Quantile regression with generated dependent variable and covariates

Jayeeta Bhattacharya*

Queen Mary University of London

September 2019

Abstract

We study linear quantile regression models when regressors and/or dependent variable are not directly observed but estimated in an initial first step and used in the second step quantile regression for estimating the quantile parameters. This general class of generated quantile regression (GQR) covers various statistical applications, for instance, estimation of endogenous quantile regression models and triangular structural equation models, and some new relevant applications are discussed. We study the asymptotic distribution of the two-step estimator, which is challenging because of the presence of generated covariates and/or dependent variable in the non-smooth quantile regression estimator. We employ techniques from empirical process theory to find uniform Bahadur expansion for the two step estimator, which is used to establish the asymptotic results. We illustrate the performance of the GQR estimator in a simulation exercise and an empirical application based on auctions.

Keywords: Two-stage estimation, generated covariates, generated dependent variable, quantile regression, asymptotic variance

JEL Codes: C13, C14, C21, C31

*I am deeply thankful to Emmanuel Guerre for his valuable comments and supervision. I also thank participants at various conferences for insightful comments. Generous funding from the School of Economics and Finance, Queen Mary University of London, is also gratefully acknowledged. Correspondence email: j.bhattacharya@qmul.ac.uk

1 Introduction

1.1 Background

Econometric analysis often requires the use of regressors that are not directly observed but have been estimated in a preliminary first step. A rich literature exists on estimation and inference in models with generated regressors. Pagan (1984) and Mammen, Rothe & Schienle (2012) study parametric and non-parametric regression with generated covariates, respectively, while Hahn & Ridder (2013) and Mammen, Rothe & Schienle (2016) study asymptotic properties of semiparametric estimators. Murphy & Topel (1985) studies two step estimators in parametric context and points out that ignoring the effect of first step estimation leads to incorrect asymptotic standard errors.

While these models are concerned with the characterization of the conditional mean, a more complete picture of the conditional distribution of a dependent variable is provided by quantile regression (QR) models. Since the seminal work of Koenker & Bassett (1978), quantile regression is widely used in both empirical studies and theoretical statistics for analysing conditional quantile functions in linear and nonlinear response models. Quantile regression applications using generated regressors abound in literature, most prominently related to models with endogenous covariates. Chernozhukov & Hansen (2005, 2006, 2008) develop identification and estimation for QR models in the presence of endogeneity. Another popular approach to deal with endogeneity uses the estimated reduced form residuals as control variables in quantile regression. This technique has been applied in endogenous censored quantile regression models by Blundell & Powell (2007) and Chernozhukov, Fernández-Val & Kowalski (2015). Estimation of quantile treatment effects or quantile parameters in triangular simultaneous equation models using the control variable approach have been considered in Chesher (2003), Koenker & Ma (2006), Lee (2007), Imbens & Newey (2009), and Chernozhukov, Fernández-Val, Newey, Stouli & Vella (2017).

There are, however, few references that develop a general theory for quantile models with generated covariates and systematically study its statistical properties. The only related work seems to be Chen, Galvao, & Song (2018), who consider estimation and inference of quantile regression when regressors are generated. However, they seem to implicitly assume that using various first stage estimators does not cause dependence among them and obtain asymptotic results under independence, while our framework does not require such an assumption. They also do not consider generated dependent variable as permitted here.

1.2 Motivation

We consider quantile regression models when either the regressors or the dependent variable (or both) are generated, and study its asymptotic behaviour. An example giving rise to generated dependent variable is quantile specifications with some constant slope parameters, as in the setup of Zou & Yuan (2008). Their composite quantile regression (CQR) method can be used to estimate the constant and quantile-varying parameters together, but its asymptotic properties have been studied for estimation of the constant parameters only. To focus on the quantile parameters, as an alternative to CQR, the constant slope parameters can be estimated by linear regression in a first step. Estimation of the quantile-varying slope parameters, thereafter, involves quantile regression with the dependent variable generated as a function of the constant slope parameters. Also, removing some parameters through the first step estimation may alleviate the computational burden of the CQR method caused by a large number of variables. Another example arises in quantile models where the dependent variable is transformed based on some transformation parameter, like Box-Cox transformation, to induce some desirable properties for statistical inference. In this example, a joint estimation of quantile varying transformation and slope parameters is computationally difficult, in addition to a numerical problem being that the objective function is not defined for all parameter values and observations (meaning estimation occurs by omitting such values). Estimating the transformation parameter in a first step will avoid such numerical problem and involve a linear quantile regression, ensuring a better performance of the numerical algorithm used to compute the estimator.

It is well known that the first step estimates impact the overall asymptotic behaviour of the final estimator, understanding which is crucial for obtaining consistent standard errors which can be used for constructing correct confidence intervals. The wide range of quantile regression applications that give rise to generated regressors or dependent variable obtained from estimation in a preliminary step suggest the need for a systematic analysis of their impact on the statistical properties of the QR estimator. The classical way in which asymptotic analysis is carried out for two step estimators with smooth objective functions relies on a Taylor expansion based technique for the second stage estimates, as applied in Murphy & Topel (1985). However, such methods are not applicable for the QR estimator, since it is difficult to differentiate the QR estimator[1]. Finding the asymptotic variance of such an estimator is not a trivial task and requires alternative techniques.

[1] This could be done in principle by applying the Implicit Function Theorem to the first-order condition that defines the estimator. However, the QR estimator is not always unique and the QR objective function is not twice differentiable, preventing the use of this approach.

1.3 Contributions of the paper

We propose a two-step estimator for QR models with generated variables, which we call the generated quantile regression (GQR) estimator. We study the asymptotic property of the GQR estimator using techniques from asymptotic analysis for quantile regression and empirical process theory. A Bahadur expansion of the GQR estimator is derived, with precise stochastic order of the remainder term, which holds uniformly with respect to the first step parameter and the quantile levels. Using the Bahadur expansion approach, under the assumption that the first stage estimation is asymptotically normal and some other regularity conditions, we establish asymptotic normality and obtain explicit expression for the asymptotic variance of the proposed GQR estimator. To the best of our knowledge, this is the first work that analyses quantile regression with generated variables without being tailored to any specific application, and systematically handles the associated issues for asymptotic analysis.

The application of the GQR estimator is illustrated through three motivating examples - quantile regression involving constant slope parameters, a variant for endogenous quantile regression model, and a Box-Cox power transformed quantile regression. In particular, the GQR estimator can be an alternative to Zou & Yuan (2008)'s composite quantile regression method for estimating QR models with some constant slope parameters. A simulation exercise based on this example illustrates the validity of the asymptotic normality result. Further analysis of QR models when some slope parameters are known to be constant shows that the GQR estimator produces an efficiency gain, except in the extreme tails, over the standard QR estimator. Finally, an empirical application based on auction models in quantile framework confirms that the GQR estimator improves the estimation of quantile parameters as compared to an unconstrained estimation using standard quantile regression.

The rest of the paper is organised as follows. Section 2 introduces the baseline model and the GQR estimator, and presents three applications to motivate the framework. Section 3 carries out the asymptotic analysis and presents the Bahadur expansion results and the central limit theorem for the GQR estimator. The asymptotic results are applied to the motivating examples in Section 4. Section 5 compares the efficiency of the GQR estimator with the standard QR estimator using the constant slope QR model example. Section 6 presents simulation results while Section 7 reports results of the empirical application to first price auctions. Proofs of the main results are given in the Appendices.

2 Quantile regression with generated variables

We consider the following linear quantile specification.

$$Y(\theta) = X(\theta)'\beta(U); \quad U|X(\theta) \sim \mathcal{U}[0,1], \tag{2.1}$$

where, provided that $\tau \mapsto X(\theta)'\beta(\tau)$ is strictly increasing and continuous in τ, $X(\theta)'\beta(\tau)$ is the τ-quantile of $Y(\theta)$ conditional on $X(\theta)$. Here, $Y(\theta)$ and $X(\theta)$ are functions of a vector of parameters θ, which includes elements that generate the dependent variable Y, or the regressor X, or both. The true value of the parameter θ in (2.1), denoted by θ_0, is not known but estimated. Hence, we propose a two-step estimation of the above quantile model.

First step: Estimation of θ_0. It is assumed that a consistent estimator $\widehat{\theta}$ is available. For the sake of generality, any estimation method is allowed at this stage, provided it satisfies an expansion typical of regular estimators, see for example Newey & McFadden (1994). As discussed for the examples, a suitable choice of $\widehat{\theta}$ can be done case by case.

Second step: Estimation of quantile parameter. The quantile parameter estimate $\widehat{\beta}(\tau)$ in (2.1) is given by

$$\widehat{\beta}(\tau) = \widehat{\beta}(\tau; \widehat{\theta}) = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau \left(Y_i(\widehat{\theta}) - X_i(\widehat{\theta})'\beta \right), \tag{2.2}$$

where $\rho_\tau(u) = (\tau - \mathbb{I}(u < 0))\, u$ is the check function of Koenker & Bassett (1978).

2.1 Motivating examples

The general framework of quantile regression with dependent variable and/or covariates obtained as a function of parameters estimated in a first step finds wide application in economics and statistics. We present three applications. The asymptotic results for these applications are discussed in a later section.

2.1.1 Quantile regression with constant slope

Consider the quantile regression (QR) model

$$Q_Y(\tau|X) = \beta_0(\tau) + \beta_1(\tau) X_1 + \beta_2(\tau) X_2 \tag{2.3}$$

and assume that $\beta_1(\tau) = \beta_1$ for all τ, ie $\beta_1(\cdot)$ is constant. This model can be estimated using Zou & Yuan (2008)'s composite quantile regression (CQR) method as follows:

$$\left(\widehat{\beta}_1, \widehat{\beta}_0(\tau_1), \widehat{\beta}_2(\tau_1), \cdots, \widehat{\beta}_0(\tau_K), \widehat{\beta}_2(\tau_K)\right) = \arg \min_{\substack{b_1, b_{0k}, b_{2k};\\ k=1,\cdots,K}} \sum_{k=1}^{K} \sum_{i=1}^{n} \rho_{\tau_k} \left(Y_i - X_{1i} b_1 - b_{0k} - X_{2i} b_{2k}\right),$$

for $0 < \tau_1 < \tau_2 < \cdots < \tau_K < 1$. This could lead to an intractable system due to very large number of variables, especially with more quantile parameters and quantile levels. Moreover, Zou & Yuan (2008) studies the asymptotic properties of the CQR estimator for estimation of constant slope parameters and compares efficiency with least squares, while the asymptotic behaviour for quantile varying slope parameters remains unstudied. As an alternative to Zou & Yuan (2008), consider a two step estimation of this model as described below.

As there exist uniform variables U_i independent of X_i such that $Y_i = Q_Y(U_i|X_i)$, it holds

$$Y_i = \overline{\beta}_0 + \overline{\beta}_1 X_{1i} + \overline{\beta}_2 X_{2i} + \varepsilon_i$$

where $\overline{\beta}_k = \mathbb{E}\left[\beta_k(U_i)\right]$, $k = \{0, 1, 2\}$, and $\varepsilon_i = \beta_0(U_i) - \overline{\beta}_0 + \left(\beta_2(U_i) - \overline{\beta}_2\right) X_{2i}$ (since $\overline{\beta}_1 = \beta_1 = \beta_1(U_i)$). It follows that the $\overline{\beta}_k$'s can be estimated using OLS, that is,

$$\left(\widehat{\overline{\beta}}_0, \widehat{\overline{\beta}}_1, \widehat{\overline{\beta}}_2\right) = \arg \min_{b_0, b_1, b_2} \sum_{i=1}^{n} \left(Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i}\right)^2. \tag{2.4}$$

Set $\widehat{\overline{\beta}}_1 = \widehat{\beta}_1$. A two step estimator of $(\beta_0(\tau), \beta_2(\tau))$ is then

$$\left(\widehat{\beta}_0(\tau), \widehat{\beta}_2(\tau)\right) = \arg \min_{b_0, b_2} \sum_{i=1}^{n} \rho_\tau \left(Y_i - \widehat{\beta}_1 X_{1i} - b_0 - b_2 X_{2i}\right). \tag{2.5}$$

Hence, in this example, the first step parameter is $\theta \equiv \beta_1$, and the dependent variable is generated as $Y_i(\beta_1) = Y_i - \beta_1 X_{1i}$.

2.1.2 Endogeneity in quantile regression - control variable approach

Control variable approach views endogeneity bias as an omitted variable bias and proceeds by estimating the 'control variable' which is the reduced form residual, conditional on which error becomes independent of the regressors (see Blundell & Powell (2003)).

