathbf{DA}$, where $\mathbf{D} \in \mathcal{R}^{p\times r}$ and $\mathbf{A} \in \mathcal{R}^{r\times K}$ contain totally free parameters. Define the

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

$(p+1)K \times \{K + r(p+K)\}$ matrix Δ as

$$\boldsymbol{\Delta} = \frac{\partial \boldsymbol{\theta}}{\partial\{\mathbf{a}^{\mathrm{T}}, \mathrm{vec}^{\mathrm{T}}(\mathbf{A}^{\mathrm{T}}), \mathrm{vec}^{\mathrm{T}}(\mathbf{D}^{\mathrm{T}})\}}.$$

Theorem 3. *Under assumptions (C1)-(C4), the estimator obtained from (2.5) satisfies*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}),$$

when $n \to \infty$. *Here,* $\boldsymbol{\Sigma} = \mathbf{U}^{-1/2}\mathbf{P}_{\boldsymbol{\Phi}}\mathbf{U}^{1/2}\{\mathbf{U}^{-1}(\mathbf{V} \otimes \mathbf{C})\mathbf{U}^{-1}\}\mathbf{U}^{1/2}\mathbf{P}_{\boldsymbol{\Phi}}\mathbf{U}^{-1/2}$, $\boldsymbol{\Phi} = \mathbf{U}^{1/2}\boldsymbol{\Delta}$, *and* $\mathbf{P}_{\boldsymbol{\Phi}} = \boldsymbol{\Phi}(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{+}\boldsymbol{\Phi}^{\mathrm{T}}$, *where* $(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{+}$ *is the Moore-Penrose inverse of* $\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}$.

Remark 1. *We decompose the matrix* $\mathbf{B}$ *into* $\mathbf{DA}$ *to explicitly accommodate the rank constraint. The parameters in* $\mathbf{A}$ *and* $\mathbf{D}$ *can be any values. However, this decomposition is not unique, and* $\boldsymbol{\Delta}$ *does not have full column rank. Thus, we need to use the Moore–Penrose inverse rather than the usual inverse matrix. Furthermore, it can be shown that* $\boldsymbol{\Sigma}$ *depends on* $\boldsymbol{\Delta}$ *only through its column span.*

Remark 2. *When* $r = \min(p, K)$, $\boldsymbol{\Delta}$ *has full row rank and* $\mathbf{P}_{\boldsymbol{\Phi}}$ *is the identity matrix* $\mathbf{I}_{(p+1)K}$. *Therefore, the variance matrix in Theorem 3 is identical to that in Theorem 1. This agrees with the fact that when the rank constraint vanishes, the estimators from (2.1) and (2.5) are identical.*

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

Remark 3. *Although (2.5) is a direct implementation of (2.1). while incorporating the additional rank constraint property, the results in Theorems 1 and 3 indicate that (2.5) does not always yield a gain over (2.1) in terms of the estimation efficiency of* $\boldsymbol{\theta}$*. Instead, the efficiency gain is in the estimation of* $\mathbf{U}^{1/2}\boldsymbol{\theta}$*, in that the total estimation variability, described as the trace of the estimation variance matrix of* $\mathbf{U}^{1/2}\widehat{\boldsymbol{\theta}}$*, decreases when the rank constraint is imposed. This phenomenon also indicates that the rank constraint is not properly taken into account to maximally benefit the estimation of the quantile regression parameters in (2.5).*

Now, we consider the estimators obtained from (2.8), of which (2.6) and (2.7) are special cases.

Theorem 4. *Under assumptions (C1)–(C4), the estimator obtained from (2.8) satisfies*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to N(\mathbf{0}, \boldsymbol{\Sigma})$$

in distribution when $n \to \infty$*. Here,*

$$\begin{aligned} \boldsymbol{\Sigma} &= \boldsymbol{\Delta}(\boldsymbol{\Delta}^{\mathrm{T}} E[(\mathbf{f} \otimes \mathbf{Z})\mathbf{V}^{-1}(\mathbf{f}^* \otimes \mathbf{Z}^{\mathrm{T}})]\mathbf{W}_f^* \\ &\quad E[(\mathbf{f}^* \otimes \mathbf{Z})\mathbf{V}^{-1}(\mathbf{f} \otimes \mathbf{Z}^{\mathrm{T}})]\boldsymbol{\Delta})^{+}\boldsymbol{\Delta}^{\mathrm{T}}. \end{aligned} \tag{3.4}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

