Statistica Sinica Preprint No: SS-2018-0060	
Title	Envelope Quantile Regression
Manuscript ID	SS-2018-0060
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0060
Complete List of Authors	Shanshan Ding Zhihua Su Guangyu Zhu and Lan Wang
Corresponding Author	Shanshan Ding
E-mail	sding@udel.edu
Notice: Accepted version subject to English editing.	

Envelope Quantile Regression *

Shanshan Ding, Zhihua Su, Guangyu Zhu and Lan Wang

Abstract

Quantile regression offers a valuable complement of classical mean regression for robust and comprehensive data analysis in a variety of applications. We propose a novel *envelope quantile regression* method (EQR) that adapts a nascent technique called *enveloping* (Cook, Li, and Chiaromonte, 2010) to improve the efficiency of standard quantile regression. The new method aims to identify material and immaterial information in a quantile regression model and use only the material information for estimation. By excluding the immaterial part, the EQR method has the potential to substantially reduce the estimation variability. Unlike existing envelope model approaches which mainly rely on the likelihood framework, our proposed estimator is defined through a set of nonsmooth estimating equations. We facilitate the estimation via the generalized method of moments (GMM) and derive the asymptotic normality of the proposed estimator by applying empirical process techniques. Furthermore, we establish that EQR is asymptotically more efficient than (or at least as asymptotically efficient as) the standard quantile regression estimators without imposing stringent conditions. Hence, our work advances the envelope model theory to general distribution-free settings. We demonstrate the effectiveness of the proposed method via Monte-Carlo simulations and real data examples.

*Shanshan Ding and Zhihua Su are co-first authors. Shanshan Ding is Assistant Professor, Department of Applied Economics and Statistics, University of Delaware. Zhihua Su is Associate Professor, Department of Statistics, University of Florida. Guangyu Zhu is Assistant Professor, Department of Computer Science and Statistics, University of Rhode Island. Lan Wang is Professor, School of Statistics, University of Minnesota. The research of Lan Wang is supported by National Science Foundation grant DMS-1512267. The research of Zhihua Su is supported by National Science Foundation grant DMS-1407460.

Key Words: sufficient dimension reduction; envelope model; reducing subspace; generalized method of moments, asymptotic efficiency.

1 Introduction

Envelopes were first proposed by Cook, Li, and Chiaromonte (2010) for response reduction and parsimonious estimation in multivariate linear regression with normal errors. In this setting, the envelope approach has been proved to achieve asymptotic efficiency and reduce estimation variability compared to standard methods. Since then, a variety of envelope models have been developed and demonstrated promising performances in multivariate statistical problems. For example, Su and Cook (2011, 2012) and Cook and Su (2013) subsequently studied envelope methods for different data structure in linear regression models. Cook, Helland, and Su (2013) used envelopes to study predictor reduction and established a connection between envelope models and partial least squares. Based on the connection, Zhu and Su (2019) derived the envelope-based sparse partial least squares method. Cook, Forzani, and Zhang (2015) applied the envelope method to reduced rank regression. Cook and Zhang (2015) extended the applicability of envelope model beyond linear regression, such as in generalized linear regression and cox proportional hazard model. Su *et al.* (2016) proposed sparse envelope models for variable selection in the multivariate linear regression setting. Khare, Pal, and Su (2017) developed Bayesian envelope approaches. Li and Zhang (2017) and Ding and Cook (2018) proposed envelopes for tensor and matrix regression problems. Envelopes for spatial and time series data were studied by Rekabdarkolaee *et al.*

(2017) and Wang and Ding (2018).

The existing works, however, were developed mainly for mean regression and likelihood-based models. The estimation and inference thus often rely on the maximum likelihood principle. In particular, the asymptotic efficiency of the envelope estimators often requires normality assumption. One major objective of this article is to develop a new non-likelihood based framework for enveloping and to advance the envelope theory to general distribution-free settings to potentially improve efficiency. Although developed in the context of quantile regression, our framework can also be extended to other statistical methods and procedures.

Quantile regression (Koenker and Bassett, 1978; Koenker, 2005) is a popular regression technique and is widely used in economics, health sciences and many other fields. It does not require distributional assumptions on error terms and thus is a flexible distribution-free regression technique. By accommodating varying covariate effects at different quantile levels, quantile regression is able to provide a more complete picture of the relationship between the response variable and covariates. In addition, It is capable to incorporate heterogeneous covariates effects and is robust to outliers. Because of its good statistical properties and flexibility in practice, quantile regression has become a popular alternative to least squares regression, and has gained considerable interest in recent years. For example, Knight (1998), He and Shao (2000), Chernozhukov (2005), He and Zhu (2011), Feng, He, and Hu (2011), Yang and He (2012) and many others have extensively studied the theoretical properties and inference tools of quantile regression under different settings. Portnoy and Koenker (1997), Chen and Wei (2005), Koenker (2011) and others have investigated the computational perspectives of quantile regression. Quantile regression has

also been extended to longitudinal and survival data analysis (He *et al.*, 2002; Portnoy, 2003; Wei *et al.*, 2006; Peng and Huang, 2008; Wang and Wang, 2009; Xu *et al.*, 2017, among many others). We refer to Koenker (2005, 2017) and Koenker *et al.* (2017) for a comprehensive review on quantile regression.

In this article, we propose a new approach called *envelope quantile regression* (EQR) that adapts a nascent technique called *enveloping* (Cook, 2018) by introducing dimension reduction into quantile modeling. In a variety of settings, it is reasonable to assume there exist linear combinations of the predictors which are irrelevant to the conditional quantiles of the response and these combinations do not affect the conditional quantiles through their association with the remaining combinations. Thus, we can focus on a subspace of the full predictor space that is directly relevant to the model fitting. We call the relevant part of the predictors as material information and the remaining part as immaterial information. Using immaterial information in model fitting is likely to increase estimation variation. The new EQR approach does not change the traditional objectives of quantile regression, but by fully utilizing information in both predictors and response, it can distinguish such material and immaterial information when modeling the conditional quantiles, and synchronously exclude immaterial information from model estimation. With this simultaneous dimension reduction and regression fitting, the EQR method can improve estimation efficiency and the efficiency gains can be substantial when the immaterial variation is large.

The main contribution of this article is three-fold. First, we develop a new EQR approach that adapts the ideas of enveloping to quantile regression and achieves efficiency gains. We prove that the EQR estimator is $\sqrt{n}$-consistent and asymptotically normal, and

more importantly, it is asymptotically more efficient than (or at least as asymptotically efficient as) the standard quantile regression estimators without imposing stringent conditions. In addition, whilst we mainly focus on linear quantile regression in this paper, our approach can be naturally extended to partially linear quantile regression, censored quantile regression and other settings. Second, our formulation offers the first non-likelihood based envelope method with theoretical justification on asymptotical efficiency. It establishes a new framework for enveloping and advances the recent development of envelopes to general distribution-free procedures with possibly nonsmooth objective functions. Third, the theoretical development of the EQR estimator is based on rather different techniques than those used in existing envelope models and quantile regression. The proposed estimator is defined through a set of nonsmooth estimating equations. We facilitate the estimation via the generalized method of moments (GMM) that not only ensures desirable theoretical properties but further improves asymptotic efficiency of the estimators. Empirical process techniques are employed for establishing asymptotics, which can be used to handle both nonsmooth and over-parametrized models, and can be potentially applied to more complex enveloping problems.

Most existing envelope approaches are focused on continuous variables, so does the EQR method. When categorical predictors are present, we develop the partial envelope quantile model that applies the enveloping idea only to the continuous predictors. We will show that the partial envelope quantile model improves the estimation efficiency of the regression coefficients, especially those of the continuous predictors.

The rest of the article is organized as follows. In Section 2, we briefly review linear

quantile regression and propose the EQR method. In Section 3, we establish theoretical properties for the EQR estimators and demonstrate their efficiency. Section 4 presents the new GMM estimation procedure and discusses dimension selection procedures for the proposed EQR. Section 5 demonstrates the empirical performance of the EQR method via simulations and real examples. Section 6 is devoted to the development of partial envelope regression quantile regression for data with categorical predictors. We conclude with a brief discussion in Section 7. Technical details, proofs, and additional simulation results are given in a supplement.

To facilitate our discussion, we introduce the following notations that will be used throughout the article. Let $\mathbb{R}^{r\times u}$ be the set of all $r \times u$ matrices and let $\mathbb{S}^{m\times m}$ be the set of all $m \times m$ real and symmetric matrices. For any $\mathbf{A} \in \mathbb{R}^{r\times u}(u \leq r)$, Span($\mathbf{A}$) is the subspace of $\mathbb{R}^r$ spanned by the columns of $\mathbf{A}$. Let $\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^\dagger\mathbf{A}^T$ be the projection onto Span($\mathbf{A}$), and let $\mathbf{Q}_{\mathbf{A}} = \mathbf{I}_r - \mathbf{P}_{\mathbf{A}}$ be the projection onto Span$(\mathbf{A})^\perp$, the orthogonal complement of Span($\mathbf{A}$), where $\dagger$ denotes the Moore-Penrose inverse and $\mathbf{I}_r$ is the identity matrix of dimension r. Note that $\mathbf{P}_{\mathbf{A}}$ and $\mathbf{Q}_{\mathbf{A}}$ can be equivalently denoted by $\mathbf{P}_{\mathcal{A}}$ and $\mathbf{Q}_{\mathcal{A}}$, where $\mathcal{A} =$ Span($\mathbf{A}$). Let 'vec' denote the vectorization operator that stacks the columns of an argument matrix. Let 'vech' represent the half-vectorization operator that vectorizes only the lower triangular of a symmetric matrix. We use $||\cdot||$ to represent Frobenius norm.

2 Envelope quantile regression

2.1 A brief review of quantile regression

Consider a univariate response variable Y and a p-dimensional predictor vector $\mathbf{X} \in \mathbb{R}^p$. Let $F_Y(y|\mathbf{X}=\mathbf{x}) = P(Y \le y|\mathbf{X}=\mathbf{x})$ be the cumulative distribution function (CDF) of Y given $\mathbf{X}=\mathbf{x}$. The τ-th conditional quantile of Y is defined as

$$Q_Y(\tau|\mathbf{X}=\mathbf{x}) = \inf\{y : F_Y(y|\mathbf{X}=\mathbf{x}) \ge \tau\}, \;\; 0<\tau<1.$$

A linear quantile regression model assumes a linear relationship between the τ-th conditional quantile of Y and the predictors, that is,

$$Q_Y(\tau|\mathbf{X}) = \mu_\tau + \boldsymbol{\beta}_\tau^T \mathbf{X}, \tag{2.1}$$

where μ_τ is the intercept and $\boldsymbol{\beta}_\tau \in \mathbb{R}^p$ is the slope vector of the τ-th conditional quantile of $Y|\mathbf{X}$. The primary objective of quantile regression is to estimate $\boldsymbol{\beta}_\tau$, for any $0<\tau<1$, and then to make statistical inference about $\boldsymbol{\beta}_\tau$. The standard method to obtain $\tilde{\boldsymbol{\beta}}_\tau$, the estimator of $\boldsymbol{\beta}_\tau$, is to solve

$$(\tilde{\mu}_\tau, \tilde{\boldsymbol{\beta}}_\tau) = \underset{\mu_\tau\in\mathbb{R}, \boldsymbol{\beta}_\tau\in\mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(Y_i - \mu_\tau - \boldsymbol{\beta}_\tau^T \mathbf{X}_i), \tag{2.2}$$

where $(Y_i, \mathbf{X}_i)$, $i=1,\ldots,n$, is a random sample of $(Y,\mathbf{X})$, and $\rho_\tau(z) = z[\tau - I(z<0)]$ is a piecewise linear loss function. This objective function can be efficiently solved

by linear programming algorithms. Furthermore, the estimator $\tilde{\boldsymbol{\beta}}_\tau$ is $\sqrt{n}$-consistent and asymptotically normal.

Note that the minimizer in (2.2) is also a root of the estimating equations

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{W}_i[I(Y_i < \mu_\tau + \boldsymbol{\beta}_\tau^T\mathbf{X}_i) - \tau] = o_p(n^{-1/2}), \tag{2.3}$$

where $\mathbf{W}_i = (1, \mathbf{X}_i^T)^T$. For a detailed background of quantile regression, we refer to Koenker (2005).

2.2 Envelope quantile regression

We now introduce the EQR approach to distinguish material and immaterial information in terms of modeling the conditional quantiles of the response. The EQR approach builds on the observations that in a variety of applications some part of the predictors are irrelevant to modeling the conditional quantile of the response and do not affect the response through the rest. For example, a disease may be related to a few genetic pathways, while these pathways are uncorrelated with other pathways that are not responsible for the disease.