Consider the two stage quantile regression model with endogeneity

$$\begin{aligned} Y &= X'\beta(U) + \eta'\lambda(U), \\ X &= Z'\gamma + \eta \end{aligned} \tag{2.6}$$

where η, $U \sim \mathcal{U}_{[0,1]}$ and Z are independent, η being centered with a finite variance. Then if $Q_{Y|X,\eta}(\tau|X,\eta) = X'\beta(\tau) + \eta'\lambda(\tau)$ is increasing in τ for all admissible X and η, the coefficients $\beta(\cdot)$ and $\lambda(\cdot)$ can be estimated from the first stage least squares estimation of the control variable η,

$$\widehat{\eta}_i = X_i - Z_i'\widehat{\gamma}, \quad \widehat{\gamma} = \left(\sum_{i=1}^{N} Z_i Z_i'\right)^{-1} \sum_{i=1}^{N} Z_i X_i. \tag{2.7}$$

The second stage estimator is

$$\left[\widehat{\beta}'(\tau), \widehat{\lambda}'(\tau)\right]' = \arg\min_{\beta,\lambda} \sum_{i=1}^{N} \rho_\tau\left(Y_i - X_i'\beta - \widehat{\eta}'\lambda\right) = \arg\min_{\beta,\lambda} \sum_{i=1}^{N} \rho_\tau\left(Y_i - X_i'\beta - (X_i - Z_i'\widehat{\gamma})'\lambda\right). \tag{2.8}$$

Hence, in this example, the first step estimator is $\theta \equiv \gamma$, and the second stage involves quantile regression of Y_i on generated regressors, $X_i(\theta) \equiv \left[X_i', (X_i - Z_i'\gamma)'\right]'$.

2.1.3 Box-Cox power transformation

Box & Cox (1964) proposes finding a transformation parameter λ such that with the following transformation on the original observations Y,

$$Y(\lambda) = \begin{cases} \frac{Y^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0, \\ \log Y, & \text{if } \lambda = 0, \end{cases} \tag{2.9}$$

$Y(\lambda)$ is normally distributed with conditional variance σ^2, and $\mathbb{E}[Y(\lambda)|X] = X'\beta$. The desirous property for quantile regression is linearity, that is,

$$Q_{Y(\lambda)}(\tau|X) = X'\beta(\tau).$$

The Box-Cox quantile regression literature has mostly focussed on finding a quantile dependent transformation parameter (see, for instance, Powell (1991), Chamberlain (1994), Buchinsky (1995), Machado & Mata (2000) and Fitzenberger, Wilke & Zhang (2009)). Owing to the equivariance property of quantiles, this leads to minimization of the non-linear function $\sum_{i=1}^{n} \rho_\tau\left(Y_i - (\lambda X_i'\beta + 1)^{1/\lambda}\right)$. Quantile varying λ adds flexibility to the model, but joint estimation of $(\lambda(\tau), \beta(\tau))$ requires effort, see Koenker (2017). Also, a basic numerical problem is that $(\lambda X_i'\beta + 1)$ needs to be positive for all λ and all observations.

A constrained estimation with a constant λ has obvious computational and numerical benefits. Mu & He (2007) considers constancy of $\lambda(\tau)$. In the empirical application of

Buchinsky (1995) studying transformation of log wages over 25 years, $\lambda(\tau)$ seems to be constant for all quantiles except the highest. A simpler approach would, therefore, involve estimating $\widehat{\lambda}$ separately in a first step and thereafter performing linear quantile regression using the transformed Y for estimating $\beta(\tau)$. $\widehat{\lambda}$ can be estimated from the linear regression $Y(\lambda) = X'\beta + \varepsilon$. A consistent estimator for $\widehat{\lambda}$ is Amemiya (1974)'s nonlinear IV (NIV) estimator,

$$\left(\widehat{\lambda}_{NIV}, \widehat{\beta}_{NIV}\right) = \arg\min_{\ell,b} \left(\sum_{i=1}^{n} \left(Y_i\left(\ell\right) - X_i'b\right) W_i'\right) \Omega \left(\sum_{i=1}^{n} W_i \left(Y_i\left(\ell\right) - X_i'b\right)\right), \tag{2.10}$$

where W_i always contains X_i as well as additional instruments (Amemiya & Powell (1981) recommends using squares and cross-products of X_i's). Set $\widehat{\lambda} = \widehat{\lambda}_{NIV}$. The dependent variables $Y_i(\widehat{\lambda})$ is, then, generated using equation (2.9). $\beta(\tau)$ is estimated from quantile regression of $Y(\widehat{\lambda})$ on X,

$$\widehat{\beta}\left(\tau\right) = \arg\min_{b} \sum_{i=1}^{n} \rho_\tau \left(Y_i(\widehat{\lambda}) - X_i'b\right). \tag{2.11}$$

3 Asymptotic analysis

Our main assumptions are as follows:

Assumption 1 (First step estimator) *There exists a function $\psi(z)$ such that the estimator of the true θ_0 is asymptotically linear:*

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Psi(z_i) + o_{\mathbb{P}}(1), \quad \mathbb{E}\left[\psi(z)\right] = 0, \quad \mathbb{E}\left[\psi(z)\psi(z)'\right] < \infty.$$

Assumption 2 (Model) *(X_i, Y_i) are i.i.d. There exists a compact set Θ with a non empty interior containing θ_0 such that $X_i\left(\theta\right) = h\left(X_i, \theta\right)$ and $Y_i\left(\theta\right) = g\left(Y_i, X_i, \theta\right)$ are continuous and differentiable with respect to θ in Θ for all (Y_i, X_i). It holds moreover that*

$$\sup_{\theta\in\Theta} \left\| \frac{\partial g\left(Y, X, \theta\right)}{\partial \theta} \right\| < \infty.$$

In the next Assumption $F\left(y|x,\theta\right)$ and $f\left(y|x,\theta\right)$ stands for the c.d.f. and p.d.f. of $Y\left(\theta\right)$ given $X\left(\theta\right)$, $f_X\left(\cdot|\theta\right)$ being the p.d.f. of $X\left(\theta\right)$. The set $\mathcal{X}\left(\theta\right)$ is the support of $X\left(\theta\right)$. All p.d.f. are defined with respect to the Lebesgue measure. The set Θ is as in Assumption 2.

Assumption 3 (Smoothness) *(i) $X(\theta)$ lies in $\mathbb{R}^d$ for each θ and $\mathcal{X}(\theta)$ is a compact subset of $\mathbb{R}^d$ with non empty interior. $f_X(x|\theta) > 0$ over the interior of $\mathcal{X}(\theta)$ and vanishes at its boundaries. $f_X(x|\theta)$ is continuously differentiable with respect to θ. (ii) the p.d.f. $f(y|x,\theta)$ of $Y(\theta)$ given X is continuously differentiable in (y,x,θ) with $f(y|x,\theta) > 0$ for all (y,x,θ) such that $(x,\theta) \in \bigcup_{\theta\in\Theta} \{\mathcal{X}(\theta) \times \theta\}$ and y is in the interior of the support of $F(\cdot|x,\theta)$.*

Asymptotically linear estimators in Assumption 1 refer to the class of extremum estimators as considered in Newey & McFadden (1994). Examples include MLE, NLS, and the GMM class. It implies $\sqrt{n}$-consistency of the first step estimator and is key to the derivation of the asymptotic normality result for the second-step estimator. The triangular structure imposed by Assumption 2 ensures that $X(\theta)$ is not a function of Y and therefore remains exogenous; it is useful in the example of Section 2.1.1. Assumption 3-(ii) is a high level assumption that can be derived from Assumption 2 and the quantile regression slope $\beta(\cdot)$ since $g(Y,X,\theta_0) = X(\theta_0)' \beta(U)$. It implicitly requests a monotone $g(\cdot,X,\theta)$ with non zero derivatives, as $f(\cdot|x,\theta)$ may diverge otherwise. Indeed, if $\partial g(y,x,\theta)/\partial y > 0$ and $f(y|x)$ is the p.d.f. of Y given X (assuming $X(\theta) = X$ for the sake of the brevity of this discussion), it holds

$$f(y|x,\theta) = \frac{1}{\frac{\partial g}{\partial y}\left[g^{-1}(y,x,\theta),x,\theta\right]} f\left[g^{-1}(y,x,\theta)|x\right]$$

which may not be bounded if $\partial g(y,x,\theta)/\partial y$ vanishes. Assumption 3-(ii) then holds if $f(y|x)$ is continuously differentiable in (x,y) and $g(y,x,\theta)$ twice differentiable with respect to y and θ with bounded partial derivatives. Assumption 3-(i) is similar, but note that the transformation $X(\theta) = h(X,\theta)$ does not need to be one to one, as $X(\theta)$ may have a smaller dimension than X.

The QR estimator of the slope coefficient is an estimator of $\beta\left(\tau;\widehat{\theta}\right)$ where

$$\beta(\tau;\theta) = \arg\min_{\beta} \mathbb{E}\left[\rho_\tau\left(Y(\theta) - X'(\theta)\beta\right)\right].$$

Assumption 3 ensures that the objective function above is strictly convex for all θ, so that $\beta(\tau;\theta)$ is the unique solution of the first order condition

$$0 = \mathbb{E}\left[\left\{\mathbb{I}\left(Y(\theta) \le X'(\theta)\beta\right) - \tau\right\} X(\theta)\right] = \mathbb{E}\left[\left\{F\left(X'(\theta)\beta\right| X,\theta) - \tau\right\} X(\theta)\right]$$

This together with the Implicit Function Theorem implies that $\beta(\tau;\theta)$ is differentiable with respect to θ, as established in the following Proposition.

Proposition 1 *Under Assumptions 2 and 3,* $\beta(\tau;\theta)$ *is continuously differentiable with respect to* θ *for any* $\theta \in \Theta$ *and* $0 < \tau < 1$*. It holds moreover*

$$\frac{\partial \beta(\tau;\theta_0)}{\partial \theta} = H(\tau;\theta_0)^{-1} D(\tau;\theta_0)$$

where

$$\begin{aligned} H(\tau;\theta_0) &= \mathbb{E}\left[f\left(X'(\theta_0)\beta(\tau;\theta_0)\right| X,\theta_0) X(\theta_0) X'(\theta_0)\right] \\ D(\tau;\theta_0) &= -\frac{\partial}{\partial\theta}\left[\mathbb{E}\left[\left\{F\left(X'(\theta)\beta\right| X,\theta) - \tau\right\} X(\theta)\right]\right]\Big|_{\theta=\theta_0,\beta=\beta(\tau;\theta_0)}. \end{aligned}$$

Proof of Proposition 1: See proof section.

The matrix $H(\tau;\theta_0)$ plays an important role in the asymptotic distribution of standard QR estimators, see below and Koenker (2005). The existence of its inverse is established in Lemma 2 of the Proof Section. The matrix $D(\tau;\theta_0)$ is specific to two stage estimation. With known θ_0, a linear representation for $\sqrt{n}\left(\widehat{\beta}(\tau;\theta_0) - \beta(\tau,\theta_0)\right)$ can be found in Koenker (2005) Section 4.3, among others. But estimating the parameter θ induces some important changes compared to a known θ_0 and requires finding an expansion for $\sqrt{n}\left(\widehat{\beta}\left(\tau;\widehat{\theta}\right) - \beta\left(\tau,\widehat{\theta}\right)\right)$. The approach used here builds on a Bahadur expansion which holds uniformly in θ and τ. For this purpose define

$$\widehat{S}(\tau;\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\mathbb{I}\left(Y_i(\theta) \leq X_i'(\theta)\beta(\tau;\theta)\right) - \tau\right] X_i(\theta), \tag{3.1}$$

$$J(\tau;\theta) = \tau(1-\tau)\mathbb{E}\left[X_i(\theta) X_i'(\theta)\right] \tag{3.2}$$

$$\widehat{\mathcal{E}}(\tau;\theta) = \sqrt{n}\left(\widehat{\beta}(\tau;\theta) - \beta(\tau;\theta)\right) - \left(-H^{-1}(\tau;\theta)\widehat{S}(\tau;\theta)\right). \tag{3.3}$$

Note that $\widehat{S}(\tau;\theta)/\sqrt{n}$ is the score of the objective function for a given θ. $\widehat{S}(\tau;\theta)$ is centered for $0 < \tau < 1$ with variance $J(\tau;\theta)$. $\widehat{\mathcal{E}}(\tau;\theta)$ is the Bahadur remainder term which is studied in the next Proposition.

Proposition 2 *Under Assumptions 1-3 it holds for any compact parameter set* Θ *and* $C > 0$

$$\sup_{(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\times\Theta}\left\|\widehat{\mathcal{E}}(\tau;\theta)\right\| = O_{\mathbb{P}}\left(\frac{\log^{3/4} n}{n^{1/4}}\right), \tag{3.4}$$

$$\sup_{(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\times\mathcal{B}\left(\theta_0,Cn^{-1/2}\right)}\left\|H^{-1}(\tau;\theta)\widehat{S}(\tau;\theta) - H^{-1}(\tau;\theta_0)\widehat{S}(\tau;\theta_0)\right\| = O_{\mathbb{P}}\left(\frac{\log^{1/2} n}{n^{1/4}}\right) \tag{3.5}$$

where $0 < \underline{\tau} \leq \overline{\tau} < 1$ *and* $\mathcal{B}(\theta_0,\varrho) = \{\theta; \|\theta - \theta_0\| \leq \varrho\}$.

Proof of Proposition 2: See proof section.

Propositions 1 and 2 give the next Theorem, which states a Central Limit Theorem for the two step estimator of the slope coefficient. Note that in absence of the parameter θ, the rate of convergence is the same as that for usual quantile regression estimator, as derived in Theorem 4.1 of Koenker (2005).