As special cases, the asymptotic variance for the estimator from (2.6) is

$$\begin{aligned}\boldsymbol{\Sigma} &= \boldsymbol{\Delta}(\boldsymbol{\Delta}^{\mathrm{T}} E[(\mathbf{f}\otimes\mathbf{Z})\mathbf{V}^{-1}(\mathbf{I}_K\otimes\mathbf{Z}^{\mathrm{T}})]\mathbf{W} \\ &\quad E[(\mathbf{I}_K\otimes\mathbf{Z})\mathbf{V}^{-1}(\mathbf{f}\otimes\mathbf{Z}^{\mathrm{T}})]\boldsymbol{\Delta})^{+}\boldsymbol{\Delta}^{\mathrm{T}}. \qquad (3.5)\end{aligned}$$

and the asymptotic variance for the estimator from (2.7) is

$$\boldsymbol{\Sigma} = \boldsymbol{\Delta}(\boldsymbol{\Delta}^{\mathrm{T}} E[(\mathbf{f}\otimes\mathbf{Z})\mathbf{V}^{-1}(\mathbf{f}\otimes\mathbf{Z}^{\mathrm{T}})]\boldsymbol{\Delta})^{+}\boldsymbol{\Delta}^{\mathrm{T}}. \qquad (3.6)$$

Corollary 2. *Among the multiple rank constrained quantile regression estimators given in (2.5), (2.6), (2.7), and (2.8), the best estimation variance of $\widehat{\boldsymbol{\theta}}$ is given in (3.6).*

The result in Corollary 2 provides a lower bound for the multiple quantile estimators, subject to the rank constraint discussed in Sections 2.2 and 2.3. Although, in theory, the variance can be as small as that given in (3.6), as in the no-constraint case, it is not necessarily easy to construct the corresponding estimator, because we have to estimate the conditional density $f_{\epsilon_\tau|\mathbf{X}}(0,\mathbf{x})$. In reality, the variance form in (3.5) is more readily achieved.

We can see that Theorems 3 and 4 contain the estimator properties under the rank constraint corresponding to those stated in Theorems 1 and 2 without any constraint. It is clear from Remarks 2 and 3 that (2.1) and (2.5) are equivalent when the rank constraint is removed, but that they differ when the rank constraint holds. The same holds for the pairs (2.8)

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

and (2.3), (2.7) and (2.2), and (2.6) and (2.4). When the rank constraint is removed, the two estimators in each pair are identical. However, when the rank constraint holds, the estimators from (2.8), (2.7), and (2.6) are more efficient than their respective counterparts, (2.3), (2.2), and (2.4). These conclusions are easily verified by noting that for any positive-definite matrix $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Delta}(\boldsymbol{\Delta}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Delta})^{+}\boldsymbol{\Delta}^{\mathrm{T}}$ is nonnegative definite because the matrix

$$\begin{pmatrix} \boldsymbol{\Sigma}^{-1} & \boldsymbol{\Delta} \\ \boldsymbol{\Delta}^{\mathrm{T}} & \boldsymbol{\Delta}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Delta} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}^{\mathrm{T}}\boldsymbol{\Sigma}^{1/2} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Delta} \end{pmatrix}$$

is nonnegative definite (see the general matrix result in the online Supplement Material), and $\boldsymbol{\Delta}$ has full row rank when $r = \min(p, K)$. We emphasize these relations in Remarks 4 and 5.

Remark 4. *When $r = \min(p, K)$, the variance matrices in (3.4), (3.5), and (3.6) are identical to those in (3.1), (3.2), and (3.3), respectively.*

Remark 5. *When $r < \min(p, K)$, the variance matrices in (3.4), (3.5), and (3.6) are smaller than those in (3.1), (3.2), and (3.3) respectively. Here, a matrix $\mathbf{A}_1$ being smaller than a matrix $\mathbf{A}_2$, we mean that $\mathbf{A}_2 - \mathbf{A}_1$ is nonnegative definite.*

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

3.4 Results on rank determination

Finally, we show that the BICs proposed in Section 2.5 can consistently estimate the rank of the true coefficient matrix $\mathbf{B}$.

Theorem 5. *Under assumptions (C1)–(C4), the BIC criteria (2.11) and (2.12) select the true rank r with probability approaching one when the sample size n approaches infinity.*

The consistency of the two BIC methods in determining the rank indicates that our rank-constrained quantile estimation procedure can be used in a model that takes full advantage of the data properties, but without imposing artificial structures. This is because if the true rank is larger than that under which we conduct our estimation, bias will occur. However, if the true rank is smaller, we are not taking full advantage of the data structure, in which case the estimation variance may be inflated. Note that this inflation may or may not happen for the estimator in (2.5), but will happen for the estimators in Section 2.3. Here, the amount of the inflation is the difference between the estimation variances given in Theorem 4 corresponding to different row sizes of $\mathbf{\Delta}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

4. Simulation Experiments

We now assess the finite-sample performance of the methods discussed in Section 2. In all of the examples, we set the dimension of $\boldsymbol{\beta}_\tau$ to $p = 7$ and we use a sample size of $n = 200$. To obtain the covariates, we first generate $(X^*_{i1}, \ldots, X^*_{ip})$, $i = 1, \ldots, n$, from a zero-mean multivariate Gaussian distribution with $\text{cov}(X^*_{ij}, X^*_{ij'}) = 0.3^{|j-j'|}$. Then, we set $X_{ij} = \Phi(X^*_{ij})$, where Φ is the cdf of the standard normal distribution. The simulation is repeated 100 times in each setting, and we use the quantile sequence $(0.1, 0.2, \ldots, 0.9)$ in all examples; hence, $K = 9$.