To mathematically formulate this statement, suppose that for the given quantile level τ of interest, there exists a subspace $\mathcal{S}_\tau = \text{Span}(\boldsymbol{\Gamma}_\tau)$ of $\mathbb{R}^p$, where $\boldsymbol{\Gamma}_\tau \in \mathbb{R}^{p\times d_\tau}(d_\tau \le p)$ is a semi-orthogonal basis of $\mathcal{S}_\tau$, such that

$$\text{i)}\ \ Q_Y(\tau|\mathbf{X}) = Q_Y(\tau|\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X})\ \ \text{and}\ \ \text{ii)}\ \text{Cov}(\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X}, \mathbf{Q}_{\mathcal{S}_\tau}\mathbf{X}) = 0, \tag{2.4}$$

where $\mathbf{P}_{\mathcal{S}_\tau}$ and $\mathbf{Q}_{\mathcal{S}_\tau}$ are projection matrices defined at the end of Section 1.

The first part of (2.4) means that $Q_Y(\tau|\mathbf{X})$ depends on $\mathbf{X}$ only through $\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X}$. Hence $\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X}$ contains full information for modeling the τ-th conditional quantile of Y. The second part indicates that $\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X}$ is uncorrelated with $\mathbf{Q}_{\mathcal{S}_\tau}\mathbf{X}$, which ensures that $\mathbf{Q}_{\mathcal{S}_\tau}\mathbf{X}$ does not provide information about the τ-th conditional quantile of Y through its association with $\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X}$. Thus, $\mathbf{X}$ affects the τ-th conditional quantile of Y only through $\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X}$. We call $\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X}$ the material part of $\mathbf{X}$, and $\mathbf{Q}_{\mathcal{S}_\tau}\mathbf{X}$ the immaterial part of $\mathbf{X}$. Let $\mathbf{\Sigma}_\mathbf{X}$ denote the covariance matrix of $\mathbf{X}$. By Cook, Li, and Chiaromonte (2010), if a subspace $\mathcal{S}_\tau$ is spanned by the eigenvectors of $\mathbf{\Sigma}_\mathbf{X}$ and it contains $\boldsymbol{\beta}_\tau$, then $\mathcal{S}_\tau$ satisfies the conditions in (2.4).

There are applications that would naturally satisfy (2.4). For example, suppose all coordinates of $\mathbf{X}$ are equally correlated such that $\mathbf{\Sigma}_\mathbf{X} = \sigma^2_\mathbf{X}\mathbf{I}_p + r1_p1_p^T$, where r is a constant and 1_p is a p-dimensional vector of 1's, and $\boldsymbol{\beta}_\tau$ has a sparse structure such as $\boldsymbol{\beta}_\tau = (1, 2, 0, \ldots, 0)^T$. Note that the eigenvectors of $\mathbf{\Sigma}_\mathbf{X}$ include 1_p and any vector in $\text{Span}(1_p)^\perp$. Let $v_1 = (r-1, -1, \ldots, -1)^T$ and $v_2 = (-1, r-1, -1, \ldots, -1)^T$. Since $v_1^T 1_p = 0$, $v_2^T 1_p = 0$, v_1 and v_2 are eigenvectors of $\mathbf{\Sigma}_\mathbf{X}$. We can take $\mathcal{S}_\tau = \text{Span}(v_1, v_2, 1_p)$. Because $\boldsymbol{\beta}_\tau \in \mathcal{S}_\tau$, $\mathcal{S}_\tau$ satisfies i) and ii) in (2.4). This example demonstrates that if $\mathbf{X}$ is equally correlated and $\boldsymbol{\beta}_\tau$ is sparse, we can find a subspace that satisfies conditions (2.4). Another example is when $\mathbf{\Sigma}_\mathbf{X}$ has a low rank decomposition. Suppose $\mathbf{\Sigma}_\mathbf{X}$ has the structure $\mathbf{\Sigma}_\mathbf{X} = \mathbf{A}\mathbf{A}^T + c\mathbf{I}_p$, where c is a constant, $\mathbf{A} \in \mathbb{R}^{p\times k}$, and $k < p$. Then $\text{Span}(\mathbf{A})$ is spanned by eigenvectors of $\mathbf{\Sigma}_\mathbf{X}$. Such a low rank covariance structure occurs in a wide range of applications such as factor analysis where the variation of the predictor vector can mostly be explained by a small number of common factors or principal components. If $\boldsymbol{\beta}_\tau$

is contained in Span($\mathbf{A}$), we can take $\mathcal{S}_\tau = \text{Span}(\mathbf{A})$. Thus, $\mathcal{S}_\tau$ satisfies conditions (2.4). If $\boldsymbol{\beta}_\tau$ is not contained in Span($\mathbf{A}$), let $\mathcal{A} = \text{Span}(\mathbf{A})$. Then any vector in the orthogonal complement of $\mathcal{A}$ is an eigenvector of $\boldsymbol{\Sigma}_\mathbf{X}$. We can write $\boldsymbol{\beta}_\tau = \mathbf{P}_\mathcal{A}\boldsymbol{\beta}_\tau + \mathbf{Q}_\mathcal{A}\boldsymbol{\beta}_\tau$, and let $v = \mathbf{Q}_\mathcal{A}\boldsymbol{\beta}_\tau$. Note that v is an eigenvector of $\boldsymbol{\Sigma}_\mathbf{X}$. Let $\mathcal{S}_\tau = \text{Span}(\{\mathbf{A}, v\})$, then $\boldsymbol{\beta}_\tau \in \mathcal{S}_\tau$ and $\mathcal{S}_\tau$ satisfies conditions in (2.4). So for any vector $\boldsymbol{\beta}_\tau$, whether it has a sparsity structure like $\boldsymbol{\beta}_\tau = (*, \cdots, *, 0, \cdots, 0)$ or not, we can find a subspace $\mathcal{S}_\tau$ with dimension at most $k+1$ that satisfies conditions (2.4). Note that we use a sparse $\boldsymbol{\beta}_\tau$ in some examples only for a simple illustration, (2.4) does not require sparsity in $\boldsymbol{\beta}_\tau$.

In fact, the subspace $\mathcal{S}_\tau$ in (2.4) always exists since it can be trivially chosen as the full space $\mathbb{R}^p$. However, the subspace might not be unique and what is of interest is the smallest subspace such that the conditions holds. To address the uniqueness of the material part, we consider the intersection of all such subspaces that satisfy (2.4), which is minimal and well defined. To see so, let's first introduce the definition of a reducing subspace given in Cook, Li, and Chiaromonte (2010).

Definition 1. *A subspace $\mathcal{R} \subseteq \mathbb{R}^p$ is said to be a reducing subspace of $\mathbf{M} \in \mathbb{R}^{p\times p}$ if $\mathcal{R}$ decomposes $\mathbf{M}$ as $\mathbf{M} = \mathbf{P}_\mathcal{R}\mathbf{M}\mathbf{P}_\mathcal{R} + \mathbf{Q}_\mathcal{R}\mathbf{M}\mathbf{Q}_\mathcal{R}$.*

This definition is commonly used in the literature of invariance subspace and functional analysis (Conway, 1990). Lemma 1 connects our formulation to reducing subspaces.

Lemma 1. *Under model (2.1), (i) $Q_Y(\tau|\mathbf{X}) = Q_Y(\tau|\boldsymbol{P}_{\mathcal{S}_\tau}\boldsymbol{X})$ if and only if $\boldsymbol{\beta}_\tau \in \mathcal{S}_\tau$; (ii) $\text{Cov}(\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X}, \mathbf{Q}_{\mathcal{S}_\tau}\mathbf{X}) = 0$ if and only if $\mathcal{S}_\tau$ is a reducing subspace of $\boldsymbol{\Sigma}_\mathbf{X}$.*

For part (i), since $Q_Y(\tau|\mathbf{X}) = \mu_\tau + \boldsymbol{\beta}_\tau^T\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X} + \boldsymbol{\beta}_\tau^T\mathbf{Q}_{\mathcal{S}_\tau}\mathbf{X} = Q_Y(\tau|\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X})$, we have

$\mathbf{Q}_{\mathcal{S}_\tau}\boldsymbol{\beta}_\tau = 0$, and therefore $\boldsymbol{\beta}_\tau \in \mathcal{S}_\tau$. For the other direction, if $\boldsymbol{\beta}_\tau \in \mathcal{S}_\tau$, $Q_Y(\tau|\mathbf{X}) = \mu_\tau + \boldsymbol{\beta}_\tau^T\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X}$. Then $Q_Y(\tau|\mathbf{X}) = Q_Y(\tau|\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X})$. Part (ii) holds as it can be shown that $\boldsymbol{\Sigma}_\mathbf{X} = \mathbf{P}_{\mathcal{S}_\tau}\boldsymbol{\Sigma}_\mathbf{X}\mathbf{P}_{\mathcal{S}_\tau} + \mathbf{Q}_{\mathcal{S}_\tau}\boldsymbol{\Sigma}_\mathbf{X}\mathbf{Q}_{\mathcal{S}_\tau}$ when $\mathrm{Cov}(\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X}, \mathbf{Q}_{\mathcal{S}_\tau}\mathbf{X}) = 0$.

Therefore, based on Lemma 1, (2.4) holds if and only if $\mathcal{S}_\tau$ is a reducing subspace of $\boldsymbol{\Sigma}_\mathbf{X}$ that contains $\boldsymbol{\beta}_\tau$. Such a reducing subspace might not be unique. However, by the property of reducing subspaces, the intersection of all reducing subspaces that contain $\boldsymbol{\beta}_\tau$ is also a reducing subspace containing $\boldsymbol{\beta}_\tau$, and it is unique and minimal. Thus, to achieve maximum reduction and efficiency gains, this smallest reducing subspace that contains $\boldsymbol{\beta}_\tau$ is of interest. We call it the $\boldsymbol{\Sigma}_\mathbf{X}$-*envelope of* $\boldsymbol{\beta}_\tau$, and denote it as $\mathcal{E}_{\boldsymbol{\Sigma}_\mathbf{X}}(\boldsymbol{\beta}_\tau)$, or $\mathcal{E}_\tau$.

To establish the EQR model, let $\boldsymbol{\Phi}_\tau \in \mathbb{R}^{p\times u_\tau}$ $(u_\tau \leq p)$ be a semi-orthogonal basis of $\mathcal{E}_\tau$ and $\boldsymbol{\Phi}_{0\tau} \in \mathbb{R}^{p\times(p-u_\tau)}$ be a semi-orthogonal basis of $\mathcal{E}_\tau^\perp$, the orthogonal subspace of $\mathcal{E}_\tau$. We first assume that the envelope dimension u_τ is known. The determination of envelope dimension will be discussed in Section 4. Since $\boldsymbol{\beta}_\tau \in \mathcal{E}_\tau$, we can write $\boldsymbol{\beta}_\tau$ in a coordinate form as $\boldsymbol{\beta}_\tau = \boldsymbol{\Phi}_\tau\boldsymbol{\eta}_\tau$, where $\boldsymbol{\eta}_\tau$ is the coordinate of $\boldsymbol{\beta}_\tau$ relative to the basis $\boldsymbol{\Phi}_\tau$. In addition, because $\mathcal{E}_\tau$ is a reducing subspace of $\boldsymbol{\Sigma}_\mathbf{X}$, $\boldsymbol{\Sigma}_\mathbf{X}$ can be decomposed into two orthogonal parts: $\boldsymbol{\Sigma}_\mathbf{X} = \mathbf{P}_{\mathcal{E}_\tau}\boldsymbol{\Sigma}_\mathbf{X}\mathbf{P}_{\mathcal{E}_\tau} + \mathbf{Q}_{\mathcal{E}_\tau}\boldsymbol{\Sigma}_\mathbf{X}\mathbf{Q}_{\mathcal{E}_\tau}$. Therefore, model (2.1) can be reparameterized into an envelope structure:

$$\begin{aligned} Q_Y(\tau|\mathbf{X}) &= \mu_\tau + \boldsymbol{\eta}_\tau^T\boldsymbol{\Phi}_\tau^T\mathbf{X} \\ \boldsymbol{\Sigma}_\mathbf{X} &= \boldsymbol{\Phi}_\tau\boldsymbol{\Omega}_\tau\boldsymbol{\Phi}_\tau^T + \boldsymbol{\Phi}_{0\tau}\boldsymbol{\Omega}_{0\tau}\boldsymbol{\Phi}_{0\tau}^T, \end{aligned} \tag{2.5}$$

where $\boldsymbol{\Omega}_\tau \in \mathbb{R}^{u_\tau\times u_\tau}$ and $\boldsymbol{\Omega}_{0\tau} \in \mathbb{R}^{(p-u_\tau)\times(p-u_\tau)}$ are positive definite matrices that serve as coordinates of $\mathbf{P}_{\mathcal{E}_\tau}\boldsymbol{\Sigma}_\mathbf{X}\mathbf{P}_{\mathcal{E}_\tau}$ and $\mathbf{Q}_{\mathcal{E}_\tau}\boldsymbol{\Sigma}_\mathbf{X}\mathbf{Q}_{\mathcal{E}_\tau}$ relative to the bases $\boldsymbol{\Phi}_\tau$ and $\boldsymbol{\Phi}_{0\tau}$, respectively. We call this model the *envelope quantile regression* (EQR) model.