Theorem 1 *Under Assumptions 1-3, it holds for any* τ *in* $(0,1)$

$$\sqrt{n}\left(\widehat{\beta}(\tau)-\beta(\tau)\right)\xrightarrow{d}\mathcal{N}(0,V(\tau))$$

where

$$\begin{aligned}V(\tau) &= H(\tau;\theta_0)^{-1}\left[J(\tau;\theta_0)+D(\tau;\theta_0)C_{\Psi S}(\tau)+C'_{\Psi S}(\tau)D'(\tau;\theta_0)\right.\\ &\quad \left.+D(\tau;\theta_0)C_{\Psi\Psi}D'(\tau;\theta_0)\right]H(\tau;\theta_0)^{-1},\\ C_{\Psi\Psi} &= \mathbb{E}\left[\Psi(Z)\Psi'(Z)\right],\quad C_{\Psi S}(\tau)=\mathbb{E}\left[\Psi(Z)X'(\theta_0)\left\{\mathbb{I}\left[Y(\theta_0)\leq X'(\theta_0)\beta(\tau;\theta_0)\right]-\tau\right\}\right].\end{aligned}$$

Proof of Theorem 1. Proposition 1 yields that

$$\begin{aligned}\sqrt{n}\left(\widehat{\beta}\left(\tau;\widehat{\theta}\right)-\beta(\tau;\theta_0)\right) &= \sqrt{n}\left(\widehat{\beta}\left(\tau;\widehat{\theta}\right)-\beta\left(\tau,\widehat{\theta}\right)\right)+\sqrt{n}\left(\beta\left(\tau;\widehat{\theta}\right)-\beta(\tau;\theta_0)\right)\\ &= \sqrt{n}\left(\widehat{\beta}\left(\tau;\widehat{\theta}\right)-\beta\left(\tau,\widehat{\theta}\right)\right)+\left(\frac{\partial\beta(\tau;\theta_0)}{\partial\theta}+o_{\mathbb{P}}(1)\right)\sqrt{n}\left(\widehat{\theta}-\theta_0\right)\\ &= \sqrt{n}\left(\widehat{\beta}\left(\tau;\widehat{\theta}\right)-\beta\left(\tau,\widehat{\theta}\right)\right)\\ &\quad+\left(\frac{\partial\beta(\tau;\theta_0)}{\partial\theta}\right)'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi(Z_i)+o_{\mathbb{P}}(1)\end{aligned}\tag{3.6}$$

where the last line holds thanks to Assumption 1. Equation (3.6) and Proposition 2 give

$$\begin{aligned}\sqrt{n}\left(\widehat{\beta}(\tau)-\beta(\tau)\right) &= H^{-1}\left(\tau;\widehat{\theta}\right)\widehat{S}\left(\tau;\widehat{\theta}\right)+\left(\frac{\partial\beta(\tau;\theta_0)}{\partial\theta}\right)'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi(Z_i)+o_{\mathbb{P}}(1)\\ &= H^{-1}(\tau;\theta_0)\widehat{S}(\tau;\theta_0)+\left(\frac{\partial\beta(\tau;\theta_0)}{\partial\theta}\right)'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi(Z_i)+o_{\mathbb{P}}(1),\end{aligned}\tag{3.7}$$

where the last line results from (3.5) since Assumption 1 and taking C large enough ensure that $\widehat{\theta}$ belongs to $\mathcal{B}\left(\theta_0,Cn^{-1/2}\right)$ with high probability. Since $\frac{\partial\beta(\tau;\theta_0)}{\partial\theta}=H(\tau;\theta_0)^{-1}D(\tau;\theta_0)$ from Proposition 1, the Limit distribution of Theorem 1 follows from the Multivariate CLT.□

Remark 1. As Propositions 1 and 2 hold uniformly in τ, the expansion (3.7) also does. Since Functional Central Limit Theorems for $\widehat{S}(\tau;\theta_0)$ can be applied, (3.7) can be used to obtain a Functional Central Limit Theorem for the two step quantile regression estimator.

Remark 2. The order of the $o_{\mathbb{P}}(1)$ remainder term in (3.7) can be made more precise, strengthening the smoothness Assumptions 2 and 3 to ensure that $\beta(\tau;\theta)$ is twice continuously differentiable using the Implicit Function Theorem as in Proposition 1. Indeed, if $\beta(\tau;\theta)$ is twice continuously differentiable with respect to θ, the $o_{\mathbb{P}}(1)$ remainder term in (3.6) is an $O_{\mathbb{P}}\left(n^{-1/2}\right)$ and the order of the $o_{\mathbb{P}}(1)$ remainder term in (3.7) follows from (3.4) and is $O_{\mathbb{P}}\left(n^{-1/4}\log^{3/4} n\right)$.

Remark 3. The proof can be easily modified for the case where θ depends upon τ.

4 Examples revisited

In this section, we apply the asymptotic theory results of Section 3 to the motivating examples introduced in Section 2.1.

4.1 Quantile regression with constant slope

For the quantile regression model (2.3), the constant paramater $\beta_1(\cdot)$ is estimated using least squares regression, and the quantile parameters $(\beta_0(\cdot), \beta_2(\cdot))$ are estimated using the generated dependent variable $Y_i(\widehat{\beta}_1) = Y_i - \widehat{\beta}_1 X_{1i}$ via the two-step quantile regression estimator of (2.5). Asymptotic normality of the first step OLS estimator is well established. Denote $X = [1, X_1, X_2]'$. Assume that $\mathbb{E}[\varepsilon^2 XX']$ is finite and $\mathbb{E}[XX']$ is full rank and finite. The OLS estimator is asymptotically linear:

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) = \sum_{i=1}^{n} \left[\mathbb{E}^{-1}\left[XX'\right] X_i \varepsilon_i\right] / \sqrt{n} + o_{\mathbb{P}}(1).$$

Denoting $i_{22} = [0, 1, 0]$, the asymptotic variance of $\widehat{\beta}_1$ is given by

$$\mathcal{V}(\beta_1) = i_{22}\left(\mathbb{E}^{-1}[XX']\mathbb{E}[\varepsilon^2 XX']\mathbb{E}^{-1}[XX']\right) i_{22}'. \tag{4.1}$$

For the second step quantile regression, the dependent variable is generated as $Y(\widehat{\beta}_1) = Y - \widehat{\beta}_1 X_1$, and the regressors are denoted as $\widetilde{X} = [1, X_2]'$. Asymptotic normality of the quantile parameters $\beta(\tau) = (\beta_0(\tau), \beta_2(\tau))'$ follows directly from Theorem 1:

$$\sqrt{n}\begin{bmatrix}\widehat{\beta}_0(\tau)-\beta_0(\tau)\\ \widehat{\beta}_2(\tau)-\beta_2(\tau)\end{bmatrix}\xrightarrow{d}\mathcal{N}(0,V(\tau)).$$

The terms of V are obtained from Theorem 1 by replacing $\theta_0\equiv\beta_1$, $\beta(\tau)\equiv(\beta_0(\tau),\beta_2(\tau))'$, $X(\theta_0)\equiv\widetilde{X}=[1,X_2]'$and $Y(\theta_0)\equiv Y(\beta_1)=Y-\beta_1X_1$. Denoting the first τ-derivative of $\beta(\tau)$ as $\beta^{(1)}(\tau)$, $V(\tau)$ comes as follows:

$$\begin{aligned}V(\tau)&=H(\tau)^{-1}\left\{J(\tau)+D(\tau)\mathcal{V}(\beta_1)D(\tau)'+C(\tau)'D(\tau)+C(\tau)D(\tau)'\right\}H(\tau)^{-1},\\ \text{where } H(\tau)&=\mathbb{E}\left[\frac{\widetilde{X}\widetilde{X}'}{\beta_0^{(1)}(\tau)+\beta_2^{(1)}(\tau)X_2}\right],\quad J(\tau)=\tau(1-\tau)\mathbb{E}\left[\widetilde{X}\widetilde{X}'\right],\\ D(\tau)&=-\mathbb{E}\left[\frac{X_1\widetilde{X}}{\beta_0^{(1)}(\tau)+\beta_2^{(1)}(\tau)X_2}\right]\text{and}\\ C(\tau)&=\mathbb{E}\left[g(X)\left\{\int_0^\tau(\beta_0(t)+\beta_2(t)X_2)\,dt-\tau\left(\beta_0(\tau)+\beta_2(\tau)X_2\right)\right\}\right],\end{aligned}\tag{4.2}$$

with $g(X)=\widetilde{X}\,[0,1,0]\,\mathbb{E}^{-1}\left[XX'\right]X$.

4.2 Endogeneity in quantile regression - control variable approach

The endogenous quantile regression model in (2.6) is estimated in two steps. The first step uses OLS estimator of (2.7) to estimate $\widehat{\gamma}$. This is used to generate the control variable $\widehat{\eta}=(X_i-Z_i'\widehat{\gamma})$, which is included as a regressor in the quantile regression estimator of (2.8) for estimating the quantile parameters $\delta(\tau)\equiv(\beta(\tau)',\lambda(\tau)')'$. Denote the generated regressors as $X(\gamma)=\left[X',(X-Z'\gamma)'\right]'$. We assume that $\mathbb{E}\left[\eta\eta'|Z\right]=\sigma^2I$ and $\mathbb{E}\left(ZZ'\right)$ is finite. The OLS estimator is asymptotically linear:

$$\sqrt{n}\left(\widehat{\gamma}-\gamma\right)=\sum_{i=1}^{n}\left[\mathbb{E}^{-1}\left[ZZ'\right]Z_i\eta_i\right]/\sqrt{n}+o_{\mathbb{P}}(1).$$

The asymptotic normality of the quantile parameters $\delta(\tau)$ follows directly from Theorem 1,

$$\sqrt{n}\left[\widehat{\delta}(\tau)-\delta(\tau)\right]\xrightarrow{d}\mathcal{N}(0,V(\tau)),$$

where

$$V(\tau)=H(\tau)^{-1}\left\{J(\tau)+D(\tau)\sigma^2\mathbb{E}^{-1}[ZZ']D(\tau)'+C(\tau)'D(\tau)+C(\tau)D(\tau)'\right\}H(\tau)^{-1}.$$

The terms of $V(\tau)$ are given by

$$H(\tau) = \mathbb{E}\left[\frac{X(\gamma)X(\gamma)'}{X(\gamma)'\delta^{(1)}(\tau)}\right], \quad J(\tau) = \tau(1-\tau)\mathbb{E}\left[X(\gamma)X(\gamma)'\right], \quad D(\tau) = -\mathbb{E}\left[\frac{X(\gamma)Z\lambda(\tau)}{X(\gamma)'\delta^{(1)}(\tau)}\right]$$
$$C(\tau) = \mathbb{E}\left[X(\gamma)\left(\mathbb{I}\left(Y \leq X(\gamma)'\delta(\tau)\right) - \tau\right)\eta' Z'\mathbb{E}^{-1}[ZZ']\right].$$

4.3 Box-Cox power transformation

The box-cox transformation parameter of (2.9) is estimated using the nonlinear IV (NIV) estimator of (2.10). The conditional quantile model for the generated dependent variable $Y(\widehat{\lambda})$ is assumed linear in parameters, which are estimated using the QR estimator of (2.11). Amemiya (1974) establishes the limiting behaviour of the NIV estimator. Assume that $\mathbb{E}\left[(Y(\lambda) - X'\beta)^2 WW'\right]$ is finite and Ω is full rank and finite.

Note that if β is a K-dimension vector, then the NIV estimator estimates $(K+1)$ parameters, denoted by $\theta = [\lambda, \beta']'$. Denote the $(K+1)$ order square matrix,

$$G = \mathbb{E}\left[W\frac{\partial Y(\lambda)}{\partial \lambda}, -WX'\right].$$

Then, the NIV estimator is asymptotically linear:

$$\sqrt{n}\left(\widehat{\theta} - \theta\right) = \sum_{i=1}^{n}\left[-\left(G'\Omega G\right)^{-1} G'\Omega W_i(Y_i(\lambda) - X_i'\beta)\right]/\sqrt{n} + o_{\mathbb{P}}(1).$$

The asymptotic variance of $\widehat{\lambda}$, denoted by $\mathcal{V}(\lambda)$, is the first term of the asymptotic variance-covariance matrix for $\widehat{\theta}$. Denoting $i_{11} = [1, \mathbf{0}_{K\times 1}]$, where $\mathbf{0}_{K\times 1}$ is a K-dimension row vector of zeros,

$$\mathcal{V}(\lambda) = i_{11}\left(\left(G'\Omega G\right)^{-1} G'\Omega\mathbb{E}\left[\left(Y(\lambda) - X'\beta\right)WW'\right]\Omega G\left(G'\Omega G\right)^{-1}\right)i_{11}'.$$

Asymptotic normality for the quantile estimates obtained from QR of $Y(\widehat{\lambda})$ on X follows directly from Theorem 1.

$$\sqrt{n}\left(\widehat{\beta}(\tau) - \beta(\tau)\right) \xrightarrow{d} \mathcal{N}\left(0, V(\tau)\right),$$

where

$$V(\tau) = H(\tau)^{-1}\left\{J(\tau) + D(\tau)\mathcal{V}(\lambda)D(\tau)' + C(\tau)'D(\tau) + C(\tau)D(\tau)'\right\}H(\tau)^{-1}.$$

The terms of $V(\tau)$ are given by

$$H(\tau) = \mathbb{E}\left[\frac{XX'}{X'\beta^{(1)}(\tau)}\right], \quad J(\tau) = \tau(1-\tau)\mathbb{E}\left[XX'\right], \quad D(\tau) = -\mathbb{E}\left[Xf(X'\beta|X,\lambda)\frac{\partial Y(\lambda)}{\partial\lambda}\right]$$
$$C(\tau) = \mathbb{E}\left[g(X)\left\{\int_0^\tau X'\beta(t)dt - \tau X'\beta(\tau)\right\}\right],$$

where $g(X) = X\left[1, \mathbf{0}_{K\times 1}\right]\left(-\left(G'\Omega G\right)^{-1} G'\Omega W\right)$.