Example 1. Location-shift model. $Y_i = 1 + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} + \epsilon_i$, where $\epsilon_i \sim N(0, 1), i = 1, \ldots, n$, and $\boldsymbol{\beta} = (1, -1, 1, -1, 1, -1, 1, -1)^{\mathrm{T}}$. The true rank of the coefficient matrix is $r = 1$.

Example 2. Location-scale shift model. $Y_i = 1 + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} + (1 + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\gamma})\epsilon_i$, where $\epsilon_i \sim N(0, \sigma^2), i = 1, \ldots, n$, $\sigma = 0.3$, $\boldsymbol{\beta} = (2, 2, \ldots, 2)^{\mathrm{T}}$, and $\boldsymbol{\gamma} = (2, 2, 2, 0, 0, 0, 0)^{\mathrm{T}}$. The true rank of the coefficient matrix is $r = 2$.

Example 3. The conditional quantile is given by $Q_\tau(\mathbf{X}) = \Phi^{-1}(\tau) + \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + (\mathbf{X}^{\mathrm{T}}\boldsymbol{\gamma})(\Phi^{-1}(\tau) - \Phi^{-1}(0.49))I\{\tau < 0.49\}$, where $\boldsymbol{\beta} = (-2, -1, 0, 1, 2, 3, 4)^{\mathrm{T}}$ and $\boldsymbol{\gamma} = (2, 2, \ldots, 2)^{\mathrm{T}}$. The true rank of the coefficient matrix is $r = 2$. We use the inverse cdf method to generate the responses.

In each of the three examples, we compare the following three sets of

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

estimators. The first set of estimators do not take into account the rank constraints. They include the standard quantile regression described in (2.1) (NAIVE), quantile estimator based on equation (2.4) (GMM), optimal quantile estimator based on equation (2.2), where the density values are estimated as in Section 2.4 (OPT), and optimal estimator without rank constraint based on (2.2), where we substitute in the true density values (ORACLE). The second set of estimators takes into account the rank constraints. They include the naive reduced-rank estimator described in (2.5) (NAIVE.RR), proposed reduced-rank GMM estimator in (2.6) (GMMRR), and optimal reduced-rank GMM estimator in (2.7), substituting in the estimated (OPT.RR) and true (ORACLE.RR) density values. Finally, we compute the composite quantile regression estimator of Zou and Yuan (2008a) (CQR) as the third set of estimators.

We use three measures to examine the performance of the various estimators. The first measure is the mean squared error of the estimated coefficients

$$MSE = \sum_{k=1}^{K} |\widehat{a}_k - a_{0k}|^2 + \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2.$$

The second measure is the integrated squared error

$$ISE = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \left\{ \widehat{a}_k + \mathbf{X}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}} - (a_{0k} + \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}_{0k}) \right\}^2.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

The third measure is the quantile prediction error

$$PE = \frac{1}{500}\sum_{i=1}^{500}\sum_{k=1}^{K}\rho_{\tau_k}\{Y_i' - \widehat{a}_k - (\mathbf{X}_i')^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_k\},$$

where $(Y_i', \mathbf{X}_i')$, for $i = 1, \ldots, 500$ are independently generated test data.

The simulation results for the different methods are reported in Table 1. We see clearly that the reduced-rank methods improve upon the full-rank methods in all cases. The GMM-based methods are generally better than the methods based on the check loss functions, although the improvement is not as large as that of the rank constraints. The optimal estimating equation using the estimated density values is asymptotically best within the GMM family. However, its finite-sample performance is similar, or even slightly worse, than the proposed GMMRR method, where the density values are simply set to one. In fact, even in the oracle case, where the true density is substituted into the GMM estimator, the performance is only slightly better than that of the GMMRR. This suggests that although the estimator (2.7) is theoretically best, the poor estimation of the density makes it uncompetitive compared with that give in (2.6). Thus we advocate using (2.6). Finally, CQR is better in Example 1, as expected, because the model satisfies the requirement set by CQR. However, it performs worse in the other two examples.

We further investigate the accuracy of the standard error estimates

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

Table 1: Simulation results for Examples 1-3 based on 100 generated data sets.