By incorporating the idea of enveloping into the formulation of quantile regression, the new EQR model utilizes underlying information in both the predictors and response to identify the material and immaterial information, and connect the parameter of interest, $\boldsymbol{\beta}_\tau$, only to the material part. This leads to efficiency gains in parameter estimation. For a simple illustration, suppose that the envelope basis $\boldsymbol{\Phi}_\tau$ is known and $\mathrm{E}(\mathbf{X}) = 0$. Let $\tilde{\boldsymbol{\beta}}_\tau$ be the standard estimator of $\boldsymbol{\beta}_\tau$ from (2.1). Then the asymptotic variance of $\tilde{\boldsymbol{\beta}}_\tau$, denoted as $\mathrm{avar}(\sqrt{n}\tilde{\boldsymbol{\beta}}_\tau)$, is $\omega^2\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}$ under an i.i.d. error model (Koenker, 2005), where ω is a constant. Since $\boldsymbol{\Phi}_\tau$ is known, the envelope estimator of $\boldsymbol{\beta}_\tau$ is then $\widehat{\boldsymbol{\beta}}_\tau = \boldsymbol{\Phi}_\tau\widehat{\boldsymbol{\eta}}_\tau$ and

$$\mathrm{avar}(\sqrt{n}\widehat{\boldsymbol{\beta}}_\tau) = \boldsymbol{\Phi}_\tau \mathrm{avar}(\sqrt{n}\widehat{\boldsymbol{\eta}}_\tau)\boldsymbol{\Phi}_\tau = \omega^2\boldsymbol{\Phi}_\tau\big[\mathrm{Var}(\boldsymbol{\Phi}_\tau^T\mathbf{X})\big]^{-1}\boldsymbol{\Phi}_\tau = \omega^2\boldsymbol{\Phi}_\tau\boldsymbol{\Omega}_\tau^{-1}\boldsymbol{\Phi}_\tau^T.$$

Thus, $\mathrm{avar}(\sqrt{n}\tilde{\boldsymbol{\beta}}_\tau) - \mathrm{avar}(\sqrt{n}\widehat{\boldsymbol{\beta}}_\tau) = \omega^2\boldsymbol{\Sigma}_{\mathbf{X}}^{-1} - \omega^2\boldsymbol{\Phi}_\tau\boldsymbol{\Omega}_\tau^{-1}\boldsymbol{\Phi}_\tau^T = \omega^2\boldsymbol{\Phi}_{0\tau}\boldsymbol{\Omega}_{0\tau}^{-1}\boldsymbol{\Phi}_{0\tau}^T \geq 0$, where the last equation holds because $\boldsymbol{\Sigma}_{\mathbf{X}}^{-1} = \boldsymbol{\Phi}_\tau\boldsymbol{\Omega}_\tau^{-1}\boldsymbol{\Phi}_\tau^T + \boldsymbol{\Phi}_{0\tau}\boldsymbol{\Omega}_{0\tau}^{-1}\boldsymbol{\Phi}_{0\tau}^T$. Therefore, the envelope estimator is asymptotically more efficient than (or at least as efficient as) the standard quantile estimator, and efficiency gains can be quite substantial when the immaterial variation $\boldsymbol{\Phi}_{0\tau}\boldsymbol{\Omega}_{0\tau}^{-1}\boldsymbol{\Phi}_{0\tau}^T$ of $\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}$ is relatively large. In Section 3, we will give more rigorous justification of the asymptotic efficiency of EQR estimators under general settings, while the estimation algorithm will be presented in Section 4.

In addition, the total number of free parameters in $\boldsymbol{\beta}_\tau$ and $\boldsymbol{\Sigma}_{\mathbf{X}}$ under the EQR model is $u_\tau + p(p+1)/2$, where u_τ for $\boldsymbol{\eta}_\tau$, $u_\tau(p-u_\tau)$ for $\mathrm{Span}(\boldsymbol{\Phi}_\tau)$, $u_\tau(u_\tau+1)/2$ for $\boldsymbol{\Omega}_\tau$, and $(p-u_\tau)(p-u_\tau+1)/2$ for $\boldsymbol{\Omega}_{0\tau}$; while without enveloping, the number of free parameters in $\boldsymbol{\beta}_\tau$ and $\boldsymbol{\Sigma}_{\mathbf{X}}$ is $p+p(p+1)/2$. The EQR model reduces the number of parameters by $p-u_\tau$.

3 Theoretical results

Consider the quantile regression model (2.1) with an arbitrary quantile level of interest τ. Denote the conditional density function of $Y|\mathbf{X}$ as $f_{Y|\mathbf{X}}$. Denote the asymptotic variance of a general statistic $\mathbf{M}_n$ as $\text{avar}(\sqrt{n}\mathbf{M}_n)$. Let $\boldsymbol{\theta} := (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T \in \mathbb{R}^{2p+1+s}$, where $\boldsymbol{\theta}_1 = (\mu_\tau, \boldsymbol{\beta}_\tau^T)^T$ represents the parameters in the conditional quantile regression, $\boldsymbol{\theta}_2 = (\text{vech}(\boldsymbol{\Sigma}_\mathbf{X})^T, \boldsymbol{\mu}_\mathbf{X}^T)^T$ contains parameters in the marginal distribution of $\mathbf{X}$, and $s = p(p+1)/2$ is the dimension of $\text{vech}(\boldsymbol{\Sigma}_\mathbf{X})$. Let $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^T, \text{vech}(\boldsymbol{\Sigma}_\mathbf{X})^T)^T$ be a collection of parameters that are directly related to the envelope model (2.5). Here $\boldsymbol{\theta}$, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}^*$ are all relevant to τ but for convenience, we omit the subscript τ for these notations.

We first consider unitizing estimating equations for the estimation of the unknown parameter vector $\boldsymbol{\theta}$. We take (2.2), and the first and second order moment conditions of $\mathbf{X}$ to be our estimation equations:

$$h_n(\boldsymbol{\theta}) = \begin{pmatrix} h_{1,n}(\boldsymbol{\theta}_1) \\ h_{2,n}(\boldsymbol{\theta}_2) \\ h_{3,n}(\boldsymbol{\theta}_2) \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \mathbf{W}_i[I(Y_i < \mu_\tau + \boldsymbol{\beta}_\tau^T\mathbf{X}_i) - \tau] \\ \text{vech}(\boldsymbol{\Sigma}_\mathbf{X}) - \text{vech}(\mathbf{S}_\mathbf{X}) \\ \boldsymbol{\mu}_\mathbf{X} - \bar{\mathbf{X}} \end{pmatrix} = \frac{1}{n}\sum_{i=1}^n g(\mathbf{Z}_i;\boldsymbol{\theta}) = o_p(n^{-1/2}), \tag{3.1}$$

where $\mathbf{S}_\mathbf{X} = \frac{1}{n}\sum_{i=1}^n(\mathbf{X}_i - \boldsymbol{\mu}_\mathbf{X})(\mathbf{X}_i - \boldsymbol{\mu}_\mathbf{X})^T$ is the sample covariance matrix of $\mathbf{X}$ given $\boldsymbol{\mu}_\mathbf{X}$, $\mathbf{Z}_i = (Y_i, \mathbf{X}_i^T)^T$, and $g(\mathbf{Z}_i;\boldsymbol{\theta}) = (g_1^T(\mathbf{Z}_i;\boldsymbol{\theta}_1), g_2^T(\mathbf{Z}_i;\boldsymbol{\theta}_2), g_3^T(\mathbf{Z}_i;\boldsymbol{\theta}_2))^T$ with $g_1(\mathbf{Z}_i;\boldsymbol{\theta}_1) = \mathbf{W}_i[I(Y_i < \mu_\tau + \boldsymbol{\beta}_\tau^T\mathbf{X}_i) - \tau]$, $g_2(\mathbf{Z}_i;\boldsymbol{\theta}_2) = \text{vech}(\boldsymbol{\Sigma}_\mathbf{X}) - \text{vech}\{(\mathbf{X}_i - \boldsymbol{\mu}_\mathbf{X})(\mathbf{X}_i - \boldsymbol{\mu}_\mathbf{X})^T\}$, and $g_3(\mathbf{Z}_i;\boldsymbol{\theta}_2) = \boldsymbol{\mu}_\mathbf{X} - \mathbf{X}_i$.

Let $\tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}^*$ denote the standard estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ from solving the estimating

equation (3.1) without enveloping, and let $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_0^*$ be the true values of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$, respectively. Since the parameters of main interest are $\boldsymbol{\beta}_\tau$ and $\boldsymbol{\Sigma}_{\mathbf{X}}$ in EQR model, and in addition the estimator of $\boldsymbol{\mu}_{\mathbf{X}}$ is $\bar{\mathbf{X}}$, which remains unchanged under enveloping and has the same asymptotic distribution in both the envelope and non-envelope settings, we neglect $\boldsymbol{\mu}_{\mathbf{X}}$ in the following theoretical development.

To investigate the asymptotic behavior of the estimator of $\boldsymbol{\theta}^*$, we require the following regularity conditions.

(C1) For any $\mathbf{x}$ in the support of $\mathbf{X}$, the conditional distribution of $\mathbf{Y}|\mathbf{X}=\mathbf{x}$ is absolutely continuous, with the continuous density $f_{\mathbf{Y}|\mathbf{X}}$ uniformly bounded away from 0 and ∞ at $\xi_0(\tau|\mathbf{x})$, the τ-conditional quantile of $Y|\mathbf{X}=\mathbf{x}$ under $\boldsymbol{\theta}_0$.

(C2) The expectation $\mathrm{E}_{\boldsymbol{\theta}_0}[g(\mathbf{Z};\boldsymbol{\theta})]$ is twice differentiable at $\boldsymbol{\theta}_0$ with $\frac{\partial \mathrm{E}_{\boldsymbol{\theta}_0}[g(\mathbf{Z};\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}^T}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ having full rank and a finite Frobenius norm. The matrix $\mathrm{E}_{\boldsymbol{\theta}_0}[g(\mathbf{Z};\boldsymbol{\theta}_0)g^T(\mathbf{Z};\boldsymbol{\theta}_0)]$ is positive definite and has a finite Frobenius norm, and in addition, the array $\frac{\partial \mathrm{E}_{\boldsymbol{\theta}_0}[g(\mathbf{Z};\boldsymbol{\theta})g^T(\mathbf{Z};\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}^T}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ has a finite Frobenius norm.

(C3) $\mathrm{E}||\mathbf{X}||^3$ is bounded. In addition, the support $\boldsymbol{\Theta}$ of $\boldsymbol{\theta}$ is compact and $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$.

Conditions (C1) and (C3) are standard in the literature of quantile regression. Condition (C2) is a regular assumption for estimating equations. The following theorem 1 establishes the asymptotic distribution of the standard estimator $\tilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}^*$.

Theorem 1. *Under the regularity conditions (C1)-(C3),* $\sqrt{n}(\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0^*)$ *converges in distribution to a multivariate normal distribution with mean zero and covariance matrix* $\mathrm{avar}(\sqrt{n}\tilde{\boldsymbol{\theta}}^*) =$

$\mathbf{U}^{-1}\mathbf{V}\mathbf{U}^{-1}$, *where*

$$\mathbf{U} = \begin{pmatrix} \mathrm{E}_{\boldsymbol{\theta}_0}[f_{Y|\mathbf{X}}(\xi_0(\tau|\mathbf{X}))\mathbf{W}\mathbf{W}^T] & 0 \\ 0 & \mathbf{I}_s \end{pmatrix}$$

and

$$\mathbf{V} = \begin{pmatrix} \tau(1-\tau)\mathrm{E}_{\boldsymbol{\theta}_0}[\mathbf{W}\mathbf{W}^T] & 0 \\ 0 & \mathrm{var}_{\boldsymbol{\theta}_0}\{\mathrm{vech}[(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X},0})(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X},0})^T]\} \end{pmatrix},$$

with $\boldsymbol{\mu}_{\mathbf{X},0}$ *being the true value of* $\boldsymbol{\mu}_{\mathbf{X}}$.

The proof of Theorems 1 is given in Section A of the supplement. Theorem 1 shows that the standard estimator $\tilde{\boldsymbol{\theta}}_1 = (\tilde{\mu}_\tau, \tilde{\boldsymbol{\beta}}_\tau^T)^T$ for the conditional quantile regression from solving $h_n(\boldsymbol{\theta}) = 0$ is asymptotically independent of the standard estimator $\mathrm{vech}(\tilde{\boldsymbol{\Sigma}}_{\mathbf{X}})$ for the marginal distribution of $\mathbf{X}$. In addition, let $\tilde{\boldsymbol{\theta}}_{1,m}$ denote the estimator of $\boldsymbol{\theta}_1$ obtained directly by minimizing (2.1). It follows from the results in Knight (1998) and Koenker (2005) that $\tilde{\boldsymbol{\theta}}_1$ is asymptotically equivalent to $\tilde{\boldsymbol{\theta}}_{1,m}$.