5 GQR vs QR: Analysis for constant slope example

To compare the asymptotic variance of the GQR estimator with that of the standard QR estimator, we analyse the constant slope QR model of Section 2.1.1 with the following true model parameters: for the model in (2.3),

$$X_1 \sim \mathcal{U}[1,5] \quad X_2 \sim \mathcal{U}[3,10], \quad \beta_0(\tau) = e^\tau, \quad \beta_1(\tau) = \beta_1 = 1\ \forall\tau, \quad \beta_2(\tau) = 2\tau^2. \tag{5.1}$$

The two step GQR estimator, which estimates the constant slope parameter β_1 and the quantile parameters $(\beta_0(\tau), \beta_2(\tau))$ separately, has asymptotic variance given by (4.2). The asymptotic variance for the standard QR estimator, where all three coefficients are estimated together in a single step quantile regression of Y on X's, is given by

$$V(\tau)_{QR} = H(\tau)_{QR}^{-1} J(\tau)_{QR} H(\tau)_{QR}^{-1} \tag{5.2}$$

where denoting $X = [1, X_1, X_2]'$,

$$H(\tau)_{QR} = \mathbb{E}\left[\frac{XX'}{\beta_0^{(1)}(\tau) + \beta_1^{(1)}(\tau)X_1 + \beta_2^{(1)}(\tau)X_2}\right], \quad J(\tau)_{QR} = \tau(1-\tau)\mathbb{E}\left[XX'\right]$$

Remark. For the GQR estimator, if the covariates X_1 and X_2 are independent, as considered here, it holds that

(i) The covariance between first and second step estimates is zero: $C(\tau) = 0$.

(ii) The first step estimation has an effect on the second-step variance for the intercept, $\widehat{\beta}_0(\tau)$, but not for the slope parameter $\widehat{\beta}_2(\tau)$, as $H(\tau)^{-1}D(\tau)$ in (4.2) evaluates to $[-\mathbb{E}[X_1], 0]'$.

Proofs are straightforward using basic matrix algebra. The outline is presented in Appendix 3.

Asymptotic variance for $\widehat{\beta}_0(\cdot)$. Under the above remark, the asymptotic variance of $\widehat{\beta}_0(\cdot)$ for GQR is obtained using (4.2) as follows:

$$V(\tau)_{GQR}^{\widehat{\beta}_0(\cdot)} = [1,0]H(\tau)^{-1}J(\tau)H(\tau)^{-1}[1,0]' + \mathbb{E}^2[X_1]\mathcal{V}(\beta_1) \tag{5.3}$$

where $H(\tau), J(\tau)$ are given by (4.2). For the true model parameters and distribution considered here, this evaluates to

$$V(\tau)_{GQR}^{\widehat{\beta}_0(\cdot)} = \frac{\tau(1-\tau)}{(ac-b^2)^2}\left(c^2 - 2bc\mathbb{E}[X_2] + b^2\mathbb{E}[X_2^2]\right) + \mathbb{E}^2[X_1]\mathcal{V}(\beta_1) \tag{5.4}$$

where

$$\begin{aligned}
a &= \mathbb{E}\left[\frac{1}{\beta_0^{(1)}(\tau)+\beta_2^{(1)}(\tau)X_2}\right] = \frac{1}{28\tau}\ln\left(\frac{e^\tau+40\tau}{e^\tau+12\tau}\right),\\
b &= \mathbb{E}\left[\frac{X_2}{\beta_0^{(1)}(\tau)+\beta_2^{(1)}(\tau)X_2}\right] = \frac{1}{7\times 16\tau^2}\left(28\tau - e^\tau\ln\left(\frac{e^\tau+40\tau}{e^\tau+12\tau}\right)\right),\\
c &= \mathbb{E}\left[\frac{X_2^2}{\beta_0^{(1)}(\tau)+\beta_2^{(1)}(\tau)X_2}\right] = \frac{1}{448\tau^3}\left(728\tau^2 - 28\tau e^\tau + e^{2\tau}\ln\left(\frac{e^\tau+40\tau}{e^\tau+12\tau}\right)\right)\\
\mathbb{E}[X_2] &= 13/2, \quad \mathbb{E}[X_2^2] = 139/3, \quad \mathbb{E}[X_1] = 3.
\end{aligned}$$

The first step asymptotic variance $\mathcal{V}(\beta_1)$ is given by (4.1), and for the model parameters considered here, evaluates to

$$\begin{aligned}
\mathcal{V}(\beta_1) &= i_{22}\left(\mathbb{E}[XX']\right)i_{22}'\mathbb{E}\left[\varepsilon_0^2\right] + i_{22}\left(\mathbb{E}^{-1}[XX']\mathbb{E}[X_2^2XX']\mathbb{E}^{-1}[XX']\right)i_{22}'\mathbb{E}\left[\varepsilon_2^2\right]\\
&= \frac{3}{4}\mathbb{E}\left[\varepsilon_0^2\right] + \frac{139}{4}\mathbb{E}\left[\varepsilon_2^2\right] = \frac{3}{4}\times\left(\frac{1}{2}(e^2-1)-(e-1)^2\right) + \frac{139}{4}\times\left(\frac{4}{5}-\left(\frac{2}{3}\right)^2\right)
\end{aligned}$$

where $\varepsilon_i = \left(\beta_i(U) - \overline{\beta}_i\right),\ i = 0, 2.$

The asymptotic variance of $\widehat{\beta}_0(\cdot)$ for the standard QR is given by the first element of (5.2), which, for the model parameters and distribution assumed in this exercise, evaluates to

$$V(\tau)_{QR}^{\widehat{\beta}_0(\cdot)} = \frac{\tau(1-\tau)}{(ac-b^2)^2}\left(c^2 - 2bc\mathbb{E}[X_2] + b^2\mathbb{E}[X_2^2]\right) + \frac{\tau(1-\tau)(bf-dc)^2}{(ac-b^2)^2\,a^2\mathrm{Var}(X_1)} \tag{5.5}$$

where $\text{Var}(X_1) = 4/3$, (a, b, c) are as in (5.4) and

$$d = \mathbb{E}\left[\frac{X_1}{\beta_0^{(1)}(\tau) + \beta_2^{(1)}(\tau)X_2}\right] = \frac{3}{28\tau}\ln\left(\frac{e^\tau + 40\tau}{e^\tau + 12\tau}\right)$$

$$f = \mathbb{E}\left[\frac{X_1 X_2}{\beta_0^{(1)}(\tau) + \beta_2^{(1)}(\tau)X_2}\right] = \frac{3}{7 \times 16\tau^2}\left(28\tau - e^\tau \ln\left(\frac{e^\tau + 40\tau}{e^\tau + 12\tau}\right)\right).$$

Comparing (5.4) and (5.5), we find that the first part of the asymptotic covariance for both GQR and QR is a common quantile varying component. GQR has a constant additional component which depends on the first step asymptotic variance, while the additional part for QR is again quantile-dependent. The following graph plots this additional component for both GQR and QR.

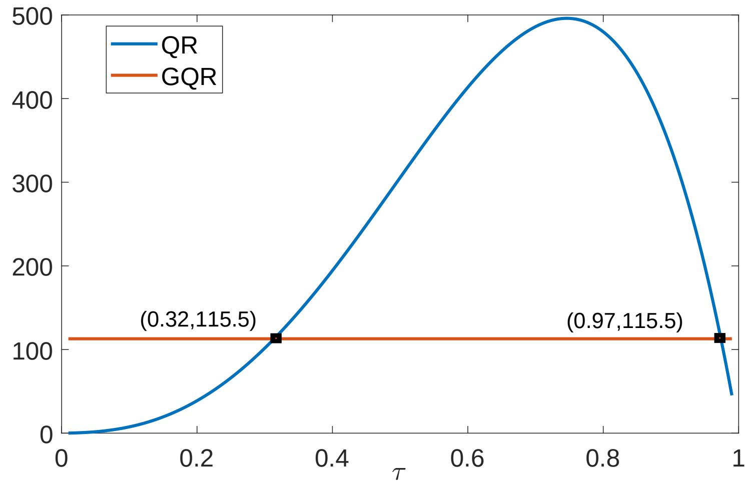

Figure 1: Additional covariance GQR vs QR

As can be seen from Figure 1, for the tails the additional variance part of QR is less than that of GQR, while the opposite is true for all other quantile levels. Hence, there isn't a clear efficiency gain of one method over the other - it depends on the quantile level. But, from our calculations for this example, it is clear that the GQR asymptotic covariance is less for most quantile levels while QR shows improvement for the tails, which is especially extreme for the right tail. Also, the GQR adds a constant contribution to the variance, while QR is very quantile dependent. For the simplicity of analysis here, we considered X_1 and X_2 to be independent. In real applications, this may not be true. But the relationship between the asymptotic covariances of GQR and QR remains similar, as we verify in the empirical application.

Asymptotic variance for $\widehat{\beta}_2(\cdot)$. Under the remark noted above, for independent X_1 and X_2, the asymptotic variance of $\widehat{\beta}_2(\cdot)$ is same for GQR and QR.

$$V(\tau)_{GQR}^{\widehat{\beta}_2(\cdot)} = V(\tau)_{QR}^{\widehat{\beta}_2(\cdot)} = [0,1]H(\tau)^{-1}J(\tau)H(\tau)^{-1}[0,1]' = [0,0,1]H(\tau)_{QR}^{-1}J(\tau)_{QR}H(\tau)_{QR}^{-1}[0,0,1]'$$

where $(H(\tau), J(\tau))$ and $(H(\tau)_{QR}, J(\tau)_{QR})$ are obtained from (4.2) and (5.2), respectively. For the true model parameters and distribution considered here, this evaluates to

$$V(\tau)^{\widehat{\beta}_2(\cdot)} = \frac{\tau(1-\tau)}{(ac-b^2)^2}\left(b^2 - 2ab\mathbb{E}[X_2] + a^2\mathbb{E}[X_2^2]\right) \tag{5.6}$$

where (a, b, c) are as in (5.4).

6 Monte Carlo Simulation

This section reports results of a simulation exercise based on the quantile regression with constant slope model described in Section 2.1.1, the asymptotic results of which are obtained in Section 4.1. The purpose of the simulation is to illustrate the performance of the proposed two-step estimator and validate the asymptotic normality result derived in Theorem 1. Besides bias-root mean squared error (RMSE) and coverage rate analysis of the GQR estimator, its performance is also compared with the standard quantile regression (QR) estimator where all parameters - both the constant and quantile varying ones - are estimated together by quantile regression. Finally, to see the impact of first step estimates on the overall variance, we compare the GQR estimator with the infeasible quantile regression (i-QR) estimator that uses the true value of the unknown parameter instead of its estimate.

Data generating process. In the quantile model of equation (2.3),

$$Q_Y(\tau|X) = \beta_0(\tau) + \beta_1(\tau)X_1 + \beta_2(\tau)X_2,$$

the true parameters are taken as in (5.1). Observations are generated as $Y_i = \beta_0(U_i) + \beta_1 X_{1i} + \beta_2(U_i)X_{2i}$, where (X_{1i}, X_{2i}) are uniform random variables as in (5.1), U_i is a $[0,1]$-uniform random variable and $i = 1, \cdots, n$. The simulation experiment is performed for the sample sizes of $n = 100$ and $n = 1000$. The number of simulation replications is set to 1000.

Estimation of quantile parameters. The estimation of quantile parameters is performed using the proposed GQR estimator, the i-QR estimator and the standard QR estimator. The GQR estimator uses the generated dependent variables $Y_i(\hat{\beta}_1) = (Y_i - \hat{\beta}_1 X_{1i})$, where

$\hat{\beta}_1$ is estimated from first step using OLS, in the quantile regression estimator of equation (2.5). The i-QR estimator uses the unknown dependent variable $Y_i^*(\beta_1) = Y_i - \beta_1 X_{1i}$ for quantile regression based estimation of the quantile parameters. The standard QR estimator estimates all coefficients together by quantile regression of Y on X's.

Estimation of asymptotic variance. We estimate asymptotic variance of the quantile parameters for the GQR, the i-QR, and the QR estimator, to compare the efficiency of the estimators and to validate the GQR asymptotic variance result obtained in Theorem 1. The estimation of asymptotic variance for quantile regression follows two common approaches: kernel based estimation following Powell (1991) or some form of bootstrap. We have used Buchinsky (1994)'s design matrix bootstrap, extensively used in empirical applications for quantile regression involving large samples, see, for instance, Buchinsky (1994) and Abrevaya (2002). Buchinsky (1995) and Koenker & Hallock (2001) compare various quantile regression variance estimators and conclude that with enough bootstrap replications, the method works satisfactorily. The bootstrapped standard errors are also used in finding confidence intervals of the GQR quantile estimates and comparing them with the nominal levels expected for normal distribution.