Method	MSE	ISE	PE
Example 1			
naive(2.1)	14.075(4.493)	6.683(0.713)	2.763(0.075)
GMM(2.4)	13.935(4.472)	6.516(0.664)	2.754(0.072)
OPT(2.2)	13.996(4.504)	6.466(0.694)	2.754(0.073)
ORACLE(2.2)	13.350(4.288)	6.232(0.639)	2.639(0.069)
naive.RR(2.5)	12.057(4.085)	5.522(0.734)	2.523(0.068)
GMMRR(2.6)	10.598(3.824)	4.832(0.710)	2.398(0.067)
OPT.RR(2.7)	10.694(3.841)	4.871(0.750)	2.399(0.065)
ORACLE.RR(2.7)	10.349(3.772)	4.791(0.732)	2.349(0.066)
CQR	8.869(2.832)	4.615(0.562)	2.344(0.063)
Example 2			
naive(2.1)	19.663(6.831)	1.479(0.453)	4.137(0.110)
GMM(2.4)	19.150(6.836)	1.423(0.462)	4.128(0.110)
OPT(2.2)	19.206(6.832)	1.437(0.466)	4.129(0.109)
ORACLE(2.2)	18.727(6.555)	1.365(0.444)	3.956(0.105)
naive.RR(2.5)	15.790(6.332)	1.164(0.422)	3.770(0.103)
GMMRR(2.6)	15.382(6.281)	1.140(0.405)	3.714(0.100)
OPT.RR(2.7)	15.490(6.307)	1.146(0.417)	3.736(0.106)
ORACLE.RR(2.7)	15.251(6.121)	1.118(0.391)	3.676(0.097)
CQR	26.261(6.798)	1.598(0.469)	4.324(0.119)
Example 3			
naive(2.1)	19.444(6.262)	1.658(0.506)	3.689(0.116)
GMM(2.4)	19.018(6.248)	1.619(0.521)	3.679(0.113)
OPT(2.2)	19.268(6.252)	1.636(0.518)	3.681(0.114)
ORACLE(2.2)	18.802(5.999)	1.567(0.485)	3.528(0.109)
naive.RR(2.5)	16.404(5.912)	1.437(0.478)	3.367(0.104)
GMMRR(2.6)	14.785(5.333)	1.302(0.440)	3.104(0.097)
OPT.RR(2.7)	14.971(5.376)	1.323(0.449)	3.109(0.093)
ORACLE.RR(2.7)	14.679(5.308)	1.282(0.431)	3.049(0.099)
CQR	30.362(4.115)	1.941(0.468)	3.894(0.125)

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

Table 2: Comparison of the standard error estimates based on asymptotic normality using (3.5) ($\widehat{\text{se}}$) and the standard error calculated based on the sample standard deviation of 100 repetitions (se).

Method	$\tau = 0.1$		$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.7$		$\tau = 0.9$	
Example 1	β_1	β_4	β_1	β_4	β_1	β_4	β_1	β_4	β_1	β_4
$\widehat{\text{se}}$	0.206	0.213	0.161	0.157	0.157	0.154	0.158	0.154	0.206	0.213
se	0.223	0.221	0.175	0.174	0.167	0.167	0.175	0.176	0.220	0.220
Example 2	β_1	β_4	β_1	β_4	β_1	β_4	β_1	β_4	β_1	β_4
$\widehat{\text{se}}$	0.276	0.282	0.211	0.212	0.211	0.209	0.215	0.209	0.275	0.287
se	0.306	0.303	0.237	0.234	0.228	0.225	0.238	0.237	0.300	0.300
Example 3	β_1	β_4	β_1	β_4	β_1	β_4	β_1	β_4	β_1	β_4
$\widehat{\text{se}}$	0.492	0.487	0.353	0.347	0.146	0.157	0.051	0.054	0.067	0.063
se	0.504	0.516	0.341	0.348	0.173	0.175	0.062	0.063	0.067	0.069

Table 3: Percentage of times that the correct rank is selected, for Simulation Examples 1–3.

Example 1			Example 2			Example 3		
naive.RR	GMMRR	OPT.RR	naive.RR	GMMRR	OPT.RR	naive.RR	GMMRR	OPT.RR
76	82	82	79	77	74	68	68	70

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

associated with the proposed GMMRR estimator, described in Theorem 4. Although the GMMRR estimation does not require the density function of Y, conditional on the covariates, an evaluation of its variability does require this quantity, which we estimate as in Section 2.4. The results are presented in Table 2. In the table, $\widehat{\text{se}}$ denotes the estimated standard error based on the asymptotic results in (3.5), and se denotes the sample standard deviation of the coefficient value based on 100 repetitions. For illustration purposes, we only show the results for β_1 and β_4 at five quantile levels $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. In general, the estimated standard errors based on the asymptotic results are reasonably close to their sample versions.

Finally, we experiment with the methods of determining the rank of the coefficient matrix. We report the percentage of times that the true rank is selected by the BIC in Table 3 in the three examples. We can see that the correct rank is selected most of the time, and that the performance improves as the sample size increases.