Under the envelope setting, we denote the parameters in the coordinate representation of the EQR model (2.5) into the vector

$$\boldsymbol{\zeta}_\tau = \left(\mu_\tau, \boldsymbol{\eta}_\tau^T, \mathrm{vec}(\boldsymbol{\Phi}_\tau)^T, \mathrm{vech}(\boldsymbol{\Omega}_\tau)^T, \mathrm{vech}(\boldsymbol{\Omega}_{0\tau})^T\right)^T = \left(\boldsymbol{\zeta}_{\tau,1}, \boldsymbol{\zeta}_{\tau,2}^T, \boldsymbol{\zeta}_{\tau,3}^T, \boldsymbol{\zeta}_{\tau,4}^T, \boldsymbol{\zeta}_{\tau,5}^T\right)^T,$$

and define the parameter of interest $\boldsymbol{\theta}^*$ as

$$\boldsymbol{\theta}^* = \begin{pmatrix} \mu_\tau \\ \boldsymbol{\beta}_\tau \\ \text{vech}(\boldsymbol{\Sigma}_{\mathbf{X}}) \end{pmatrix} = \begin{pmatrix} \mu_\tau \\ \boldsymbol{\Phi}_\tau \boldsymbol{\eta}_\tau \\ \text{vech}(\boldsymbol{\Phi}_\tau \boldsymbol{\Omega}_\tau \boldsymbol{\Phi}_\tau^T + \boldsymbol{\Phi}_{0\tau} \boldsymbol{\Omega}_{0\tau} \boldsymbol{\Phi}_{0\tau}^T) \end{pmatrix} := \begin{pmatrix} \psi_1(\boldsymbol{\zeta}_\tau) \\ \psi_2(\boldsymbol{\zeta}_\tau) \\ \psi_3(\boldsymbol{\zeta}_\tau) \end{pmatrix} = \psi(\boldsymbol{\zeta}_\tau). \quad (3.2)$$

Note that under enveloping, the estimating equations in (3.1) are reparameterized as:

$$h_n(\boldsymbol{\theta}) = \begin{pmatrix} h_{1,n}(\boldsymbol{\theta}_1) \\ h_{2,n}(\boldsymbol{\theta}_2) \\ h_{3,n}(\boldsymbol{\theta}_2) \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \mathbf{W}_i[I(Y_i < \mu_\tau + \boldsymbol{\eta}_\tau^T \boldsymbol{\Phi}_\tau^T \mathbf{X}_i) - \tau] \\ \text{vech}(\boldsymbol{\Phi}_\tau \boldsymbol{\Omega}_\tau \boldsymbol{\Phi}_\tau^T + \boldsymbol{\Phi}_{0\tau} \boldsymbol{\Omega}_{0\tau} \boldsymbol{\Phi}_{0\tau}^T) - \text{vech}(\mathbf{S}_{\mathbf{X}}) \\ \boldsymbol{\mu}_{\mathbf{X}} - \bar{\mathbf{X}} \end{pmatrix}. \quad (3.3)$$

The number of equations in (3.3) is $1 + 2p + p(p+1)/2$. It is greater than the number of free parameters in μ_τ, $\boldsymbol{\beta}_\tau$, $\mu_{\mathbf{X}}$ and $\boldsymbol{\Sigma}_{\mathbf{X}}$ under the envelope parameterization, which is $1 + u_\tau + p + p(p+1)/2$. Therefore it cannot be guaranteed that all equations can be made to zero simultaneously. Hence the solution of (3.3) may not exist. Instead, we propose to estimate the parameters by utilizing the idea of generalized method of moments (GMM; Hansen, 1982) for the parsimonious envelope model. Let $\boldsymbol{\zeta}_\tau^{'} = (\boldsymbol{\zeta}_\tau^T, \boldsymbol{\mu}_{\mathbf{X}}^T)^T$ and $\psi_0(\boldsymbol{\zeta}_\tau^{'}) := (\psi^T(\boldsymbol{\zeta}_\tau), \boldsymbol{\mu}_{\mathbf{X}}^T)^T = \boldsymbol{\theta}$. The *envelope GMM estimator* $\widehat{\boldsymbol{\theta}}_g$ of $\boldsymbol{\theta}$ is then defined as

$$\widehat{\boldsymbol{\theta}}_g = \underset{\boldsymbol{\theta}:\boldsymbol{\theta}=\psi_0(\boldsymbol{\zeta}_\tau^{'})}{\text{argmin}} \; h_n^T(\boldsymbol{\theta}) \widehat{\boldsymbol{\Delta}} h_n(\boldsymbol{\theta}), \quad (3.4)$$

where $\widehat{\boldsymbol{\Delta}}$ is chosen to be any $\sqrt{n}$-consistent estimator of $\left\{\text{E}_{\boldsymbol{\theta}_0}[g(\mathbf{Z};\boldsymbol{\theta}_0) g^T(\mathbf{Z};\boldsymbol{\theta}_0)]\right\}^{-1}$, for example, $\widehat{\boldsymbol{\Delta}} = \left\{n^{-1}\sum_{i=1}^n g(\mathbf{Z}_i;\tilde{\boldsymbol{\theta}}) g^T(\mathbf{Z}_i;\tilde{\boldsymbol{\theta}})\right\}^{-1}$. In Section 4, we will propose an estimation

procedure to attain the envelope GMM estimator $\widehat{\boldsymbol{\theta}}_g$.

Let $\widehat{\boldsymbol{\theta}}_g^*$ denote the envelope GMM estimator of $\boldsymbol{\theta}^*$, the parameter of interest. We next establish asymptotic theory for $\widehat{\boldsymbol{\theta}}_g^*$ and compare it with the standard estimator $\tilde{\boldsymbol{\theta}}^*$.

Theorem 2. *(1) Under the regularity conditions (C1)-(C3), assume that the support of the envelope parameter vector $\boldsymbol{\zeta}_\tau$ is compact, then $\sqrt{n}(\widehat{\boldsymbol{\theta}}_g^* - \boldsymbol{\theta}_0^*)$ converges in distribution to a multivariate normal distribution with mean zero and covariance matrix*

$$\text{avar}(\sqrt{n}\widehat{\boldsymbol{\theta}}_g^*) = \boldsymbol{\Psi}(\boldsymbol{\Psi}^T\mathbf{U}\mathbf{V}^{-1}\mathbf{U}\boldsymbol{\Psi})^\dagger\boldsymbol{\Psi}^T,$$

where $\boldsymbol{\Psi} = \partial\psi(\boldsymbol{\zeta}_\tau)/\partial\boldsymbol{\zeta}_\tau^T$ is the gradient matrix of $\psi(\boldsymbol{\zeta}_\tau)$ relative to $\boldsymbol{\zeta}_\tau$. Its explicit expression is given in the supplement (B.6).

(2) In addition, $\text{avar}(\sqrt{n}\widehat{\boldsymbol{\theta}}_g^*) \leq \text{avar}(\sqrt{n}\tilde{\boldsymbol{\theta}}^*)$.

The proof of Theorem 2 is given in Section B of the supplement. The main challenge of the proof lies in the fact that objective function in (3.4) is not only nonsmooth but also over-parameterized. We employ empirical process techniques (Van Der Vaart and Wellner, 1996; Van der Vaart, 1998) and the results in Newey and McFadden (1994) and Shapiro (1986) for the derivation. The theorem shows asymptotic normality for the envelope GMM estimator $\widehat{\boldsymbol{\theta}}_g^*$ of the joint parameters in the quantile regression and the covariance matrix of $\mathbf{X}$. More importantly, it establishes the asymptotic efficiency of $\widehat{\boldsymbol{\theta}}_g^*$ relative to the standard estimator $\tilde{\boldsymbol{\theta}}^*$. Thus, by utilizing information in both the predictors and response, the new EQR approach can lead to gains in efficiency in the quantile regression estimation.

To illustrate the efficiency gains, we consider a special case of i.i.d error models and

assume that $\mathbf{X}$ is multivariate normal and $\mathrm{E}(\mathbf{X}) = 0$. After some simplification of the form of the asymptotic variance given in Theorem 1 and Theorem 2 (see Section B of the supplement), we have

$$\mathrm{avar}(\sqrt{n}\tilde{\boldsymbol{\beta}}_\tau) = \frac{\tau(1-\tau)}{f^2(\xi(\tau))}\boldsymbol{\Sigma}_{\mathbf{X}}^{-1},$$

and

$$\mathrm{avar}(\sqrt{n}\widehat{\boldsymbol{\beta}}_{g,\tau}) = \frac{\tau(1-\tau)}{f^2(\xi(\tau))}\boldsymbol{\Phi}_\tau\boldsymbol{\Omega}_\tau^{-1}\boldsymbol{\Phi}_\tau^T + (\boldsymbol{\eta}_\tau^T \otimes \boldsymbol{\Phi}_{0\tau})\mathbf{T}^{-1}(\boldsymbol{\eta}_\tau \otimes \boldsymbol{\Phi}_{0\tau}^T)$$

where

$$\mathbf{T} = \frac{f^2(\xi(\tau))}{\tau(1-\tau)}(\boldsymbol{\eta}_\tau\boldsymbol{\eta}_\tau^T) \otimes \boldsymbol{\Omega}_{0\tau} + \boldsymbol{\Omega}_\tau \otimes \boldsymbol{\Omega}_{0\tau}^{-1} + \boldsymbol{\Omega}_\tau^{-1} \otimes \boldsymbol{\Omega}_{0\tau} - 2\mathbf{I}_{u_\tau} \otimes \mathbf{I}_{p-u_\tau}.$$

Compared to the simple illustration example given in Section 2.2, the asymptotic variance of $\widehat{\boldsymbol{\beta}}_{g,\tau}$ has an additional term $(\boldsymbol{\eta}_\tau^T \otimes \boldsymbol{\Phi}_{0\tau})\mathbf{T}^{-1}(\boldsymbol{\eta}_\tau \otimes \boldsymbol{\Phi}_{0\tau}^T)$, which can be considered as the cost for estimating the envelope as it is unknown in general. Theorem 2 shows that even with this estimation cost, the envelope GMM estimator $\widehat{\boldsymbol{\beta}}_{g,\tau}$ is still asymptotically more efficient than (or at least as asymptotically efficient as) the standard estimator $\tilde{\boldsymbol{\beta}}_\tau$.

Statistical inference for the envelope GMM estimator can be performed based on the asymptotic distribution in Theorem 2. One can estimate the asymptotic variance with $\widehat{\mathrm{avar}}(\sqrt{n}\widehat{\boldsymbol{\theta}}_g^*) = \hat{\boldsymbol{\Psi}}(\hat{\boldsymbol{\Psi}}^T\hat{\mathbf{U}}\hat{\mathbf{V}}^{-1}\hat{\mathbf{U}}\hat{\boldsymbol{\Psi}})^\dagger\hat{\boldsymbol{\Psi}}^T$, where $\hat{\boldsymbol{\Psi}}$, $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ are consistent estimators of $\boldsymbol{\Psi}$, $\mathbf{U}$ and $\mathbf{V}$, respectively. Correspondingly, $\widehat{\mathrm{avar}}(\sqrt{n}\widehat{\boldsymbol{\theta}}_g^*) \to \mathrm{avar}(\sqrt{n}\widehat{\boldsymbol{\theta}}_g^*)$ in probability. Then statistical inference can be made based on asymptotic normality. The consistent estimator of $\boldsymbol{\Psi}$ can be easily obtained using the estimated envelope parameters, and the consistent estimator of $\mathbf{V}$ can be provided by moment estimation. The estimation of $\mathbf{U}$ is not straightforward

as it involves unknown density function, which is a problem that also occurs in standard quantile regression inference. We can adopt the kernel-based estimation approach (Powell, 1991; Koenker, 2005) to achieve consistent estimation of $\mathbf{U}_{(1)} = \mathrm{E}_{\boldsymbol{\theta}_0}[f_{Y|\mathbf{X}}(\xi_0(\tau|\mathbf{X}))\mathbf{W}\mathbf{W}^T]$ with $\hat{\mathbf{U}}_{(1)} = (nh_n)^{-1}\sum_{i=1}^{n} K(\hat{\xi}_i(\tau|\mathbf{X})/h_n)\mathbf{W}_i\mathbf{W}_i^T$ under certain Lipschitz continuity conditions on f, where $\hat{\xi}_i(\tau|\mathbf{X}) = Y_i - \hat{\mu}_{g,\tau} - \hat{\boldsymbol{\beta}}_{g,\tau}^T\mathbf{X}_i$, and $K(\cdot)$ and h_n are kernel function and bandwidth satisfying $h_n \to 0$ and $\sqrt{n}h_n \to \infty$. For example, Powell (1991) used the kernel $K(\hat{\xi}_i(\tau|\mathbf{X})) = I(|\hat{\xi}_i(\tau|\mathbf{X})| < h_n)/2$. One might refer to Powell (1991) and Koenker (2005) for more details on the choices of $K(\cdot)$ and h_n.