For $B = 1000$ bootstrap replications, each of size of m (drawn with replacement from the overall sample size of n), we get $(b = 1, \cdots, B)$ bootstrap quantile estimates at each quantile level. When $n = 1000$, the bootstrap sample size is $m = 300$, while for $n = 100$, we have $m = n$. This follows the so-called m out of n bootstrap technique which provides significant computational advantage when sample size is large. Following Buchinsky (1994), the sample covariance of these estimates, rescaled by (m/n), constitutes a valid estimator of the covariance matrix of the quantile regression estimator. Hence, the estimate for asymptotic covariance $V(\tau)$ with quantile parameters $\beta(\cdot)$ and the bootstrap estimates denoted by $\widehat{\beta}^b(\tau)$, $b = 1, \cdots, B$, is given by

$$\widehat{V}(\tau) = n\left(\frac{m}{n}\right)\frac{1}{B}\sum_{b=1}^{B}\left(\widehat{\beta}^b(\tau) - \widehat{\beta}^b_{\mathcal{A}}(\tau)\right)\left(\widehat{\beta}^b(\tau) - \widehat{\beta}^b_{\mathcal{A}}(\tau)\right)', \tag{6.1}$$

where $\widehat{\beta}^b_{\mathcal{A}}(\tau)$ is the average of the B bootstrap estimates. The choice of bootstrap replications and sample size are consistent with Buchinsky (1995) and Andrews & Buchinsky (2000). We estimate $\widehat{V}(\tau)$ from (6.1) for each of the 1000 simulations and report the average.

Results. The estimation results for the quantile parameters, along with comparisons of the GQR, standard QR and i-QR estimation methods, are presented in Tables 1-6. The first step least squares regression gives the mean of $\hat{\beta}_1$ as 1.007 (with average standard deviation

= 0.3953) for a sample size of 100, and 1.001 (with average standard deviation = 0.1242) for a sample size of 1000, respectively. The fact that OLS is unbiased is expected but the standard deviation is meaningful as it gives an idea of how much the first step impacts the overall variance.

Table 1: Bias and RMSE of $\hat{\beta}_0(\cdot)$ for $n = 100$ and 1000

		$n = 100$		$n = 1000$	
τ		Bias	RMSE	Bias	RMSE
0.2	GQR	0.0320	1.4076	−0.0004	0.4357
	QR	0.0328	0.9152	0.0135	0.2868
	i-QR	0.0108	0.6971	−0.0124	0.2256
0.4	GQR	0.1374	1.9558	−0.0126	0.6405
	QR	0.0125	2.0384	0.0272	0.6662
	i-QR	0.1546	1.5300	−0.0137	0.4960
0.6	GQR	0.1654	2.5208	−0.0486	0.8433
	QR	0.0470	2.8203	0.0409	0.9770
	i-QR	0.0922	2.2092	0.0252	0.7233
0.8	GQR	0.0581	2.7213	−0.0481	0.8440
	QR	−0.1241	3.0094	0.0129	1.0336
	i-QR	0.0412	2.4851	0.0006	0.7875

Table 1 shows the Bias-RMSE for $\widehat{\beta}_0(\cdot)$ using all three estimation methods, for varying n. It can be seen that all methods of estimation have low biases, and the RMSE falls with increasing sample size. But while all estimation procedures have similar biases, the RMSE with GQR is greater than that of QR for the first quantile, and the opposite is true for the rest of the quantiles, as expected from the analysis in Section 5. Also, the RMSE with GQR are greater than those of the i-QR method at each quantile, with substantial difference in some. This is as expected from theory: the asymptotic variance for i-QR is given by $H^{-1}JH^{-1}$ but when the first step is not known but estimated, the first step estimate's variance increases the overall variance (by $H^{-1}D\mathcal{V}(\beta_1)D'H^{-1}$, from (4.2), with $C(\tau) = 0$). Table 2 reports the Bias-RMSE results for the slope parameter $\widehat{\beta}_2(\cdot)$. The analysis in Section 5 shows that when X_1 and X_2 are independent, the variance of $\widehat{\beta}_2(\cdot)$ is unaffected by first step estimation so that it is the same for GQR, i-QR, as well as QR, methods. It can be seen in Table 2 that the bias and RMSE are similar for all three methods of estimation and the RMSE falls with increase in sample size.

Tables 3-4 compare the estimated asymptotic standard errors with their true values as a means of validating our asymptotic variance result. The true asymptotic variance of $\widehat{\beta}_0(\cdot)$

Table 2: Bias and RMSE of $\hat{\beta}_2(\cdot)$ for $n = 100$ and 1000

		$n = 100$		$n = 1000$	
τ		Bias	RMSE	Bias	RMSE
0.2	GQR	0.0148	0.1362	0.0017	0.0397
	QR	0.0053	0.1235	−0.0034	0.0375
	i-QR	0.0009	0.1242	−0.0003	0.0395
0.4	GQR	−0.0089	0.2756	0.0014	0.0918
	QR	−0.0028	0.2759	−0.0073	0.0863
	i-QR	−0.0257	0.2692	−0.0003	0.0892
0.6	GQR	−0.0371	0.3972	0.0066	0.1348
	QR	−0.0346	0.3948	−0.0177	0.1286
	i-QR	−0.0314	0.3897	−0.0097	0.1288
0.8	GQR	−0.0425	0.4400	0.0036	0.1371
	QR	−0.0452	0.4308	−0.0148	0.1433
	i-QR	−0.0486	0.4481	−0.0082	0.1408

for GQR and i-QR are calculated using (5.4), where for the latter $\mathcal{V}(\beta_1) = 0$, while that of QR is given by (5.5). The true asymptotic variance of $\widehat{\beta}_2(\cdot)$ for all methods is given by (5.6). The correctness of our asymptotic covariance result is verified by comparing $V(\tau)$ with its bootstrapped estimate as given by (6.1) (mean of $\widehat{V}(\tau)$ over the 1000 simulations is reported). It can be seen in Table 3 that the true asymptotic SE of $\widehat{\beta}_0(\cdot)$ is greater for GQR than QR for $\tau = 0.2$ and the trend changes for all other quantile levels, while it is always greater than that of i-QR, which are as predicted by theory and discussed above. Bootstrap estimation of asymptotic standard error works well even for small sample size of 100 (except for $\tau = 0.2$ using GQR) and the estimation accuracy for GQR improves with samples size (for QR and i-QR, the estimates are mostly similar for both sample sizes). Table 3 also reports the coefficient of variation (CoV) for $\widehat{V}(\tau)$, which is the ratio of the standard deviation to the mean of $\widehat{V}(\tau)$ over the 1000 simulations. CoV measures the precision in estimation of the asymptotic standard errors (or variability among the estimated values in each run of the simulation). Looking at CoV, it is interesting to note that for GQR the estimates of asymptotic SE have lesser variation across simulations relative to their mean values, and CoV is very similar across quantiles, than that of i-QR or QR. This suggests that the GQR asymptotic SE estimates are less dispersed around the mean than that of i-QR or QR. The CoV falls for all methods with sample size; for sample size of 1000, it is well within 10% for GQR and slightly greater than 10% for i-QR and QR. Table 4 confirms that variance of $\hat{\beta}_2(\cdot)$ is unaffected by the two step procedure, as QR, i-QR and GQR yield identical true values, and similar bootstrapped estimates as well as CoV. Also, in Table 3 and, to a lesser degree in Table 4, we find a slight overestimation of the GQR variance and underestimation of the QR one.

Table 3: Asymptotic standard error for $\sqrt{n}\hat{\beta}_0(\cdot)$, $B = 1000$, simulations = 1000

			$n = 100$		$n = 1000$	
τ		True	Estimated	CoV	Estimated	CoV
0.2	GQR	12.8272	15.2959	18.43%	13.8210	7.63%
	QR	9.5117	10.5633	33.22%	9.2487	13.66%
	i-QR	7.1905	7.8174	31.51%	7.0263	12.52%
0.4	GQR	19.1989	20.0756	18.03%	19.6835	7.80%
	QR	21.2165	20.6893	24.49%	20.6343	11.01%
	i-QR	15.9926	15.8358	24.46%	15.6959	10.73%
0.6	GQR	25.5869	25.7885	18.03%	25.8453	7.87%
	QR	30.9088	28.9673	22%	29.7102	10.25%
	i-QR	23.2778	22.5501	22.53%	22.8857	10.37%
0.8	GQR	27.2149	28.1044	19.08%	27.4708	8.55%
	QR	33.2814	32.3425	22.83%	32.5182	10.29%
	i-QR	25.0563	25.0403	22.82%	24.9103	10.3%

The true asymptotic standard error for GQR and QR are computed using (5.4) and (5.5), respectively, while for i-QR, the first step variance=0 in the formula for GQR. Mean over 1000 simulations of the bootstrapped asymptotic standard error estimate of (6.1) is reported. CoV denotes coefficient of variation and indicates the extent of variability in the estimates for each run of the simulation.

Tables 5-6 demonstrate that the 90%, 95% and 99% confidence intervals of the quantile estimates obtained from GQR are close to that in theory for normal approximation. For $\tau = \{0.2, 0.4, 0.6, 0.8\}$, t-stat of the quantile parameters is computed using bootstrapped standard errors and its absolute value is compared with the critical values for $(1-\alpha)$ confidence level of the normal approximation, $(1-\alpha) = 0.9, 0.95$ and 0.99, to find if the true quantile parameter is inside the corresponding confidence interval. Repeating the exercise for 1000 times, we find the percentage of times when the true parameter is inside the $(1-\alpha)$ confidence interval. We also report the coverage rate from QR and i-QR methods. The GQR variance overestimation as noted in Tables 3-4, especially for $n = 100$ and $\tau = 0.2$, is consistent with the high coverage rate corresponding to the 90% and 95% nominal confidence intervals for GQR (and QR) for the starting quantile when $n = 100$, but it improves for $n = 1000$. Overall, the empirical levels for confidence intervals are close to $(1-\alpha)$ and improves with increasing sample size, which suggests that the estimation procedure gives accurate central limit theorem based confidence intervals.

Table 4: Asymptotic standard error for $\sqrt{n}\hat{\beta}_2(\cdot)$, $B = 1000$, simulations $= 1000$

			$n = 100$		$n = 1000$	
τ		True	Estimated	CoV	Estimated	CoV
0.2	GQR	1.2450	1.6212	25.13%	1.2617	11.48%
	QR	1.2450	1.3722	33.06%	1.2084	12.56%
	i-QR	1.2450	1.3328	32.32%	1.2183	12.44%
0.4	GQR	2.8129	2.8139	23.43%	2.7845	10.16%
	QR	2.8129	2.7485	24.07%	2.7447	10.03%
	i-QR	2.8129	2.7656	23.82%	2.7697	10.51%
0.6	GQR	4.1156	4.0434	21.58%	4.0755	9.63%
	QR	4.1156	3.9354	21.50%	4.0237	9.37%
	i-QR	4.1156	4.0175	22.66%	4.0526	10.06
0.8	GQR	4.4389	4.4844	22.36%	4.3899	10.07%
	QR	4.4389	4.4241	21.47%	4.3901	9.92%
	i-QR	4.4389	4.4241	23.14%	4.4223	10.65%

The true asymptotic standard error for all methods is given by (5.6). The rest of explanation is as in Table 3.

Table 5: Confidence intervals: nominal vs. empirical, $n = 100$, simulations $= 1000$

		CI for $\beta_0(\cdot)$			CI for $\beta_2(\cdot)$		
Nominal level		0.90	0.95	0.99	0.90	0.95	0.99
Empirical level for $\tau = 0.2$	GQR	0.940	0.977	0.997	0.960	0.982	0.995
	QR	0.949	0.978	0.996	0.923	0.970	0.991
	i-QR	0.928	0.972	0.994	0.907	0.952	0.986
Empirical level for $\tau = 0.4$	GQR	0.897	0.951	0.991	0.888	0.942	0.980
	QR	0.899	0.948	0.987	0.888	0.940	0.987
	i-QR	0.885	0.940	0.986	0.878	0.933	0.981
Empirical level for $\tau = 0.6$	GQR	0.895	0.944	0.985	0.882	0.934	0.984
	QR	0.893	0.936	0.981	0.883	0.940	0.976
	i-QR	0.878	0.929	0.979	0.877	0.934	0.981
Empirical level for $\tau = 0.8$	GQR	0.900	0.940	0.989	0.895	0.942	0.981
	QR	0.905	0.950	0.984	0.890	0.942	0.986
	i-QR	0.867	0.928	0.978	0.875	0.932	0.975

Table 6: Confidence intervals: nominal vs. empirical, $n = 1000$, simulations $= 1000$

		CI for $\beta_0(\cdot)$			CI for $\beta_2(\cdot)$		
Nominal level		0.90	0.95	0.99	0.90	0.95	0.99
Empirical level for $\tau = 0.2$	GQR	0.897	0.952	0.992	0.906	0.954	0.989
	QR	0.903	0.953	0.989	0.911	0.947	0.981
	i-QR	0.884	0.940	0.985	0.893	0.939	0.983
Empirical level for $\tau = 0.4$	GQR	0.885	0.933	0.987	0.882	0.944	0.984
	QR	0.886	0.942	0.987	0.907	0.954	0.987
	i-QR	0.901	0.947	0.988	0.895	0.946	0.986
Empirical level for $\tau = 0.6$	GQR	0.886	0.952	0.985	0.891	0.937	0.981
	QR	0.874	0.945	0.980	0.887	0.935	0.984
	i-QR	0.893	0.946	0.985	0.893	0.942	0.986
Empirical level for $\tau = 0.8$	GQR	0.902	0.952	0.991	0.899	0.941	0.986
	QR	0.897	0.946	0.983	0.876	0.935	0.984
	i-QR	0.892	0.946	0.990	0.883	0.942	0.989

7 Empirical Application

The two step estimation procedure of Section 2.1.1 can be useful in estimating auction models as in the quantile regression approach of Gimenes (2017). In first price auctions, a quantile regression specification for the private value generates a quantile regression specification for the bid, see Gimenes & Guerre (2016). The linear regression approach of Haile, Hong & Shum (2003) for estimating first price auction models uses the 'homogenized bid' technique, which implies constant slope parameters in the bid quantile regression model. It is shown here that the two approaches can be combined, as in the example of Section 2.1.1. We apply the GQR estimator for the estimation of bid quantile specification containing both quantile-constant and quantile-dependent slope parameters. In the first step, following Haile et al. (2003), the constant slope parameter is estimated by regressing the bids on the observed covariates. This is then used to generate the dependent variable for the quantile regression for estimating the quantile parameters (as detailed in Section 2.1.1). The aim of our empirical exercise is to see how imposing a constant slope for a given set of variables can improve the estimation of the other slope functions.