Experiments with sample sizes of $n = 100$ and 400 show similar comparative results for the estimators. The results are not reported here to save space.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

5. Real-Data Application

We now use various quantile regression procedures to analyze the Barro growth data used in Koenker and Machado (1999). The data consist of a pooled sample of 161 observations related to national growth rates during two periods, 1965–1975 and 1975–1985. The goal of the analysis is to understand the effect of 13 covariates in terms of their influence on the growth process at different conditional quantiles levels.

We apply the standard quantile regression (2.1) and the proposed methods to the data set. All variables are standardized before the analysis. To compare the performance of the different methods, we consider predictions in which we randomly sample 100 observations for training, and use the rest for testing, with 100 repetitions. The prediction errors are reported in Table 4, as well as the average rank (mostly 2 or 3) selected for each reduced-rank method. We find that the reduced-rank methods perform uniformly better, with smaller prediction errors, than the methods that do not take into account the possibility of a reduced rank.

Table 4: Prediction errors and selected rank for the Barro data.

	naive (2.1)	GMM(2.4)	OPT(2.2)	naive.RR(2.5)	GMMRR(2.6)	OPT.RR(2.7)	CQR
PE	14.31	14.14	14.12	13.74	13.60	13.65	13.90
rank	NA	NA	NA	2.30	2.08	2.20	NA

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

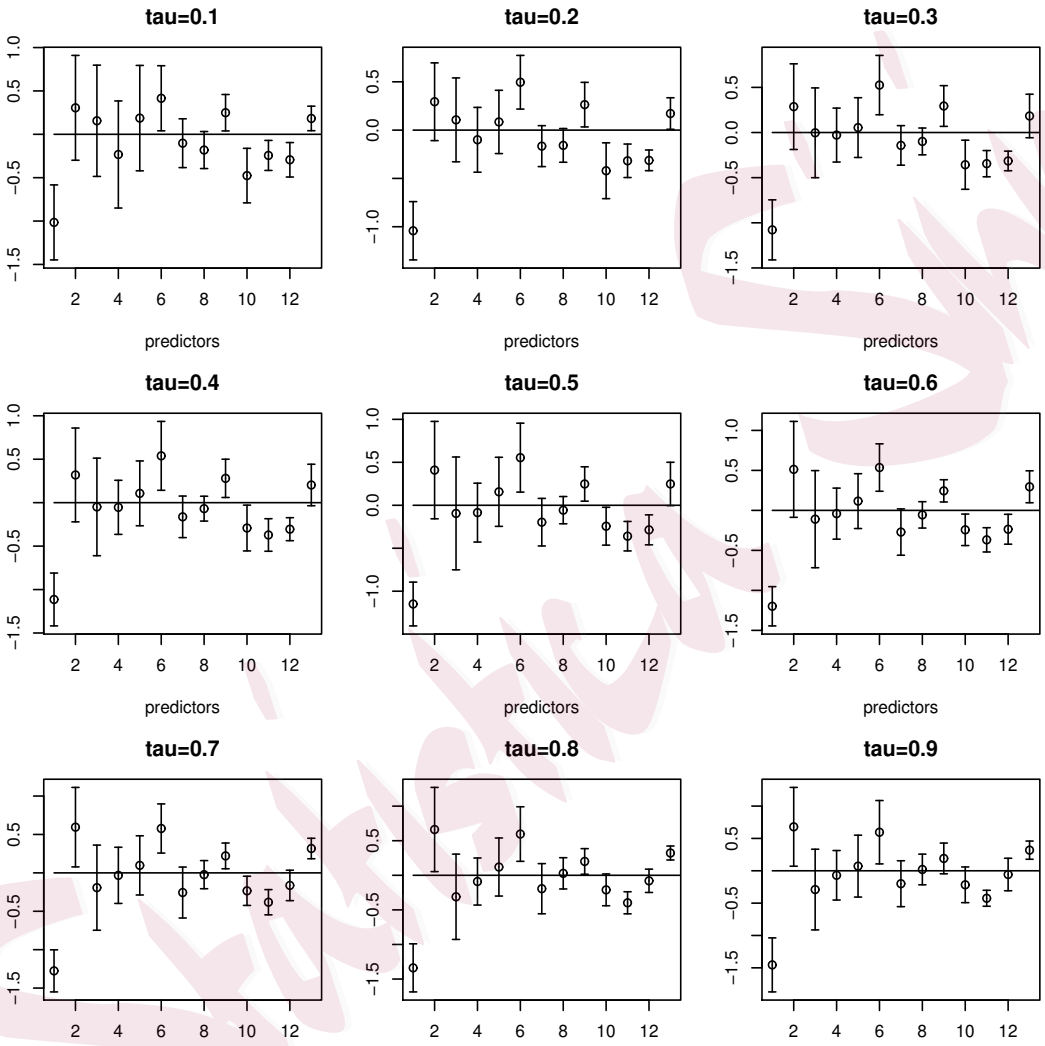

Figure 1: Estimates and 95% confidence interval for the Barro data, using the standard quantile regression in (2.1).