On the other hand, bootstrap is a useful alternative for the inference of the EQR estimator and is widely used in standard quantile regression (QR) inference (Knight, 1999; Koenker, 2005; Wang and Wang, 2009; Feng *et al.*, 2011, among many others). For example, one can apply paired bootstrap or wild bootstrap to make inference for the EQR estimator under heteroscedastic errors. These methods have been shown to achieve consistency in QR inference (Knight, 1999; Feng *et al.*, 2011). We applied paired bootstrap in our numerical studies (see Figures 1 and 2). It performs fairly well and shows accurate estimation of the standard deviations as compared to those obtained from repeated samples.

For statistical inference, the EQR estimator might lose some efficiency compared to its theoretical asymptotic variance due to the estimation uncertainty of the unknown parameters. In this circumstance, the performance of the EQR estimator might fall into either one of the following two scenarios. First, when the immaterial variation of the data is substantial, even if the envelope dimension is relatively large (e.g. close to the full dimension), EQR could still outperform the standard QR. In this case, the efficiency gains from iden-

tifying and removing immaterial information could overcome the estimation uncertainty, leading to more efficient estimators and smaller mean squared errors (MSE). On the other hand, when the immaterial variation is relatively small while the envelope dimension is large, the efficiency gains from enveloping might be inadequate to overcome the cost of uncertainty in estimating the envelope subspace and parameters. In this case, the estimation uncertainty (including both the estimation of the envelope dimension and envelope parameters) could counteract and surpass the efficiency gains, resulting in relatively close or worse performance of the EQR estimator compared to the QR estimator. Simulation studies illustrating these two cases are provided in Section D.1 of the supplement.

If parameters in (2.1) do not have the envelope structure, the EQR estimator $\widehat{\boldsymbol{\beta}}_{g,\tau}$ may still have a smaller MSE compared to the standard QR estimator $\tilde{\boldsymbol{\beta}}_{\tau}$ based on the bias-variance tradeoff. To be more specific, although the EQR estimator might be biased, it could have a smaller estimation variance. Then if the reduction of the estimation variance is substantial, the EQR estimator will have a smaller MSE. A simulation is included in Section D.2 of the supplement for illustration.

4 Estimation

In the literature on envelope models, estimation is routinely performed by optimizing a standard objective function, such as the log likelihood function, under the envelope parameterization. These existing envelope estimation techniques usually require the first two derivatives of the objective function (Cook, Li, and Chiaromonte, 2010; Cook and Zhang,

2016; Cook, Forzani, and Su, 2016). However, the objective function for quantile regression (2.2) is non-smooth.

We start with estimating equation (3.3). In (3.3), $\mathbf{\Phi}_\tau$ is not estimable as it can be any orthogonal basis of $\mathcal{E}_{\mathbf{\Sigma_X}}(\boldsymbol{\beta}_\tau)$, and only $\mathcal{E}_{\mathbf{\Sigma_X}}(\boldsymbol{\beta}_\tau) = \text{Span}(\mathbf{\Phi}_\tau)$ is estimable. To obtain an estimator of $\mathcal{E}_{\mathbf{\Sigma_X}}(\boldsymbol{\beta}_\tau)$, we have to perform a Grassmann manifold optimization, which can be slow and difficult in sizable problems. Cook, Forzani, and Su (2016) proposed a reparameterization of $\mathbf{\Phi}_\tau$ such that the Grassmann manifold optimization problem can be converted to an unconstrained matrix optimization problem. It is shown that the computing speed is greatly improved under the new parameterization. Therefore we adopt this reparameterization for our problem and this does not affect our theoretical results. Without loss of generality, we assume that the upper $u_\tau \times u_\tau$ block is invertible. Write

$$\mathbf{\Phi}_\tau = \begin{pmatrix} \mathbf{\Phi}_{\tau 1} \\ \mathbf{\Phi}_{\tau 2} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{u_\tau} \\ \mathbf{\Phi}_{\tau 2}\mathbf{\Phi}_{\tau 1}^{-1} \end{pmatrix} \mathbf{\Phi}_{\tau 1} \equiv \begin{pmatrix} \mathbf{I}_{u_\tau} \\ \mathbf{A} \end{pmatrix} \mathbf{\Phi}_{\tau 1} \equiv \mathbf{\Phi}_\tau^* \mathbf{\Phi}_{\tau 1}. \tag{4.1}$$

Then $\mathcal{E}_{\mathbf{\Sigma_X}}(\boldsymbol{\beta}_\tau)$ and $\mathbf{A}$ have a one-to-one correspondence. To be more specific, for a u_τ-dimensional subspace of $\mathcal{R}^p$, we have a unique representing basis $\mathbf{\Phi}_\tau^*$ whose first u_τ rows form an identity matrix. So if we obtain an estimator of $\mathbf{A}$, say $\widehat{\mathbf{A}}$, we can easily get $\widehat{\mathbf{\Phi}}_\tau^*$ following the structure in (4.1), and $\widehat{\mathcal{E}}_{\mathbf{\Sigma_X}}(\boldsymbol{\beta}_\tau) = \text{Span}(\widehat{\mathbf{\Phi}}_\tau^*)$. Now let $\boldsymbol{\eta}_\tau^* = \mathbf{\Phi}_{\tau 1}\boldsymbol{\eta}_\tau$ and $\mathbf{\Omega}_\tau^* = \mathbf{\Phi}_{\tau 1}\mathbf{\Omega}_\tau\mathbf{\Phi}_{\tau 1}^T$ be the coordinates of $\boldsymbol{\beta}_\tau$ and $\mathbf{\Sigma_X}$ with respect to $\mathbf{\Phi}_\tau^*$. Let $\boldsymbol{\zeta}_\tau^* = \{\mu_\tau, \text{vec}(\boldsymbol{\eta}_\tau^*)^T, \text{vec}(\mathbf{A})^T, \text{vech}(\mathbf{\Omega}_\tau^*)^T, \text{vech}(\mathbf{\Omega}_{0\tau})^T, \text{vec}(\boldsymbol{\mu}_\mathbf{X})^T\}^T$. Under this parameterization,

(3.3) becomes

$$h_n^*(\boldsymbol{\zeta}_\tau^*) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \mathbf{W}_i\{I[Y_i < \mu_\tau + (\boldsymbol{\Phi}_\tau^*\boldsymbol{\eta}_\tau^*)^T\mathbf{X}_i] - \tau\} \\ \frac{1}{n}\sum_{i=1}^n\{\text{vech}(\boldsymbol{\Phi}_\tau^*\boldsymbol{\Omega}_\tau^*\boldsymbol{\Phi}_\tau^{*T} + \boldsymbol{\Phi}_{0\tau}\boldsymbol{\Omega}_{0\tau}\boldsymbol{\Phi}_{0\tau}^T) - \text{vech}[(\mathbf{X}_i - \boldsymbol{\mu}_\mathbf{X})(\mathbf{X}_i - \boldsymbol{\mu}_\mathbf{X})^T]\} \\ \frac{1}{n}\sum_{i=1}^n(\boldsymbol{\mu}_\mathbf{X} - \mathbf{X}_i) \end{pmatrix}$$

$$\equiv \frac{1}{n}\sum_{i=1}^n g^*(\mathbf{Z}_i; \boldsymbol{\zeta}_\tau^*). \tag{4.2}$$

To obtain the GMM estimator, we use the following two-step algorithm:

Step 1. Obtain the estimator of $\boldsymbol{\zeta}_\tau^*$ by minimizing $h_n^*(\boldsymbol{\zeta}_\tau^*)^T h_n^*(\boldsymbol{\zeta}_\tau^*)$, and denote it as $\tilde{\boldsymbol{\zeta}}_\tau^*$.

Step 2. Estimate the optimal weight matrix as

$$\widehat{\boldsymbol{\Delta}}^{-1} = \left[\frac{1}{n}\sum_{i=1}^n g^*(\mathbf{Z}_i; \tilde{\boldsymbol{\zeta}}_\tau^*) g^*(\mathbf{Z}_i; \tilde{\boldsymbol{\zeta}}_\tau^*)^T\right]^{-1},$$

then obtain the GMM estimator $\widehat{\boldsymbol{\zeta}_\tau^*}$ by minimizing the following quadratic form

$$Q_n(\boldsymbol{\zeta}_\tau^*) = h_n^*(\boldsymbol{\zeta}_\tau^*)^T \widehat{\boldsymbol{\Delta}}^{-1} h_n^*(\boldsymbol{\zeta}_\tau^*).$$

Then $\widehat{\mu}_\tau$, $\widehat{\boldsymbol{\beta}}_\tau = \widehat{\boldsymbol{\Phi}}_\tau^*\widehat{\boldsymbol{\eta}}_\tau^*$ and $\widehat{\boldsymbol{\Sigma}}_\mathbf{X} = \widehat{\boldsymbol{\Phi}}_\tau^*\widehat{\boldsymbol{\Omega}}_\tau^*(\widehat{\boldsymbol{\Phi}}_\tau^*)^T + \widehat{\boldsymbol{\Phi}}_{0\tau}\widehat{\boldsymbol{\Omega}}_{0\tau}\widehat{\boldsymbol{\Phi}}_{0\tau}^T$ are the envelope GMM estimators of μ_τ, $\boldsymbol{\beta}_\tau$ and $\boldsymbol{\Sigma}_\mathbf{X}$.

To optimize the discontinuous GMM objective function, we use the function `fminsearch` in the R package `neldermead`. This function does not require the derivative of the objective function, and is also applicable to discontinuous objective functions. It uses Nelder-Mead

method or downhill simplex method (Nelder and Mead, 1965) to find the minima of the objective function. More information on the method can be found in Section E of the supplement. The Nelder-Mead method has also been used for fitting other quantile regression models (e.g. Koenker and Park, 1996; Otsu, 2003; Noufaily and Jones, 2013).

To select the dimension of the envelope $\mathcal{E}_{\mathbf{\Sigma_X}}(\boldsymbol{\beta}_\tau)$, we apply the robust cross-validation approach (RCV) (Oh *et al.*, 2004). More specifically, we randomly divide the data into K folds, use the kth fold for testing and the remaining $K-1$ folds for training. We repeat this for $k = 1, \ldots, K$, and aggregate the prediction error based on the quantile loss function. For a fixed u_τ, the RCV criterion is

$$\text{RCV}(u_\tau) = \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(Y_i - \widehat{\mu}_{\tau,-k(i)} - \widehat{\boldsymbol{\beta}}^T_{\tau,-k(i)}\mathbf{X}_i),$$

where $\widehat{\mu}_{\tau,-k(i)}$ and $\widehat{\boldsymbol{\beta}}_{\tau,-k(i)}$ are computed using the data excluding the kth fold that the ith observation resides. Since cross-validation often overfits, we pick u_τ according to the "one-standard error" rule. That is, we choose the smallest u_τ whose error is no more than one standard error above the minimum value of RCV. In our numerical studies, we found that the performance of RCV is stable, even with small sample size.