We illustrate our proposed methodology using data from first price timber auctions conducted by the US Forest Services (USFS) covering the western half of US in the year 1979. This is the same data used by Lu & Perrigne (2008). The data consists of 214 first price auctions with 2 bidders, and the covariates listed are appraisal value and timber volume (in log).

Bid homogenization. Figure 2 shows the bid quantile parameter estimates obtained from quantile regression of bids on the covariates along with the corresponding linear regression estimates. Figure 3 plots the 95% confidence interval of the difference of the QR and OLS estimates for each quantile. If the interval contains zero, the corresponding parameter is likely to be constant across quantiles, while zero lying outside the confidence interval suggests potential misspecification of bid homogenization. Intercept and appraisal value quantile slope coefficients seem to satisfy the assumption of constancy across quantiles, except for higher quantiles of above 0.95 for appraisal value. However, the volume quantile parameter does not seem to be constant.

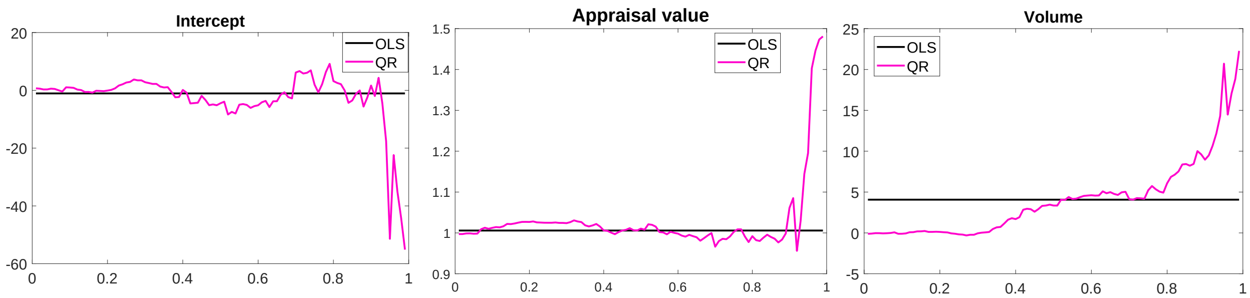

Figure 2: Bid quantile parameter estimates

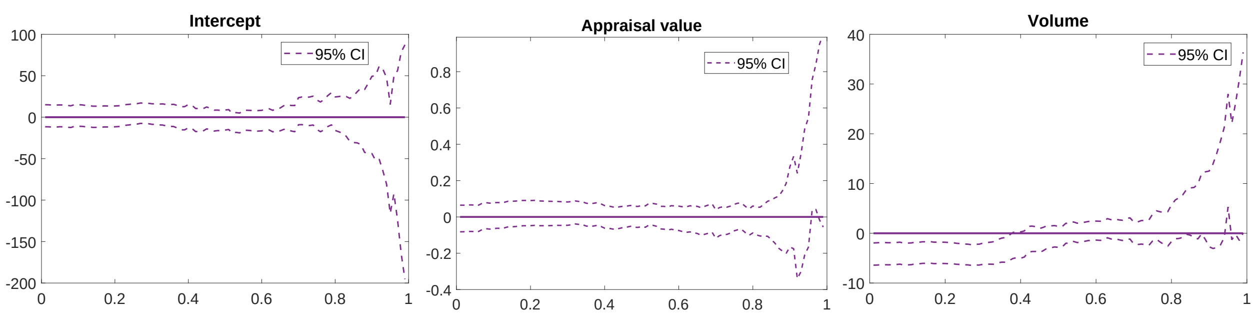

Figure 3: 95% CI for difference of QR and regression

Bid quantile estimation using GQR. The GQR estimator involves constrained estimation assuming the intercept and appraisal value slope to be constant across quantiles, while the volume parameter is considered to be varying with quantile levels. Table 7 reports the result of linear regression of bids on the covariates. The first step estimates constitute the intercept and appraisal value slope regression estimates, while the quantile estimates for slope of volume is obtained through quantile regression of the generated dependent variable $(bids_i - (-1.07) - 1.01 \times appraisal\ value_i)$ on $volume_i$. The second step GQR bid quantile estimate for slope of volume is shown in Figure 4. For comparison purpose, we also plot the results of unconstrained estimation of quantile parameters of volume. Table 8 also reports the bootstrapped standard error (SE) of the constrained and the unconstrained estimator obtained from 10,000 bootstrap replications.

Table 7: First step - bid regression

Intercept	Appraisal value	Volume	R^2
-1.07	1.01	4.07	0.77
(6.72)	(0.04)	(1.12)	

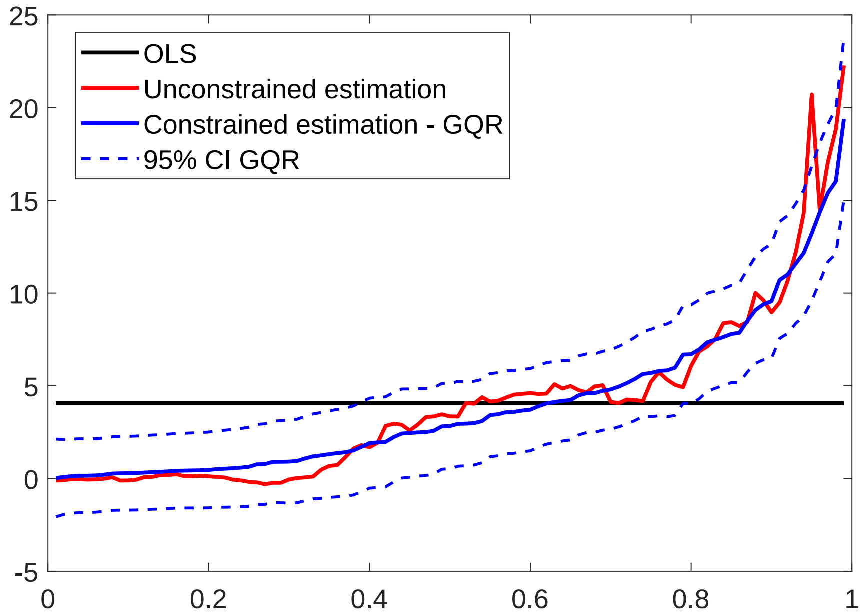

Figure 4: Second step - Bid QR estimates for volume

As can be seen in Figure 4 and Table 8, the GQR slope estimate is more regular than that of the unconstrained estimation; the GQR estimates increase with quantile level, which is consistent with an increasing bid conditional quantile function. The SE pattern observed in Table 8 is as expected from the analysis in Section 5, although the covariates are no longer independent: SE for the constrained estimator are similar across quantiles and lesser than that for the unconstrained case except for the first three quantile levels. The SE obtained for the unconstrained estimator varies quite a lot across quantiles and is quite high for the higher quantiles, which are particularly important for auction models as winners reside here. An intuitive explanation for the SE pattern observed here is as follows. The asymptotic variance of the unconstrained estimator will have the form given by (5.2): in the tails, while $\tau(1-\tau)$ tends to make the quantile estimate more precise, the derivative of

Table 8: Bootstrapped SE for constrained vs unconstrained quantile estimation of volume

	Constrained (GQR)		Unconstrained	
τ	Estimate	SE	Estimate	SE
0.1	0.2837	1.0457	−0.1023	0.3050
0.2	0.4631	1.0735	0.1219	0.3492
0.3	0.9131	1.1677	−0.0491	0.8013
0.4	1.9080	1.2727	1.6894	1.5634
0.5	2.8312	1.2024	3.3503	1.3285
0.6	3.7115	1.1624	4.6088	1.2478
0.7	4.8065	1.1281	4.1329	1.4609
0.8	6.7040	1.3835	6.0770	2.5452
0.9	9.5601	1.6093	8.9608	4.5484

the quantile slope parameter has an opposite effect. In higher quantiles, as is typical with quantile regression, the latter effect dominates making the quantile estimates in that region less precise. The asymptotic variance of the GQR estimator will have the form of (4.2): there will be a constant effect of the first step variance at each quantile level, but in addition to the corresponding $H^{-1}JH^{-1}$ term which increases with τ for increasing slope parameter, there is a negative quantile effect due to the covariance term being negative in τ for an increasing slope parameter. So, the net quantile-dependent effect is reduced. Hence, at lower quantile levels, the SE of the GQR estimator is greater than that of the unconstrained one because of the constant contribution of first step variance. But at higher quantiles, SE of the unconstrained estimator is much greater. The exact comparison of asymptotic variance will, however, depend on the model specifics.

In general, the unconstrained quantile regression involves fitting the model at each quantile level, for estimating both the constant and the quantile-dependent model parameters, and thus loses out on the information that some covariate effects are common across quantiles. The GQR estimator utilizes the commonality information and improves upon efficiency, except at the extreme tails. It is well noted in literature that for estimating quantile models that have some common covariate effects, efficiency gain can be achieved by aggregating the information across multiple quantiles, as in the combined quantile regression approach of Zou & Yuan (2008).

8 Conclusion

This paper presents a two step estimation method for estimating quantile regression models with generated covariates and/or dependent variable. The asymptotic properties of this generated quantile regression (GQR) estimator is systematically studied using Bahadur expansion and the asymptotic normality result is obtained. The results are verified using simulation and an application based on auctions is carried out. We mention some further areas of application. A key technical contribution of the paper is to provide Bahadur expansion which holds uniformly with respect to first step parameter and quantile levels, which can be utilised for developing specification tests (like those developed in Koenker & Machado (1999) and Koenker & Xiao (2002)) as well as to obtain functional central limit theorem for the two step quantile regression estimator. A further work is to prove the validity of bootstrap for the estimation of asymptotic variance and develop the theory for bootstrapped confidence interval.

A slightly different problem that can be studied using techniques developed here relates to quantile specifications where a first step estimation impacts the quantile level for the second stage quantile regression. Such specifications arise in Arellano & Bonhomme (2017)'s method of quantile regression with "rotated" check function to correct for sample selection in quantile regression models. A more challenging problem open for future research is to relax the assumption of $\sqrt{n}$-consistency in both the steps of estimation and consider models where they may be different, like in quantile regression models for panel data where the first step within estimator is usually $\sqrt{n}$-consistent and the quantile estimator is $\sqrt{nT}$-consistent.

Appendix 1. Proof section

Notations. The notation $\asymp$ is defined as: sequences $\{x_n\}$ and $\{y_n\}$ satisfy $x_n \asymp y_n$ if $|x_n|/C \leq |y_n| \leq C|x_n|$, for some $C > 0$ and n large enough. $||\cdot||$ is the Euclidean norm. The largest eigenvalue in absolute value for a symmetric matrix A is $||A|| = \sup_{u \in \mathcal{B}(0,1)} ||Au|| = \sup_{u \in \mathcal{B}(0,1)} |u'Au|$. Also, for any matrix or vector B, $||AB|| \leq ||A|| \; ||B||$. We denote $||f(\cdot|\cdot)||_\infty = \sup_{y,x} |f(y|X)|$. And the notation $\succ$ denotes that, for two symmetric matrices A_1, A_2, $A_1 \succ A_2$ if and only if $A_1 - A_2$ is a positive definite symmetric matrix.