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

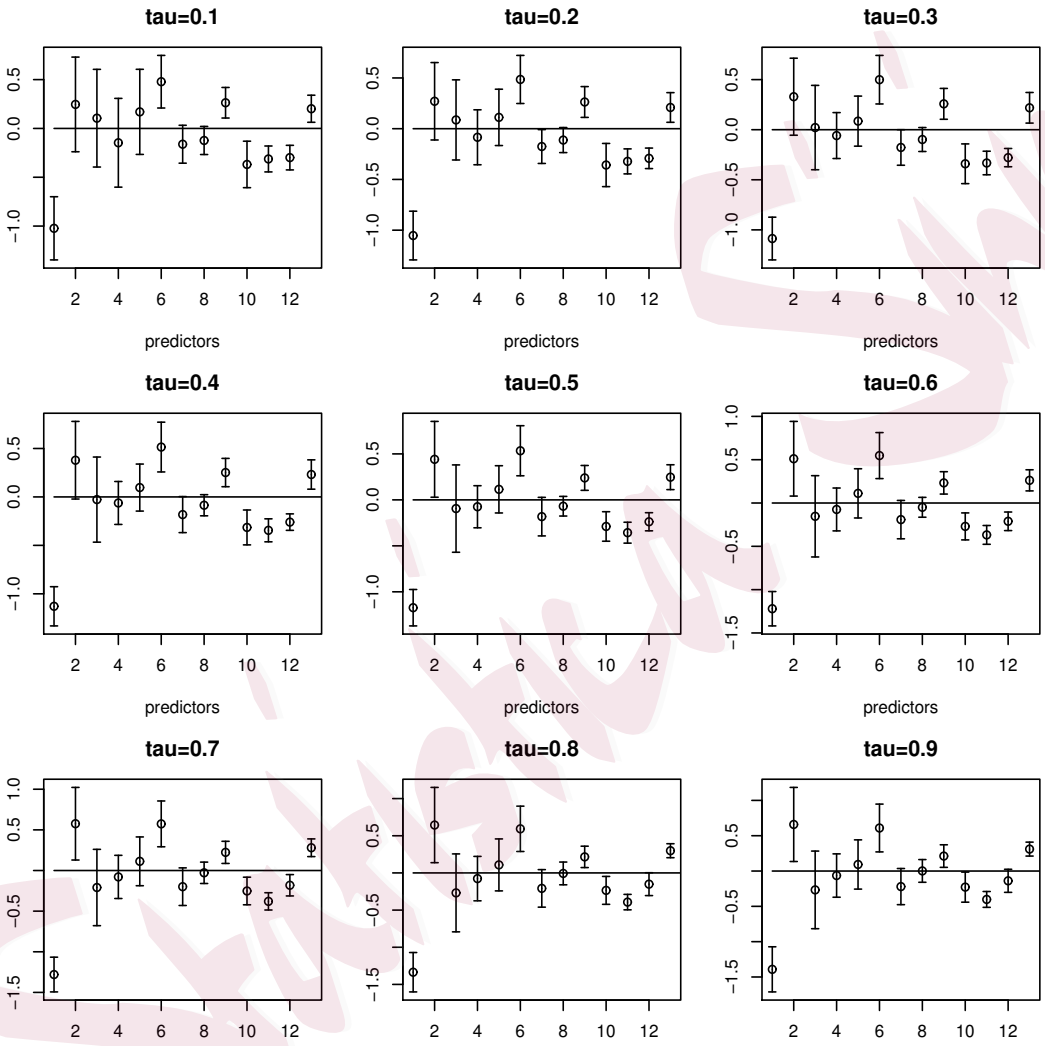

Figure 2: Estimates and 95% confidence interval for the Barro data, using the GMMRR method in (2.6).

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

We provide the estimated coefficients with 95% confidence intervals for the standard quantile regression (2.1) and the proposed GMMRR method (2.6) in Figures 1 and 2, respectively. Here, we find that the intervals for the GMMRR method are generally shorter, resulting in more variables that are statistically significant. For example, using the standard quantile regression, the last variable (growth rate terms trade) is found to be significant at the 0.05 level for $\tau = 0.1, 0.2, 0.6, 0.7, 0.8, 0.9$, whereas it is found to be significant at all $\tau \in \{0.1, \ldots, 0.9\}$ when using GMMRR.

6. Discussion

In this study, we investigated a new approach for simultaneously estimating multiple conditional quantiles, motivated by a reduced-rank regression. We derived the most efficient estimating equations based on the GMM principle, and proposed an estimator that is easy to compute and performs nearly as well as the most efficient estimator. The proposed method improves the efficiency by sharing information across quantiles levels and is shown empirically to attain estimates for the coefficients that are more accurate.