5 Simulation and data analysis

In this section, we demonstrate the efficiency gains of the EQR model with a numerical experiment and a real data example. We consider the following simulation setting

$$Y_i = \mu + \boldsymbol{\alpha}^T \mathbf{X}_i + (5 + \boldsymbol{\gamma}^T \mathbf{X}_i)\epsilon_i, \quad \text{for } i = 1, \ldots, n,$$

where $\boldsymbol{\alpha} = \boldsymbol{\Phi}\boldsymbol{\eta}_1$, $\boldsymbol{\gamma} = \boldsymbol{\Phi}\boldsymbol{\eta}_2$, and the error ϵ follows the standard normal distribution with distribution function denoted by F_ϵ. Here $\boldsymbol{\Phi} \in \mathbb{R}^{p\times u}(u < p)$ is a semi-orthogonal matrix. Hence $\mu_\tau = \mu + 5F_\epsilon^{-1}(\tau)$, $\boldsymbol{\beta}_\tau = \boldsymbol{\Phi}(\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 F_\epsilon^{-1}(\tau)) = \boldsymbol{\Phi}\boldsymbol{\eta}_\tau$, $\boldsymbol{\Phi}_\tau = \boldsymbol{\Phi}$ and $u_\pi = u$, for $0 < \tau < 1$. We set $p = 10$, $u = 2$ and varied the sample size n from 50 to 1000. We set $\mathbf{X}$ to follow a multivariate normal distribution with mean 0 and variance having the structure $\boldsymbol{\Phi}\boldsymbol{\Omega}\boldsymbol{\Phi}^T + \boldsymbol{\Phi}_0\boldsymbol{\Omega}_0\boldsymbol{\Phi}_0^T$, where $\boldsymbol{\Phi}_0$ is a completion of $\boldsymbol{\Phi}$, and $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ are the coordinate matrices. We generated $\boldsymbol{\Phi}$ with the first $p/2$ rows to be $(-1/\sqrt{p/2}, 0)$ and the other rows to be $(0, -1/\sqrt{p/2})$. The matrix $\boldsymbol{\Omega}$ was a diagonal matrix with diagonal elements 50 and 100, $\boldsymbol{\Omega}_0$ was an identity matrix, $\boldsymbol{\eta}_1$ was $(-5\sqrt{p/2}, -5\sqrt{p/2})^T$, $\boldsymbol{\eta}_2$ was $(0, -\sqrt{2p}/20)^T$, and μ was 5. Therefore $\boldsymbol{\alpha}$ was a vector of 5 and $\boldsymbol{\gamma}$ was a vector with the first $p/2$ elements to be 0 and the rest to be 0.1. For each sample size, 200 replications were generated. For each replication, we fit the standard QR model (2.1) and the EQR model with $u_\pi = 2$. For each element in $\boldsymbol{\beta}_\tau$, we computed the estimation standard deviation from the 200 estimators. We also generated 200 bootstrap repetitions using paired bootstrap, and computed the bootstrap standard deviation. We considered $\tau = 0.5$ and $\tau = 0.9$. The results for a randomly chosen element in $\boldsymbol{\beta}_\tau$ are summarized in Figures 1 and 2. The EQR model

achieves obvious efficiency gains in this example. We compared the estimation standard deviations of the standard QR estimator and the EQR estimator for each element in $\boldsymbol{\beta}_\tau$. We found that, at sample size 50, the EQR estimator reduced the estimation standard deviation by 57.1% to 65.9% for $\tau = 0.5$. Under the standard QR model, to reduce the standard deviation by 65.9%, we need to increase the sample size by approximately 8.6 times the original sample size. The efficiency gain is more pronounced for $\tau = 0.9$, the EQR estimator reduced the estimation standard deviation by 71.0% to 75.9%. To achieve a reduction of 75.9% in estimation standard deviation, we need to increase the sample size by seventeen times the original sample size under the standard QR model. Figures 1 and 2 also show that the bootstrap standard deviation is a very good approximation to the estimation standard deviation.

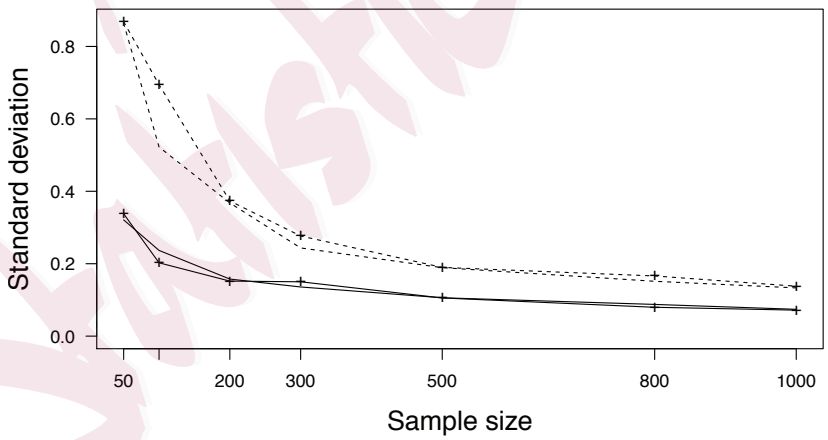

Figure 1: Comparison of the EQR estimator and the standard QR estimator with u_τ fixed at true value ($\tau = 0.5$). Lines — mark the standard deviations of the EQR estimator and lines – – mark the standard deviations of the standard QR estimator. The lines with "+" mark the bootstrap standard deviations for the corresponding estimators.

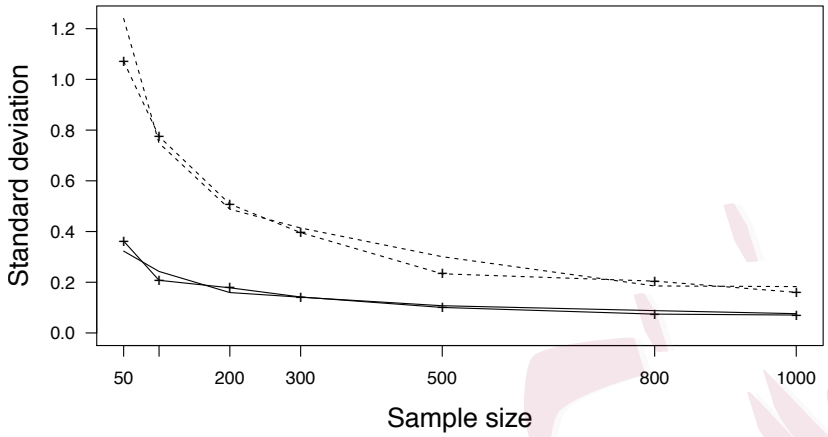

Figure 2: Comparison of the EQR estimator and the standard QR estimator with u_τ fixed at true value ($\tau = 0.9$). Lines — mark the standard deviations of the EQR estimator and lines – – mark the standard deviations of the standard QR estimator. The lines with "+" mark the bootstrap standard deviations for the corresponding estimators.

We also investigated the selection performance of five-fold RCV for each sample size. For different sample sizes, the fraction in 200 replications that RCV selects the true dimension is summarized in Table 1. It is quite stable across all the sample sizes in Table 1. With small sample sizes, when it fails to select the true u_τ, it tends to overestimate and picks a larger dimension than the truth. In that case, we may achieve less efficiency gains, but we do not lose any material information. Therefore we consider the performance of RCV to be reasonable with small sample sizes.

Table 1: The fraction that RCV selects the true dimension

n	50	100	200	500	1000
$\tau = 0.5$	89%	96%	99%	99%	100%
$\tau = 0.9$	93%	97%	96%	99%	100%

Now we compute the estimation standard deviation of the EQR estimator again, but using the selected u_τ instead of the true u_τ. This estimation standard deviation includes the

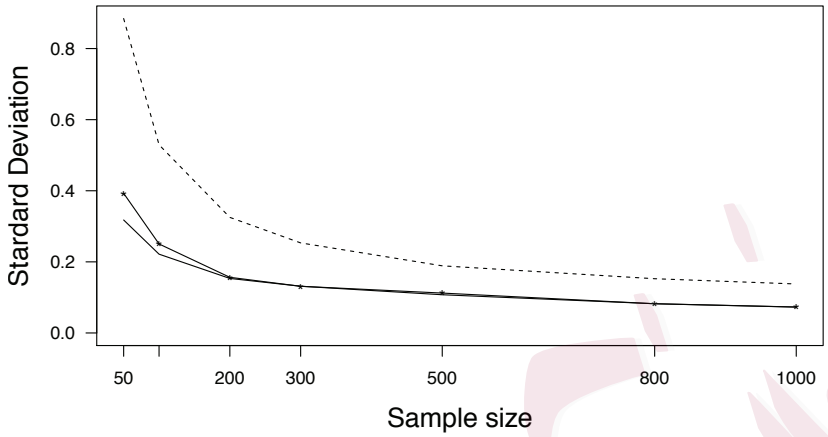

Figure 3: Comparison of the EQR estimators and the standard QR estimator with $\tau = 0.5$. The line — marks the standard deviations of the EQR estimator with true u_τ, the line — with $*$ marks the standard deviations of the EQR estimator with selected u_τ and the line – – marks the standard deviations of the standard QR estimator.

variability in model selection and the variability of the EQR estimator given the selected u_τ. The results are included in Figures 3 and 4. To ease the comparison, we also included the lines for the EQR estimators with u_τ fixed at the true value. At sample size $n = 50$, the EQR estimator with selected u_τ reduces the estimation standard deviation of the QR estimator by 51.7% to 59.3% for $\tau = 0.5$ and by 63.9% to 72.6% for $\tau = 0.9$. Compared with the results with true u_τ, the EQR estimator loses some efficiency gains due to the variability in the selection but the EQR estimator is still more efficient than the standard QR estimator. We also included the MSE for the EQR estimator and the standard QR estimator in Figures 5 and 6. With $n = 50$, the EQR estimator with true u_τ reduces the MSE by 81.2% to 88.0% for $\tau = 0.5$ and by 91.7% to 94.0% for $\tau = 0.9$. The EQR estimator with selected u_τ reduces the MSE by 75.9% to 83.2% for $\tau = 0.5$ and by 87.1% to 92.1% for $\tau = 0.9$. The reduction in MSE is mainly due to the efficiency gains. In this simulation,

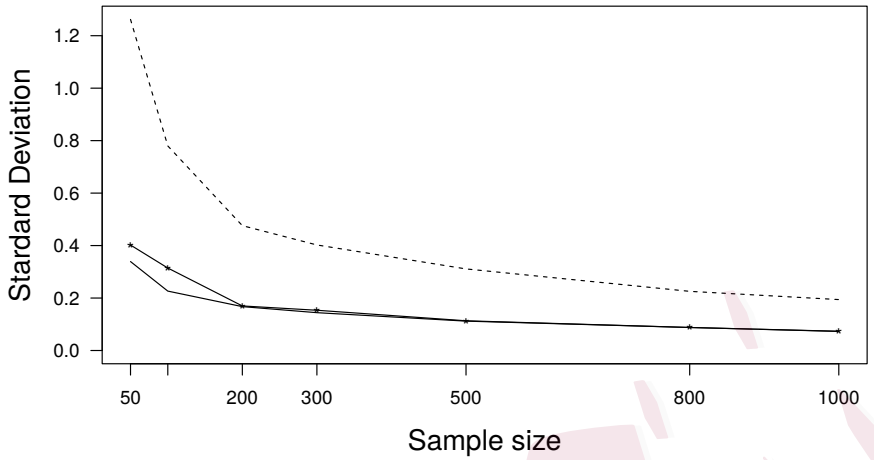

Figure 4: Comparison of the EQR estimators and the standard QR estimator with $\tau = 0.9$. The line — marks the standard deviations of the EQR estimator with true u_τ, the line — with $*$ marks the standard deviations of the EQR estimator with selected u_τ and the line – – marks the standard deviations of the standard QR estimator.

RCV always overestimates u_τ, which loses some efficiency but does not bring in bias. In fact, the squared bias of the EQR estimator is about the same as the QR estimator (see the results in Section D.3 of the supplement).

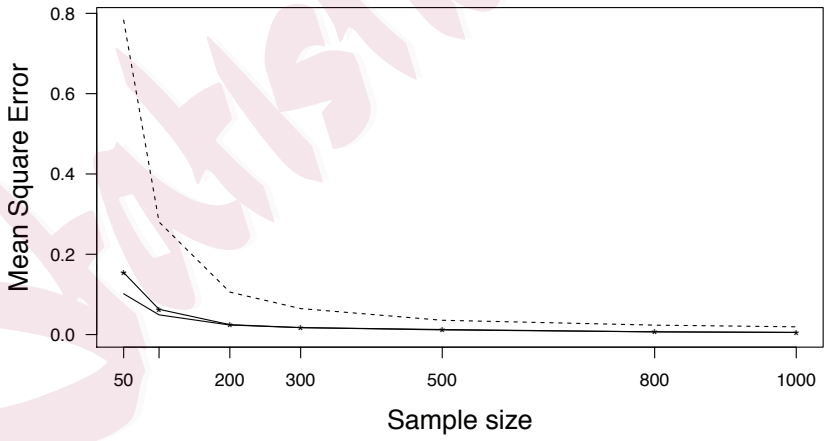

Figure 5: Comparison of the EQR estimators and the standard QR estimator with $\tau = 0.5$. The line — marks the MSE of the EQR estimator with true u_τ, the line — with $*$ marks the MSE of the EQR estimator with selected u_τ and the line – – marks the MSE of the standard QR estimator.

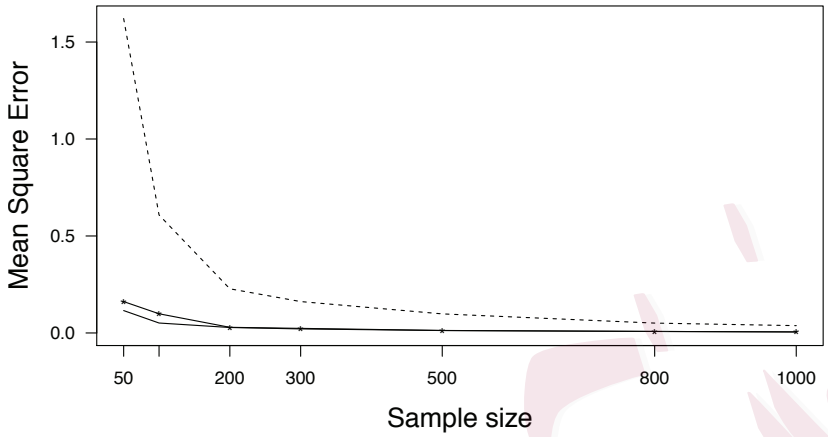

Figure 6: Comparison of the EQR estimators and the standard QR estimator with $\tau = 0.9$. The line — marks the MSE of the EQR estimator with true u_τ, the line — with $*$ marks the MSE of the EQR estimator with selected u_τ and the line – – marks the MSE of the standard QR estimator.