Define

$$Q(\beta; \tau, \theta) = \mathbb{E}\left[\rho_\tau\left(Y(\theta) - X'(\theta)\beta\right)\right] - \mathbb{E}\left[\rho_\tau\left(Y(\theta)\right)\right].$$

As $\rho_\tau(\cdot)$ is almost everywhere differentiable with bounded derivatives, $\beta \mapsto Q(\beta; \tau, \theta)$ is differentiable with first derivative

$$\begin{aligned} Q^{(1)}(\beta; \tau, \theta) &= \mathbb{E}\left[\left\{\mathbb{I}\left(Y(\theta) \leq X'(\theta)\beta\right) - \tau\right\} X(\theta)\right] \\ &= \mathbb{E}\left[\left\{F\left(X'(\theta)\beta | X, \theta\right) - \tau\right\} X(\theta)\right]. \end{aligned}$$

Hence $Q(\cdot; \tau, \theta)$ is twice continuously differentiable with respect to β, with a second derivative

$$\begin{aligned} Q^{(2)}(\beta; \tau, \theta) &= \mathbb{E}\left[f\left(X'(\theta)\beta | X, \theta\right) X(\theta) X'(\theta)\right] \\ &= \int f\left(x'\beta | x, \theta\right) x x' f_X(x|\theta)\, dx. \end{aligned}$$

Let $B(\theta)$ be the set of β' such that $0 < F\left(x'(\theta)\beta | x, \theta\right) < 1$ for some inner x of $\mathcal{X}(\theta)$,

$$B(\theta) = \left\{\beta; \text{there is an inner } x \text{ of } \mathcal{X}(\theta) \text{ such that } \underline{y}(\theta|x) < x'\beta < \overline{y}(\theta|x)\right\}$$

where $\underline{y}(\theta|x) = F^{-1}(0|x)$ and $\overline{y}(\theta|x) = F^{-1}(1|x)$. The next Lemma describes some key properties of $Q^{(2)}(\beta; \tau, \theta)$ and $Q(\beta; \tau, \theta)$. Note that Proposition 1 follows from Lemma 2-(ii).

Lemma 2 *Under Assumption 3 it holds*

(i) $Q^{(2)}(\beta; \tau, \theta)$ *is continuous with respect to its three arguments, with*

$$\left\|Q^{(2)}(\beta_1; \tau, \theta) - Q^{(2)}(\beta_0; \tau, \theta)\right\| \leq C\left\|\beta_1 - \beta_0\right\|$$

for all β_0 *and* β_1, $\theta \in \Theta$ *and* $\tau \in [0, 1]$.

(ii) $Q^{(2)}(\beta;\tau,\theta)$ *is strictly positive for all* $\beta \in \mathrm{B}(\theta)$, $\theta \in \Theta$ *and* $\tau \in [0,1]$.

(iii) For $\theta \in \Theta$ *and* $\tau \in (0,1)$, $Q(\beta;\tau,\theta)$ *has a unique minimizer* $\beta(\tau;\theta)$ *which is continuously differentiable in* θ *and* τ *with*

$$\begin{aligned}\frac{\partial \beta(\tau;\theta)}{\partial \theta'} &= H(\tau;\theta)^{-1} D(\tau;\theta), \\ \frac{\partial \beta(\tau;\theta)}{\partial \tau} &= H(\tau;\theta)^{-1} \mathbb{E}[X(\theta)],\end{aligned}$$

where $H(\tau;\theta)$ *and* $D(\tau;\theta)$ *are as in Proposition 1.*

Proof of Lemma 2. (i) directly follows from Assumption 3 and the Lebesgue Dominated Convergence Theorem. For (ii), Assumption 3 gives that, for each β in $B(\theta)$, there is an open subset $\mathcal{O} = \mathcal{O}_{\beta,\theta}$ of $\mathcal{X}(\theta)$ such that

$$Q^{(2)}(\beta;\tau,\theta) \succeq \int_{\mathcal{O}} xx'dx.$$

Hence, $H(\tau;\theta) = Q^{(2)}(\beta(\tau;\theta);\tau,\theta)$ has an inverse. For (iii), observe that $Q(\beta;\tau,\theta)$ is bounded away from $-\infty$, so that it has local minimizers which must satisfy the first order condition

$$0 = Q^{(1)}(\beta;\tau,\theta) = \mathbb{E}\left[\{F(X'(\theta)\beta|X,\theta) - \tau\} X(\theta)\right]. \tag{A1.1}$$

Hence these minimizers must lie in $B(\theta)$ as outside this set it holds $F(X'(\theta)\beta|X,\theta) = 1$ a.s, or $F(X'(\theta)\beta|X,\theta) = 0$ a.s. Now, if there are two such local minimizers $\beta_0(\tau;\theta)$ and $\beta_1(\tau;\theta)$, convexity implies that all $\beta_\pi(\tau;\theta) = (1-\pi)\beta_0(\tau;\theta) + \pi\beta_1(\tau;\theta)$, $0 \le \pi \le 1$, must be global minimizers, contradicting that $Q^{(2)}(\beta_\pi(\tau;\theta);\tau,\theta)$ is strictly positive as $Q^{(1)}(\beta_\pi(\tau;\theta);\tau,\theta) = 0$ for all π in $[0,1]$. The rest of (iii) follows from (i), (ii) and the Implicit Function Theorem.□

Proof of Proposition 1. follows from Lemma 2-(iii). □

Proof of Proposition 2-(i). This proof conducts a uniform order study of the Bahadur error term (3.4). Define the following

$$\mathcal{L}_n(\gamma,\tau;\theta) = \sum_{i=1}^{n} \left\{ \rho_\tau\left(Y_i(\theta) - X_i(\theta)'\left(\frac{\gamma}{\sqrt{n}} + \beta(\tau;\theta)\right)\right) - \rho_\tau\left(Y_i(\theta) - X_i(\theta)'\beta(\tau;\theta)\right)\right\},$$

such that

$$\sqrt{n}\left(\widehat{\beta}(\tau;\theta) - \beta(\tau;\theta)\right) = \arg\min_{\gamma} \mathcal{L}_n(\gamma,\tau;\theta).$$

In what follows, we write

$$\widehat{\alpha}(\tau;\theta) \equiv -H^{-1}(\tau;\theta)\,\widehat{S}(\tau;\theta) \tag{A1.2}$$

$$\widehat{S}(\tau;\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i(\tau;\theta). \tag{A1.3}$$

It follows from equation (3.3) that

$$\widehat{\mathcal{E}}(\tau;\theta) = \arg\min_{\epsilon} \mathbb{L}_n(\widehat{\alpha}(\tau;\theta),\epsilon,\tau;\theta), \text{ where}$$
$$\mathbb{L}_n(\gamma,\epsilon,\tau;\theta) = \mathcal{L}_n(\gamma+\epsilon,\tau;\theta) - \mathcal{L}_n(\gamma,\tau;\theta) \tag{A1.4}$$
$$.$$

Consider the following decomposition of $\mathbb{L}_n(\gamma,\epsilon,\tau;\theta)$.

$$\begin{aligned}\mathbb{L}_n(\gamma,\epsilon,\tau;\theta) &= \mathbb{L}_n^0(\gamma,\epsilon,\tau;\theta) + \mathbb{R}_n(\gamma,\epsilon,\tau;\theta), \text{ where}\\ \mathbb{L}_n^0(\gamma,\epsilon,\tau;\theta) &= \widehat{S}(\tau;\theta)'(\gamma+\epsilon) + \frac{1}{2}(\gamma+\epsilon)'H(\tau;\theta)(\gamma+\epsilon) - \widehat{S}(\tau;\theta)'\gamma - \frac{1}{2}\gamma' H(\tau;\theta)\gamma \\ &= \widehat{S}(\tau;\theta)'\epsilon + \frac{1}{2}\epsilon' H(\tau;\theta)(\epsilon+2\gamma).\end{aligned} \tag{A1.5}$$

$\mathbb{L}_n^0(\gamma,\epsilon,\tau;\theta)$ is the quadratic approximation of $\mathbb{L}_n(\gamma,\epsilon,\tau;\theta)$ and $\mathbb{R}_n(\gamma,\epsilon,\tau;\theta)$ is the remainder term. A uniform order for $\widehat{\mathcal{E}}(\tau;\theta)$ relies on a uniform order study for the remainder term $\mathbb{R}_n(\gamma,\epsilon,\tau;\theta)$, using concepts of maximal inequality under bracketing conditions given in Massart (2007), and on linearization techniques to study $\widehat{\mathcal{E}}(\tau;\theta)$ given in Hjort & Pollard (2011). The remainder term is $\mathbb{R}_n(\gamma,\epsilon,\tau;\theta) = \mathbb{L}_n(\gamma,\epsilon,\tau;\theta) - \mathbb{L}_n^0(\gamma,\epsilon,\tau;\theta) = \sum_{i=1}^{n} \mathbb{R}_i(\gamma,\epsilon,\tau;\theta)$, where

$$\begin{aligned}\mathbb{R}_i(\gamma,\epsilon,\tau;\theta) = &\left\{\rho_\tau\left(Y_i(\theta) - X_i(\theta)'\left(\frac{\gamma+\epsilon}{\sqrt{n}} + \beta(\tau;\theta)\right)\right) - \rho_\tau\left(Y_i(\theta) - X_i(\theta)'\left(\frac{\gamma}{\sqrt{n}} + \beta(\tau;\theta)\right)\right)\right\}\\ &- \frac{s_i(\tau;\theta)'}{\sqrt{n}}\epsilon - \frac{1}{2}\epsilon'\frac{H(\tau;\theta)}{n}(\epsilon+2\gamma).\end{aligned} \tag{A1.6}$$

Define also

$$R_i(\gamma,\epsilon,\tau;\theta) = \mathbb{R}_i(\gamma,\epsilon,\tau;\theta) + \frac{1}{2}\epsilon'\frac{H(\tau;\theta)}{n}(\epsilon+2\gamma), \tag{A1.7}$$

$$\mathbb{R}_i^1(\gamma,\epsilon,\tau;\theta) = R_i(\gamma,\epsilon,\tau;\theta) - \mathbb{E}[R_i(\gamma,\epsilon,\tau;\theta)|X_i(\theta)], \tag{A1.8}$$

$$\mathbb{R}_i^2(\gamma,\epsilon,\tau;\theta) = \mathbb{E}[R_i(\gamma,\epsilon,\tau;\theta)|X_i(\theta)] - \frac{1}{2}\epsilon'\frac{H(\tau;\theta)}{n}(\epsilon+2\gamma), \tag{A1.9}$$

such that

$$\begin{aligned}&\mathbb{R}_n(\gamma,\epsilon,\tau;\theta)=\mathbb{R}_n^1(\gamma,\epsilon,\tau;\theta)+\mathbb{R}_n^2(\gamma,\epsilon,\tau;\theta),\ \text{with},\\&\mathbb{R}_n^j(\gamma,\epsilon,\tau;\theta)=\sum_{i=1}^n\mathbb{R}_i^j(\gamma,\epsilon,\tau;\theta),\ j=1,2.\end{aligned}\tag{A1.10}$$

We now present some intermediary results in Lemma 3, 4 and 5, on which the proof depends.

Lemma 3 *Under Assumption 3, for real numbers* t_γ,$t_\epsilon>0$ *with* $t_\gamma\asymp\log^{1/2}n$, $t_\gamma\geq1$, $t_\epsilon=\left(t\log^{3/4}n\right)/n^{1/4}$ *for some* $t>0$, *such that* $(t_\gamma+t_\epsilon)^{1/2}/t_\epsilon\leq O\left(n^{1/4}/\log^{1/2}n\right)$, *for large* n,

$$\mathbb{E}\left[\sup_{(\gamma,\epsilon,\tau;\theta)\in\mathcal{B}(0,t_\gamma)\times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta}\left|\mathbb{R}_n^1(\gamma,\epsilon,\tau;\theta)\right|\right]\leq C\frac{\log^{1/2}n}{n^{1/4}}t_\epsilon(t_\gamma+t_\epsilon)^{1/2}.$$

Lemma 4 *Under Assumption 3, for real numbers* t_γ,$t_\epsilon>0$ *defined as in Lemma 3, such that* $t_\gamma/t_\epsilon=O\left(n/\log^{1/2}n\right)$, *for large* n,

$$\mathbb{E}\left[\sup_{(\gamma,\epsilon,\tau;\theta)\in\mathcal{B}(0,t_\gamma)\times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta}\left|\mathbb{R}_n^2(\gamma,\epsilon,\tau;\theta)\right|\right]\leq C\frac{t_\epsilon(t_\gamma+t_\epsilon)^2}{n^{1/2}}.$$

Lemma 5 *Under Assumption 3,*

$$\sup_{(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\times\Theta}\left|\left|\widehat{S}(\tau;\theta)\right|\right|=O_{\mathbb{P}}(\log^{1/2}n).$$

Proofs of Lemma 3, 4 and 5 are provided in Appendix 2. In what follows,

$$t_n=t\frac{\log^{3/4}n}{n^{1/4}},\ t>0\ ,$$

and since $(\log n)/n=o(1)$, $t_n=o\left(\log^{1/2}n\right)$. t_n plays the role of t_ϵ in the Lemmas, while t_γ is chosen such that $t_\gamma\asymp\log^{1/2}n$. Hence,

$$\begin{aligned}\frac{(t_\gamma+t_\epsilon)^{1/2}}{t_\epsilon}&\asymp\frac{n^{1/4}\log^{1/4}n}{t\log^{3/4}n}=\frac{1}{t}O\left(\frac{n^{1/4}}{\log^{1/2}n}\right)\\\frac{t_\gamma}{t_\epsilon}&\leq C\frac{n^{1/4}\log^{1/2}n}{t\log^{3/4}n}\leq C\frac{n^{1/2}}{\log^{1/2}n}\leq C\frac{n^{1/2}n^{1/2}}{\log^{1/2}n}=O\left(\frac{n}{\log^{1/2}n}\right),\end{aligned}$$

for large n. These choices for t_γ and t_ϵ satisfy the requirements for the Lemmas. Lemma 2-(ii), which proves existence of H^{-1} for all $\tau\in[\underline{\tau},\overline{\tau}]$ and $\theta\in\Theta$, implies that $\widehat{\alpha}(\tau;\theta)$ is well defined with a probability tending to 1. Lemma 5 implies

$$\sup_{(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\times\Theta} ||\widehat{\alpha}(\tau;\theta)|| = O_{\mathbb{P}}\left(\log^{1/2} n\right). \tag{A1.11}$$