Empirically, we do not encounter any problems with numerical convergence. However, we are not able to provide an additional convergence analysis of the algorithm. For (2.9), in addition to the derivative of the loss

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

not being smooth, another technical problem is to show that the computed quantities are bounded. For (2.10), the problem is nonconvex in $\mathbf{D}$ or $\mathbf{A}$ (when the other is fixed), making it even more difficult to remark on the convergence. Fortunately, for our problem, it is trivial to obtain a good initial estimator using a standard quantile regression, and, thus we are working in a small neighborhood of the optimum. As a check-loss-based problem (2.9), as suggested by a reviewer, using an existing MM algorithm for the quantile regression, as in Hunter and Lange (2000), the quantile loss can be majorized by a quadratic function. Thus, it might be possible to update $\mathbf{B}$ directly, which may make it easier to conduct a convergence analysis. However, this does not seem to work for the GMM estimator and, thus, we choose to use the alternating update algorithm for both the loss-based and the GMM estimator.

In this study, we focus on the prediction performance of the proposed method, which serves as its main motivation. By the well-known relationships between a reduced-rank regression and a factor analysis, it is tempting to interpret $\mathbf{X}^{\mathrm{T}}\mathbf{D}$ as some type of low-dimensional factor. However, it is not clear whether this is useful in practice, and the fact that solutions with different rank values are not nested, in general, makes interpretation even more difficult. Further investigations in this direction may be worthwhile.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

Several other extensions of the proposed methodology are of interest. For example, it can be extended to the semiparametric quantile models studied in Kim (2007); Wang et al. (2009); Lian (2012), among others, to improve efficiency. A problem that is not addressed in the current study is the problem of crossing quantiles. Many methods have been proposed in the literature to deal with this (He; 1997; Dette and Volgushev; 2008; Bondell et al.; 2010; Chernozhukov et al.; 2010). How to combine these methods with our approach is an interesting research topic. Finally, it is worthwhile considering the case when p is diverging with the sample size, given the level of interest in high-dimensional analyses in the statistical community.

Supplementary Material

The online supplementary material contains the proofs of all theorems and corollaries.

Acknowledgements The authors sincerely thank the Co-Editor Professor Zhiliang Ying, an Associate Editor, and two reviewers for their insightful comments, which have lead to significant improvements of the manuscript. The research of Heng Lian is partially supported by City University of Hong Kong Start Up Grant 7200521, by Hong Kong RGC general research fund 11301718, and by Project 11871411 from NSFC and the Shenzhen Research

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

Institute of the City University of Hong Kong. The research of Weihua Zhao is supported by National Social Science Fund of China (15BTJ027).

References

Anderson, T. (1999). Asymptotic distribution of the reduced rank regression estimator under general conditions, *The Annals of Statistics* **27**(4): 1141–1154.

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions, *The Annals of Mathematical Statistics* **29**: 327–351.

Belloni, A. and Chernozhukov, V. (2011). l1-penalized quantile regression in high-dimensional sparse models, *The Annals of Statistics* **39**(1): 82–130.

Bondell, H. D., Reich, B. J. and Wang, H. (2010). Noncrossing quantile regression curve estimation, *Biometrika* **97**(4): 825–838.

Brown, B. and Wang, Y.-G. (2005). Standard errors and covariance matrices for smoothed rank estimators, *Biometrika* **92**(1): 149–158.

Bunea, F., She, Y. and Wegkamp, M. H. (2011). Optimal selection of reduced rank estimators of high-dimensional matrices, *The Annals of Statistics* **39**(2): 1282–1309.

Bunea, F., She, Y. and Wegkamp, M. H. (2012). Joint variable and rank selection for parsimonious estimation of high-dimensional matrices, *The Annals of Statistics* **40**(5): 2359–2388.

Cade, B. and Noon, B. (2003). A gentle introduction to quantile regression for ecologists, *Frontiers in Ecology and the Environment* **1**(8): 412–420.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

Chen, K., Chan, K. S. and Stenseth, N. C. (2012). Reduced rank stochastic regression with a sparse singular value decomposition, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74**(2): 203–221.

Chen, K., Dong, H. and Chan, K. S. (2013). Reduced rank regression via adaptive nuclear norm penalization, *Biometrika* **100**: 901–920.

Chen, L. and Huang, J. Z. (2012). Sparse reduced-rank regression for simultaneous dimension reduction and variable selection, *Journal of the American Statistical Association* **107**(500): 1533–1545.

Chernozhukov, V., Fernandez-Val, I. and Galichon, A. (2010). Quantile and probability curves without crossing, *Econometrica* **78**(3): 1093–1125.

De Gooijer, J. G. and Zerom, D. (2003). On additive conditional quantiles with high-dimensional covariates, *Journal of the American Statistical Association* **98**(461): 135–146.

Dette, H. and Volgushev, S. (2008). Noncrossing nonparametric estimates of quantile curves, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**(3): 609–627.