We further examined the EQR model with the baseball salary data (Watnik, 1998). The data contain salaries for 337 non-pitchers for 1992 Major League Baseball season. The histogram of the salaries is right-skewed, which means that some of the players have much higher salaries than the others. The dataset also includes 12 measures of the players' performance in the previous year, including batting average, on-base percentage, number of runs, hits, doubles, triples, home runs, batted in, walks, strike-outs, stolen bases and errors. Each predictor was scaled to have standard deviation 1. We fit the EQR model to the data. RCV suggested $u_\tau = 4$ for $\tau = 0.5$. Across all elements in $\boldsymbol{\beta}_\tau$, the ratios of bootstrap standard deviations of the standard QR estimator versus the EQR estimator range from 0.99 to 6.78 with an average of 2.90. For $\tau = 0.9$, $u_\tau = 2$ was selected by RCV. The ratios of the bootstrap standard deviations range from 1.88 to 29.48 with an average of 8.30. To get an efficient estimator whose estimation standard deviation is 1/8.3

of the original standard deviation under the standard QR, we need to increase the sample to $8.30^2 \approx 70$ times the original sample size. The efficiency gains of the EQR model is massive in this example.

6 Partial envelope quantile regression model

Partial envelope quantile regression model is motived by applications where some predictors are categorical. For example, in medical studies, gender and race are often measured as covariates along with continuous variables such as gene expression intensities to study causes of a certain disease. If categorical predictors are present, the EQR model cannot be applied directly. To resolve this issue, we propose to envelop on the continuous predictors and leave the categorical predictors intact. In this way, the coefficients of the continuous variables can be estimated more efficiently, and the coefficients of the categorical variables are estimated with about the same efficiency as the QR model. Specifically, let $\mathbf{X} = (\mathbf{X}_1^T, \mathbf{X}_2^T)^T$, where $\mathbf{X}_1 \in \mathbb{R}^{p_1}$ contains the continuous predictors and $\mathbf{X}_2 \in \mathbb{R}^{p_2}$ contains the categorical predictors, $p_1 + p_2 = p$. Then the QR model (2.1) can be written as

$$Q_Y(\tau \mid \mathbf{X}) = \mu_\tau + \boldsymbol{\beta}_{1,\tau}^T \mathbf{X}_1 + \boldsymbol{\beta}_{2,\tau}^T \mathbf{X}_2, \tag{6.1}$$

where $\boldsymbol{\beta}_{1,\tau} \in \mathbb{R}^{p_1}$ is the coefficient vector of $\mathbf{X}_1$ and $\boldsymbol{\beta}_{2,\tau} \in \mathbb{R}^{p_2}$ is the coefficient vector of $\mathbf{X}_2$. Let $\boldsymbol{\mu}_{\mathbf{X}_1}$ and $\boldsymbol{\Sigma}_{\mathbf{X}_1}$ denote the mean and covariance matrix of $\mathbf{X}_1$. With the presence

of $\mathbf{X}_2$, suppose $\mathcal{S}_\tau$ is a subspace of $\mathbb{R}^{p_1}$ that satisfies the following two conditions

$$\text{i) } Q_Y(\tau \mid \mathbf{X}) = Q_Y(\tau \mid \mathbf{P}_{\mathcal{S}_\tau}\mathbf{X}_1, \mathbf{X}_2) \quad \text{and} \quad \text{ii) } \mathrm{Cov}(\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X}_1, \mathbf{Q}_{\mathcal{S}_\tau}\mathbf{X}_1) = 0. \tag{6.2}$$

Then it can shown that $\mathcal{S}_\tau$ is a reducing subspace of $\mathbf{\Sigma}_{\mathbf{X}_1}$ that contains $\boldsymbol{\beta}_{1,\tau}$. The intersection of all such $\mathcal{S}_\tau$ is called the partial $\mathbf{\Sigma}_{\mathbf{X}_1}$-envelope of $\boldsymbol{\beta}_{1,\tau}$, denoted by $\mathcal{E}_{\mathbf{\Sigma}_{\mathbf{X}_1}}(\boldsymbol{\beta}_{1,\tau})$ or $\mathcal{E}_{1,\tau}$ for short. We denote the dimension of $\mathcal{E}_{\mathbf{\Sigma}_{\mathbf{X}_1}}(\boldsymbol{\beta}_{1,\tau})$ as d_τ ($d_\tau \leq p_1$). Since we only consider the envelope on $\boldsymbol{\beta}_{1,\tau}$, $\boldsymbol{\beta}_{2,\tau}$ remains intact. We call (6.1) a partial envelope quantile regression (PEQR) model if conditions (6.2) are incorporated. Let $\mathbf{\Psi}_\tau \in \mathbb{R}^{p_1 \times d_\tau}$ be an orthonormal basis of $\mathcal{E}_{\mathbf{\Sigma}_{\mathbf{X}_1}}(\boldsymbol{\beta}_{1,\tau})$ and $\mathbf{\Psi}_{0,\tau} \in \mathbb{R}^{p_1 \times (p_1 - d_\tau)}$ be a completion of $\mathbf{\Psi}_\tau$. Then the coordinate form of the PEQR model is

$$\begin{aligned} Q_Y(\tau|\mathbf{X}) &= \mu_\tau + \boldsymbol{\eta}_\tau^T \mathbf{\Psi}_\tau^T \mathbf{X}_1 + \boldsymbol{\beta}_{2,\tau}^T \mathbf{X}_2 \\ \mathbf{\Sigma}_{\mathbf{X}_1} &= \mathbf{\Psi}_\tau \mathbf{\Omega}_\tau \mathbf{\Psi}_\tau^T + \mathbf{\Psi}_{0\tau} \mathbf{\Omega}_{0\tau} \mathbf{\Psi}_{0\tau}^T, \end{aligned} \tag{6.3}$$

where $\boldsymbol{\beta}_{1,\tau} = \mathbf{\Psi}_\tau \boldsymbol{\eta}_\tau$, $\boldsymbol{\eta}_\tau \in \mathbb{R}^{d_\tau}$ carries the coordinates of $\boldsymbol{\beta}_{1,\tau}$ with respect to $\mathbf{\Psi}_\tau$, $\mathbf{\Omega}_\tau \in \mathbb{R}^{d_\tau \times d_\tau}$ and $\mathbf{\Omega}_{0,\tau} \in \mathbb{R}^{(p_1 - d_\tau) \times (p_1 - d_\tau)}$ carry the coordinates of $\mathbf{\Sigma}_{\mathbf{X}_1}$ with respect to $\mathbf{\Psi}_\tau$ and $\mathbf{\Psi}_{0\tau}$. Let $s_1 = p_1(p_1+1)/2$. Then the number of parameters in this model is $1 + p_2 + d_\tau + s_1$, reduced from $1 + p_1 + p_2 + s_1$ without enveloping, and the parameter vector is

$$\boldsymbol{\zeta}_{1,\tau} = (\mu_\tau, \mathrm{vec}(\boldsymbol{\eta}_\tau)^T, \mathrm{vec}(\mathbf{\Psi}_\tau)^T, \boldsymbol{\beta}_{2,\tau}^T, \mathrm{vech}(\mathbf{\Omega}_\tau)^T, \mathrm{vech}(\mathbf{\Omega}_{0,\tau})^T)^T.$$

The estimation of parameters in PEQR is similar to that in EQR. We adopt the

reparametrization in (4.1). Let $\boldsymbol{\Psi}_{\tau,1}$ be the matrix that contains the first d_τ rows in $\boldsymbol{\Psi}_\tau$, and let $\boldsymbol{\Psi}_{\tau,2}$ be the matrix that contains the remaining rows in $\boldsymbol{\Psi}_\tau$. Without loss of generality, we assume that $\boldsymbol{\Psi}_{\tau,1}$ is nonsingular. Let $\boldsymbol{\Psi}_\tau^* = \boldsymbol{\Psi}_\tau \boldsymbol{\Psi}_{\tau,1}^{-1}$, $\boldsymbol{\eta}_\tau^* = \boldsymbol{\Psi}_{\tau,1}\boldsymbol{\eta}_\tau$ and $\boldsymbol{\Omega}_\tau^* = \boldsymbol{\Psi}_{\tau,1}\boldsymbol{\Omega}_\tau\boldsymbol{\Psi}_{\tau,1}^T$. Then $\boldsymbol{\Psi}_\tau^* = (\mathbf{I}_{d_\tau}, \mathbf{A}_1^T)^T$, where $\mathbf{A}_1 = \boldsymbol{\Psi}_{\tau,2}\boldsymbol{\Psi}_{\tau,1}^{-1}$. We write $\mathbf{X}_i = (\mathbf{X}_{1,i}^T, \mathbf{X}_{2,i}^T)^T$ and $\mathbf{W}_i = (1, \mathbf{X}_i^T)^T$, $i = 1, \cdots, n$. Under the PEQR model, define

$$
\begin{aligned}
h_n^*(\boldsymbol{\zeta}_{1,\tau}^*) &= \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \mathbf{W}_i\{I[Y_i < \mu_\tau + (\boldsymbol{\Psi}_\tau^*\boldsymbol{\eta}_\tau^*)^T\mathbf{X}_{1,i} + \boldsymbol{\beta}_{2,\tau}^T\mathbf{X}_{2,i}] - \tau\} \\ \frac{1}{n}\sum_{i=1}^n\{\mathrm{vech}(\boldsymbol{\Psi}_\tau^*\boldsymbol{\Omega}_\tau^*\boldsymbol{\Psi}_\tau^{*T} + \boldsymbol{\Psi}_{0\tau}\boldsymbol{\Omega}_{0\tau}\boldsymbol{\Psi}_{0\tau}^T) - \mathrm{vech}[(\mathbf{X}_{1,i} - \boldsymbol{\mu}_{\mathbf{X}_1})(\mathbf{X}_{1,i} - \boldsymbol{\mu}_{\mathbf{X}_1})^T]\} \\ \frac{1}{n}\sum_{i=1}^n(\boldsymbol{\mu}_{\mathbf{X}_1} - \mathbf{X}_{1,i}) \end{pmatrix} \\
&\equiv \frac{1}{n}\sum_{i=1}^n g_n^*(\boldsymbol{\zeta}_{1,\tau}^*), \qquad (6.4)
\end{aligned}
$$

where $\boldsymbol{\zeta}_{1,\tau}^* = (\mu_\tau, \mathrm{vec}(\boldsymbol{\eta}_\tau^*)^T, \mathrm{vec}(\mathbf{A}_1)^T, \boldsymbol{\beta}_{2,\tau}^T, \mathrm{vech}(\boldsymbol{\Omega}_\tau^*)^T, \mathrm{vech}(\boldsymbol{\Omega}_{0,\tau})^T, \boldsymbol{\mu}_{\mathbf{X}_1}^T)^T$. We follow the procedures in Section 4, and use a two-step algorithm to get the GMM estimator of $\boldsymbol{\zeta}_{1,\tau}^*$.

Step 1. Find the estimator $\boldsymbol{\zeta}_{1,\tau}^*$ by minimizing $h_n^*(\boldsymbol{\zeta}_{1,\tau}^*)^T h_n^*(\boldsymbol{\zeta}_{1,\tau}^*)$, and denote it as $\tilde{\boldsymbol{\zeta}}_{1,\tau}^*$.

Step 2. Estimate the optimal weight matrix as

$$
\widehat{\boldsymbol{\Delta}}^{-1} = \left[\frac{1}{n}\sum_{i=1}^n g_n^*(\tilde{\boldsymbol{\zeta}}_{1,\tau}^*) g_n^*(\tilde{\boldsymbol{\zeta}}_{1,\tau}^*)^T\right]^{-1},
$$

and get the GMM estimator $\widehat{\boldsymbol{\zeta}}_{1,\tau}^*$ as the minimizer of the following quadratic form

$$
Q_n(\boldsymbol{\zeta}_{1,\tau}^*) = h_n^*(\boldsymbol{\zeta}_{1,\tau}^*)^T \widehat{\boldsymbol{\Delta}}^{-1} h_n^*(\boldsymbol{\zeta}_{1,\tau}^*).
$$

Then the envelope GMM estimators of $\boldsymbol{\beta}_{1,\tau}$ and $\boldsymbol{\Sigma}_{\mathbf{X}_1}$ are $\widehat{\boldsymbol{\beta}}_{1,\tau} = \widehat{\boldsymbol{\Psi}}^*_\tau \widehat{\boldsymbol{\eta}}^*_\tau$ and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{X}_1} = \widehat{\boldsymbol{\Psi}}^*_\tau \widehat{\boldsymbol{\Omega}}^*_\tau \widehat{\boldsymbol{\Psi}^*}^T_\tau + \widehat{\boldsymbol{\Psi}}_{0\tau} \widehat{\boldsymbol{\Omega}}_{0\tau} \widehat{\boldsymbol{\Psi}}^T_{0\tau}$.

The selection of the dimension for $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}_1}}(\boldsymbol{\beta}_{1,\tau})$ can be performed by RCV.

The asymptotic variance of the envelope GMM estimator can be derived similarly as in Theorem 2. As $\boldsymbol{\beta}_\tau = (\boldsymbol{\beta}^T_{1,\tau}, \boldsymbol{\beta}^T_{2,\tau})^T$, $\boldsymbol{\theta}^*$ in the PEQR setting is then $\boldsymbol{\theta}^* = (\mu_\tau, \boldsymbol{\beta}^T_{1,\tau}, \boldsymbol{\beta}^T_{2,\tau}, \text{vech}(\boldsymbol{\Sigma}_{\mathbf{X}_1})^T)^T$. Let $\widehat{\boldsymbol{\theta}}^*_{pe}$ denote the PEQR estimator of $\boldsymbol{\theta}^*$ and let $\tilde{\boldsymbol{\theta}}^*$ denote the standard estimator of $\boldsymbol{\theta}^*$ by directly solving the estimating equations without enveloping. Let $\boldsymbol{\theta}^*_0$ be the true value of $\boldsymbol{\theta}^*$. As discussed in the EQR model, we ignore $\boldsymbol{\mu}_{\mathbf{X}_1}$ with no loss of generality.

Theorem 3. *Under the same conditions as in Theorem 2, (1)* $\sqrt{n}(\widehat{\boldsymbol{\theta}}^*_{pe} - \boldsymbol{\theta}^*_0)$ *converges in distribution to a normal distribution with mean zero and covariance matrix* $\text{avar}(\sqrt{n}\widehat{\boldsymbol{\theta}}^*_{pe}) = \mathbf{G}(\mathbf{G}^T\mathbf{U}_{pe}\mathbf{V}^{-1}_{pe}\mathbf{U}_{pe}\mathbf{G})^\dagger\mathbf{G}^T$, *where* $\mathbf{G} = \partial\boldsymbol{\theta}^*/\partial\boldsymbol{\zeta}^T_{1,\tau}$ *is the gradient matrix of* $\boldsymbol{\theta}^*$ *relative to* $\boldsymbol{\zeta}_{1,\tau}$,

$$\mathbf{U}_{pe} = \begin{pmatrix} \text{E}_{\boldsymbol{\theta}_0}[f_{Y|\mathbf{X}}(\xi_0(\tau|\mathbf{X}))\mathbf{W}\mathbf{W}^T] & 0 \\ 0 & \mathbf{I}_{s_1} \end{pmatrix}$$

and

$$\mathbf{V}_{pe} = \begin{pmatrix} \tau(1-\tau)\text{E}_{\boldsymbol{\theta}_0}[\mathbf{W}\mathbf{W}^T] & 0 \\ 0 & \text{var}_{\boldsymbol{\theta}_0}\{\text{vech}[(\mathbf{X}_1 - \boldsymbol{\mu}_{\mathbf{X}_1,0})(\mathbf{X}_1 - \boldsymbol{\mu}_{\mathbf{X}_1,0})^T]\} \end{pmatrix},$$

with $\boldsymbol{\mu}_{\mathbf{X}_1,0}$ *being the true value of* $\boldsymbol{\mu}_{\mathbf{X}_1}$.

(2) In addition, $\text{avar}(\sqrt{n}\widehat{\boldsymbol{\theta}}^*_{pe}) \leq \text{avar}(\sqrt{n}\tilde{\boldsymbol{\theta}}^*)$.

Theorem 3 suggests that PEQR improves the estimation efficiency of $\boldsymbol{\beta}_{1,\tau}$, while it does

not sacrifice the estimation efficiency of $\boldsymbol{\beta}_{2,\tau}$. The proof of Theorem 3 is briefly described in Section C of the supplement.

We then demonstrate the performance of PEQR with a simulation and an example. To save space, we present the simulation setting and results in Section D.4 of the supplement, where PEQR demonstrates efficiency gains in estimating the parameters compared to the standard QR. We provide details on the real data analysis below.

We applied the PEQR model to the Boston housing data (Harrison and Rubinfeld, 1978). The data contain housing value and 13 attributes for 506 owner-occupied homes in suburbs of Boston. The 13 attributes include one categorical variable: Charles River dummy variable, which takes value 1 if tract bounds river and 0 otherwise. The 12 continuous variables include crime rate, nitric oxides concentration, pupil-teacher ratio by town and others. Each continuous variable was scaled to have sample standard deviation 1. We took the house value of the homes as response and the 13 attributes as predictors. As the distribution of the response is right skewed, we fit the standard QR model and the PEQR model to the data. RCV suggested $d_\tau = 3$ for $\tau = 0.5$ and $d_\tau = 2$ for $\tau = 0.9$. We computed the bootstrap standard deviation from the standard QR model and the PEQR model for each element in $\boldsymbol{\beta}_\tau$, and took the ratio. The ratios ranged from 0.88 to 3.44 with an average of 2.12 for $\tau = 0.5$, and they ranged from 0.83 to 5.50 with an average of 3.57 for $\tau = 0.9$. The PEQR model demonstrates efficiency gains in this example.

7 Discussion

In this article, the EQR approach along with its variant PEQR are developed to reduce estimation variation and improve efficiency for quantile regression. The new EQR method utilizes information in both the predictors and response by connecting the covariance matrix $\mathbf{\Sigma_X}$ of $\mathbf{X}$ to the parameter of interest $\boldsymbol{\beta}_\tau$ for identifying material and immaterial information in estimating $\boldsymbol{\beta}_\tau$, while synchronously excluding immaterial information from estimation. With this simultaneous dimension reduction and regression fitting, the new method can lead to gains in efficiency. It also advances the recent development of envelopes to general distribution-free procedures with possibly nonsmooth objective functions, and offers new technical tools for justification of asymptotic efficiency. The idea of EQR can be naturally extended to other quantile regression settings, such as censored quantile regression and partially linear quantile regression, for survival and other complex data analysis. On the other hand, since $\mathbf{\Sigma_X}$ is incorporated in the estimation procedure, the number of parameters in EQR can be large when the number of predictors increases. Hence a direct application of EQR to high dimensional settings is hard. To overcome this issue, a penalized EQR model can be considered by imposing sparsity on the parameters $\boldsymbol{\beta}_\tau$, $\mathbf{\Sigma_X}$, and the weighted matrix $\mathbf{\Delta}$ in the GMM estimation, inspired by Su *et al.* (2016) and Qian *et al.* (2018). The theoretical properties of the associated estimators require further investigation. We leave the penalized EQR model as a potentially interesting future research project.

SUPPLEMENTARY MATERIALS

The supplementary materials contain proofs, technical details, and additional simula-

tions.

ACKNOWLEDGEMENT

The authors sincerely thank the Editor, the Associate Editor, and one anonymous referee for their insightful and valuable comments that help improve the article substantially.

REFERENCES

Chen, C. and Wei, Y. (2005). Computational issues for quantile regression. *Sankhyā: The Indian Journal of Statistics*, **67**, 399–417.

Chernozhukov, V. (2005). Extremal quantile regression. *Annals of Statistics*, **33**(2), 806–839.

Conway, J. B. (1990). *A course in functional analysis (2nd ed.)*. New York: Springer.

Cook, R., Helland, I., and Su, Z. (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society: Series B*, **75**(5), 851–877.

Cook, R. D. (2018). *An Introduction to Envelopes: Dimension Reduction for Efficient Estimation in Multivariate Statistics*. John Wiley & Sons.

Cook, R. D. and Su, Z. (2013). Scaled envelopes: scale-invariant and efficient estimation in multivariate linear regression. *Biometrika*, **100**(4), 939–954.

Cook, R. D. and Zhang, X. (2015). Foundations for envelope models and methods. *Journal of the American Statistical Association*, **110**(510), 599–611.

Cook, R. D. and Zhang, X. (2016). Algorithms for envelope estimation. *Journal of Computational and Graphical Statistics*, **25**(1), 284–300.

Cook, R. D., Li, B., and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica*, **20**, 927–1010.

Cook, R. D., Forzani, L., and Zhang, X. (2015). Envelopes and reduced-rank regression. *Biometrika*, **102**(2), 439–456.

Cook, R. D., Forzani, L., and Su, Z. (2016). A note on fast envelope estimation. *Journal of Multivariate Analysis*, **150**, 42–54.

Ding, S. and Cook, R. D. (2018). Matrix variate regressions and envelope models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **80**, 387–408.

Feng, X., He, X., and Hu, J. (2011). Wild bootstrap for quantile regression. *Biometrika*, **98**(4), 995–999.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, **50**(4), 1029–1054.

Harrison, D. and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, **5**(1), 81–102.

He, X. and Shao, Q.-M. (2000). On parameters of increasing dimensions. *Journal of Multivariate Analysis*, **73**(1), 120–135.

He, X. and Zhu, L.-X. (2011). A lack-of-fit test for quantile regression. *Journal of the American Statistical Association*, **98**(464), 1013–1022.

He, X., Zhu, Z.-Y., and Fung, W.-K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika*, **89**(3), 579–590.

Khare, K., Pal, S., and Su, Z. (2017). A bayesian approach for envelope models. *The Annals of Statistics*, **45**(1), 196–222.

Knight, K. (1998). Limiting distributions for l_1 regression estimators under general conditions. *The Annals of Statistics*, **26**(2), 755–770.

Knight, K. (1999). Asymptotics for l1-estimators of regression parameters under heteroscedasticity. *Canadian Journal of Statistics*, **27**(3), 497–507.

Koenker, R. (2005). *Quantile regression*. Cambridge university press.

Koenker, R. (2011). Additive models for quantite regression: Model selection and confidence bandaids. *Brazilian Journal of Probability and Statistics*, **25**(3), 239–262.

Koenker, R. (2017). Quantile regression 40 years on. *Annual Reviews in Economics*, **9**.

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

Koenker, R. and Park, B. J. (1996). An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, **71**(1-2), 265–283.

Koenker, R., Chernozhukov, V., He, X., and Peng, L. (2017). *Handbook of Quantile Regression*. CRC press.

Li, L. and Zhang, X. (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association*, **112**(519), 1131–1146.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, **7**(4), 308–313.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, **4**, 2111–2245.

Noufaily, A. and Jones, M. (2013). Parametric quantile regression based on the generalized gamma distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **62**(5), 723–740.

Oh, H.-S., Nychka, D., Brown, T., and Charbonneau, P. (2004). Period analysis of variable stars by robust smoothing. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **53**(1), 15–30.

Otsu, T. (2003). Empirical likelihood for quantile regression. *University of Wisconsin, Madison Department of Economics Discussion Paper*.

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, **103**(482), 637–649.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, **98**(464), 1001–1012.

Portnoy, S. and Koenker, R. (1997). The gaussian hare and the laplacian tortoise: Computation of squared-errors vs. absolute-errors estimators. *Statistical Science*, **1**, 279–300.

Powell, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. *Nonparametric and semiparametric methods in Econometrics*, pages 357–384.

Qian, W., Ding, S., and Cook, R. D. (2018). Sparse minimum discrepancy approach to sufficient dimension reduction with simultaneous variable selection in ultrahigh dimension. *Journal of the American Statistical Association*, pages 1–48.

Rekabdarkolaee, H. M., Wang, Q., Naji, Z., and Fuentes, M. (2017). New parsimonious multivariate spatial model: Spatial envelope. *arXiv preprint arXiv:1706.06703*.

Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. *Journal of the American Statistical Association*, **81**(393), 142–149.

Su, Z. and Cook, R. D. (2011). Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika*, **98**(1), 133–146.

Su, Z. and Cook, R. D. (2012). Inner envelopes: efficient estimation in multivariate linear regression. *Biometrika*, **99**(3), 687–702.

Su, Z., Zhu, G., Chen, X., and Yang, Y. (2016). Sparse envelope model: efficient estimation and response variable selection in multivariate linear regression. *Biometrika*, **103**(3), 579–593.

Van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press.

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

Wang, H. and Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association*, **104**(487), 1117–1128.

Wang, L. and Ding, S. (2018). Vector autoregression and envelope model. *Stat*, **7**(1), e203.

Watnik, M. R. (1998). Pay for play: Are baseball salaries based on performance? *Journal of Statistics Education*, **6**(2).

Wei, Y., Pere, A., Koenker, R., and He, X. (2006). Quantile regression methods for reference growth charts. *Statistics in Medicine*, **25**(8), 1369–1382.

Xu, G., Sit, T., Wang, L., and Huang, C.-Y. (2017). Estimation and inference of quantile regression for survival data under biased sampling. *Journal of the American Statistical Association*, **112**(520), 1571–1586.

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

Zhu, G. and Su, Z. (2019). Envelope-based sparse partial least squares. *Annals of Statistics*. To appear.