Geweke, J. (1996). Bayesian reduced rank regression in econometrics, *Journal of Econometrics* **75**(1): 121–146.

Hall, P. and Sheather, S. J. (1988). On the distribution of a studentized quantile, *Journal of the Royal Statistical Society. Series B (Methodological)* **50**(3): 381–391.

He, Q., Kong, L., Wang, Y., Wang, S., Chan, T. A. and Holland, E. (2016). Regularized

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

quantile regression under heterogeneous sparsity with application to quantitative genetic traits, *Computational Statistics & Data Analysis* **95**: 222–239.

He, X. (1997). Quantile curves without crossing, *The American Statistician* **51**(2): 186–192.

He, X. and Shi, P. (1994). Convergence rate of b-spline estimators of nonparametric conditional quantile functions, *Journal of Nonparametric Statistics* **3**: 299–308.

He, X., Wang, L. and Hong, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data, *The Annals of Statistics* **41**(1): 342–369.

Hendricks, W. and Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity, *Journal of the American Statistical Association* **87**(417): 58–68.

Horowitz, J. L. (1998). Bootstrap methods for median regression models, *Econometrica* **66**(6): 1327–1351.

Horowitz, J. L. and Lee, S. (2005). Nonparametric estimation of an additive quantile regression model, *Journal of the American Statistical Association* **100**(472): 1238–1249.

Hunter, D. R. and Lange, K. (2000). Quantile regression via an mm algorithm, *Journal of Computational and Graphical Statistics* **9**(1): 60–77.

Jiang, L., Bondell, H. D. and Wang, H. J. (2014). Interquantile shrinkage and variable selection in quantile regression, *Computational Statistics & Data Analysis* **69**: 208–219.

Jiang, L., Wang, H. J. and Bondell, H. D. (2013). Interquantile shrinkage in regression models, *Journal of Computational and Graphical Statistics* **22**(4): 970–986.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

Kim, M. (2007). Quantile regression with varying coefficients, *Annals of Statistics* **35**(1): 92–108.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles, *Econometrica: Journal of the Econometric Society* **1**(46): 33–50.

Koenker, R. and Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression, *Journal of the American Statistical Association* **94**(448): 1296–1306.

Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for bayesian quantile regression, *Journal of Statistical Computation and Simulation* **81**(11): 1565–1578.

Lee, S. (2003). Efficient semiparametric estimation of a partially linear quantile regression model, *Econometric Theory* **19**: 1–31.

Lian, H. (2012). Semiparametric estimation of additive quantile regression models by two-fold penalty, *Journal of Business & Economic Statistics* **30**(3): 337–350.

Lian, H. and Ma, S. (2013). Reduced-rank regression in sparse multivariate varying-coefficient models with high-dimensional covariates, *arXiv preprint arXiv:1309.6058* .

Newey, W. K. and Powell, J. L. (1990). Efficient estimation of linear and type i censored regression models under conditional quantile restrictions, *Econometric Theory* **6**: 295–317.

Wang, H. and He, X. (2007). Detecting differential expressions in genechip microarray studies: a quantile approach, *Journal of the American Statistical Association* **102**(477): 104–112.

Wang, H. J., Zhu, Z. and Zhou, J. (2009). Quantile regression in partially linear varying coefficient models, *The Annals of Statistics* **37**(6B): 3841–3866.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

Wang, L. and Qu, A. (2009). Consistent model selection and data-driven smooth tests for longitudinal data in the estimating equations approach, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**(1): 177–190.

Wang, L., Wu, Y. and Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension, *Journal of the American Statistical Association* **107**(497): 214–222.

Yang, Y. and He, X. (2012). Bayesian empirical likelihood for quantile regression, *The Annals of Statistics* **40**(2): 1102–1131.

Yu, K., Lu, Z. and Stander, J. (2003). Quantile regression: applications and current research areas, *Journal of the Royal Statistical Society: Series D* **52**(3): 331–350.

Yu, K. and Moyeed, R. (2001). Bayesian quantile regression, *Statistics & Probability Letters* **54**(4): 437–447.

Zhao, Z. and Xiao, Z. (2014). Efficient regressions via optimally combining quantile information, *Econometric heory* **30**(06): 1272–1314.

Zou, H. and Yuan, M. (2008a). Composite quantile regression and the oracle model selection theory, *The Annals of Statistics* **36**(3): 1108–1126.

Zou, H. and Yuan, M. (2008b). Regularized simultaneous model selection in multiple quantiles regression, *Computational Statistics & Data Analysis* **52**(12): 5296–5304.

Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong

E-mail: henglian@cityu.edu.hk

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0426

School of Science, Nantong University, Nantong, China, 226019

E-mail: zhaowhstat@163.com

Department of Statistics, Pennsylvania State University, University Park, PA, USA 16802

E-mail: yanyuanma@gmail.com