Consider $\xi > 0$ arbitrarily small. Then there exists a C_ξ such that, for large n and some $\underline{\varphi} > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{(\epsilon,\tau,\theta)\in\mathcal{B}(0,t_n)\times[\underline{\tau},\overline{\tau}]\times\Theta} |\mathbb{R}_n(\widehat{\alpha}(\tau;\theta),\epsilon,\tau;\theta)| \geq \frac{\underline{\varphi}t_n^2}{4}\right) \\
&\leq \mathbb{P}\left(\sup_{(\epsilon,\tau,\theta)\in\mathcal{B}(0,t_n)\times[\underline{\tau},\overline{\tau}]\times\Theta} |\mathbb{R}_n(\widehat{\alpha}(\tau;\theta),\epsilon,\tau;\theta)| \geq \frac{\underline{\varphi}t_n^2}{4}, \sup_{\tau,\theta\in[\underline{\tau},\overline{\tau}]\times\Theta} ||\widehat{\alpha}(\tau;\theta)|| \leq C_\xi \log^{1/2} n\right) \\
&\quad + \mathbb{P}\left(\sup_{\tau,\theta\in[\underline{\tau},\overline{\tau}]\times\Theta} ||\widehat{\alpha}(\tau;\theta)|| > C_\xi \log^{1/2} n\right) \\
&\leq \mathbb{P}\left(\sup_{(\gamma,\epsilon,\tau,\theta)\in\mathcal{B}(0,C_\xi\log^{1/2} n)\times\mathcal{B}(0,t_n)\times[\underline{\tau},\overline{\tau}]\times\Theta} |\mathbb{R}_n(\gamma,\epsilon,\tau;\theta)| \geq \frac{\underline{\varphi}t_n^2}{4}\right) + \xi.
\end{aligned}$$

Since $\mathbb{R}_n = \mathbb{R}_n^1 + \mathbb{R}_n^2$ from (A1.10), Lemmas 3 and 4, and Markov inequality give

$$\begin{aligned}
&\mathbb{P}\left(\sup_{(\gamma,\epsilon,\tau,\theta)\in\mathcal{B}(0,C_\xi\log^{1/2} n)\times\mathcal{B}(0,t_n)\times[\underline{\tau},\overline{\tau}]\times\Theta} |\mathbb{R}_n(\gamma,\epsilon,\tau;\theta)| \geq \frac{\underline{\varphi}t_n^2}{4}\right) \\
&\leq \frac{C}{t_n^2}\left(\mathbb{E}\left[\sup_{\substack{(\gamma,\epsilon,\tau,\theta)\in\mathcal{B}(0,C_\xi\log^{1/2} n)\\ \times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta}} \left|\mathbb{R}_n^1(\gamma,\epsilon,\tau;\theta)\right|\right] + \mathbb{E}\left[\sup_{\substack{(\gamma,\epsilon,\tau,\theta)\in\mathcal{B}(0,C_\xi\log^{1/2} n)\\ \times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta}} \left|\mathbb{R}_n^2(\gamma,\epsilon,\tau;\theta)\right|\right]\right) \\
&\leq \frac{C}{t_n^2}\left(\frac{t_n\left(C_\xi\log^{1/2} n + t_n\right)^{1/2}\log^{1/2} n}{n^{1/4}} + \frac{t_n\left(C_\xi\log^{1/2} n + t_n\right)^2}{n^{1/2}}\right) \\
&= \frac{C}{t_n}\frac{\log^{3/4} n}{n^{1/4}}\left(\left(C_\xi + \frac{t_n}{\log^{1/2} n}\right)^{1/2} + \left(\frac{\log n}{n}\right)^{1/4}\left(C_\xi + \frac{t_n}{\log^{1/2} n}\right)^2\right).
\end{aligned}$$

Using $t_n = \left(t\log^{3/4} n\right)/n^{1/4}$ and since $(\log n)/n = o(1)$, we get

$$\lim_{n\to\infty}\mathbb{P}\left(\sup_{(\epsilon,\tau,\theta)\in\mathcal{B}(0,t_n)\times[\underline{\tau},\overline{\tau}]\times\Theta} |\mathbb{R}_n(\widehat{\alpha}(\tau;\theta),\epsilon,\tau;\theta)| \geq \frac{\underline{\varphi}t_n^2}{4}\right) = \xi + O\left(\frac{{C_\xi}^{1/2}}{t}\right). \tag{A1.12}$$

We now find a uniform order for $\widehat{\mathcal{E}}(\tau;\theta)$. Consider $\mathcal{T}_n \geq t_n$ and $\epsilon = \mathcal{T}_n e$, $||e|| = 1$ so that $||\epsilon|| \geq t_n$. Since $\rho_\tau(\cdot)$ is convex, from the definition in (A1.4), $\mathbb{L}_n\left(\beta(\tau;\theta), \epsilon, \tau;\theta\right)$ is also convex. Recall that from (A1.4) and (A1.5), $\mathbb{L}_n\left(\beta(\tau;\theta), 0, \tau;\theta\right) = 0$ and $\mathbb{L}_n = \mathbb{L}_n^0 + \mathbb{R}_n$. Then, using convexity property, we have

$$
\begin{aligned}
\frac{t_n}{\mathcal{T}_n}\mathbb{L}_n\left(\widehat{\alpha}(\tau;\theta), \epsilon, \tau;\theta\right) &= \frac{t_n}{\mathcal{T}_n}\mathbb{L}_n\left(\widehat{\alpha}(\tau;\theta), \epsilon, \tau;\theta\right) + \left(1 - \frac{t_n}{\mathcal{T}_n}\right)\mathbb{L}_n\left(\widehat{\alpha}(\tau;\theta), 0, \tau;\theta\right) \\
&\geq \mathbb{L}_n\left(\widehat{\alpha}(\tau;\theta), \frac{t_n \epsilon}{\mathcal{T}_n}, \tau;\theta\right) = \mathbb{L}_n\left(\widehat{\alpha}(\tau;\theta), t_n e, \tau;\theta\right) \\
&= \mathbb{L}_n^0\left(\widehat{\alpha}(\tau;\theta), t_n e, \tau;\theta\right) + \mathbb{R}_n\left(\widehat{\alpha}(\tau;\theta), t_n e, \tau;\theta\right).
\end{aligned}
$$

Since from (A1.4), $\widehat{\mathcal{E}}(\tau;\theta) = \arg\min_\epsilon \mathbb{L}_n\left(\widehat{\alpha}(\tau;\theta), \epsilon, \tau;\theta\right)$, we have

$$
\begin{aligned}
\left\{\left|\left|\widehat{\mathcal{E}}(\tau;\theta)\right|\right| \geq t_n\right\} &\subset \left\{\inf_{\epsilon;||\epsilon||\geq t_n} \mathbb{L}_n\left(\widehat{\alpha}(\tau;\theta), \epsilon, \tau;\theta\right) \leq \inf_{\epsilon;||\epsilon||< t_n} \mathbb{L}_n\left(\widehat{\alpha}(\tau;\theta), \epsilon, \tau;\theta\right)\right\} \\
&\subset \left\{\inf_{\epsilon;||\epsilon||\geq t_n} \mathbb{L}_n\left(\widehat{\alpha}(\tau;\theta), \epsilon, \tau;\theta\right) \leq \mathbb{L}_n\left(\widehat{\alpha}(\tau;\theta), 0, \tau;\theta\right) = 0\right\} \\
&\subset \left\{\inf_{e;||e||=1} \left[\mathbb{L}_n^0\left(\widehat{\alpha}(\tau;\theta), t_n e, \tau;\theta\right) + \mathbb{R}_n\left(\widehat{\alpha}(\tau;\theta), t_n e, \tau;\theta\right)\right] \leq 0\right\} \\
&\subset \left\{\inf_{||\epsilon||=t_n} \mathbb{L}_n^0\left(\widehat{\alpha}(\tau;\theta), \epsilon, \tau;\theta\right) - \sup_{||\epsilon||=t_n}\left|\mathbb{R}_n\left(\widehat{\delta}(\tau;\theta), \epsilon, \tau;\theta\right)\right| \leq 0\right\}.
\end{aligned}
$$

Then, it follows for supremum of $\left|\left|\widehat{\mathcal{E}}(\tau;\theta)\right|\right|$ that

$$
\begin{aligned}
&\left\{\sup_{(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\times\Theta}\left|\left|\widehat{\mathcal{E}}(\tau;\theta)\right|\right| \geq t_n\right\} = \bigcup_{(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\times\Theta}\left\{\left|\left|\widehat{\mathcal{E}}(\tau;\theta)\right|\right| \geq t_n\right\} \\
&\subset \bigcup_{(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\times\Theta}\left\{\inf_{||\epsilon||=t_n} \mathbb{L}_n^0\left(\widehat{\alpha}(\tau;\theta), \epsilon, \tau;\theta\right) - \sup_{||\epsilon||=t_n}\left|\mathbb{R}_n\left(\widehat{\alpha}(\tau;\theta), \epsilon, \tau;\theta\right)\right| \leq 0\right\} \\
&\subset \left\{\inf_{(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\times\Theta}\inf_{||\epsilon||=t_n} \mathbb{L}_n^0\left(\widehat{\alpha}(\tau;\theta), \epsilon, \tau;\theta\right) \leq \sup_{||\epsilon||=t_n}\left|\mathbb{R}_n\left(\widehat{\alpha}(\tau;\theta), \epsilon, \tau;\theta\right)\right|\right\}. \qquad \text{(A1.13)}
\end{aligned}
$$

Under Assumption 3, there exists a $C > 0$ such that for all $\tau \in [\underline{\tau}, \overline{\tau}]$ and $\theta \in \Theta$,

$$
H(\tau;\theta) \succ CM; \text{ where } M = \mathbb{E}\left[X(\theta)X(\theta)'\right],
$$

and for all u in $\mathbb{R}^P$,

$$\begin{aligned}u'Mu &= \mathbb{E}\left[u'X(\theta)X(\theta)'u\right] = \mathbb{E}\left[\left(u'X(\theta)\right)^2\right] \\ &= \int \left(u'x\right)^2 f_X(x|\theta)dx \geq C\int_{\mathcal{H}} \left(u'x\right)^2 dx \geq C\left|\left|u\right|\right|^2,\end{aligned}$$

where the last bound uses the fact that $u \mapsto \left(\int \left(u'x\right)^2 dx\right)^{1/2}$ is a norm and norm over $\mathbb{R}^P$ are equivalent. Hence, for any non-zero $u \in \mathbb{R}^P$, M is a positive definite matrix. This implies that if $\underline{\phi}_M$ is the smallest eigenvalue of M, then, $\underline{\phi}_M > 0$. Since $H(\tau;\theta) \succ CM$, it follows for the smallest eigenvalue of the positive definite symmetric matrix $H(\tau;\theta)$, denoted by $\underline{\phi}_n(\tau;\theta)$, that

$$\inf_{(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\times\Theta} \underline{\phi}_n(\tau;\theta) \geq C\underline{\phi}_M + o_{\mathbb{P}}(1);\ \text{for some } \underline{\phi}_M > 0. \tag{A1.14}$$

Consider $\inf_{(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\times\Theta} \inf_{||\epsilon||=t_n} \mathbb{L}_n^0\left(\widehat{\alpha}(\tau;\theta),\epsilon,\tau;\theta\right)$. The definition of $\mathbb{L}_n^0$ in (A1.5) and the result obtained in (A1.14) give, for any ϵ with $||\epsilon|| \geq t_n$,

$$\mathbb{L}_n^0\left(\widehat{\alpha}(\tau;\theta),\epsilon,\tau;\theta\right) = \frac{1}{2}\epsilon' H(\tau;\theta)\epsilon \geq \frac{1}{2}\underline{\phi}_n t_n^2.$$

The above result, (A1.12) and (A1.13) give

$$\begin{aligned}\lim_{n\to\infty}\mathbb{P}\left(\sup_{\substack{(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\\ \times\Theta}} \left|\left|\widehat{\mathcal{E}}(\tau;\theta)\right|\right| \geq t_n\right) &\leq \lim_{n\to\infty}\mathbb{P}\left(\sup_{\substack{(\epsilon,\tau,\theta)\in\mathcal{B}(0,t_n)\\ \times[\underline{\tau},\overline{\tau}]\times\Theta}} \left|\mathbb{R}_n\left(\widehat{\alpha}(\tau;\theta),\epsilon,\tau;\theta\right)\right| \geq \frac{1}{2}\underline{\phi}_n t_n^2\right) \\ &\leq \lim_{n\to\infty}\mathbb{P}\left(\sup_{\substack{(\epsilon,\tau,\theta)\in\mathcal{B}(0,t_n)\\ \times[\underline{\tau},\overline{\tau}]\times\Theta}} \left|\mathbb{R}_n\left(\widehat{\alpha}(\tau;\theta),\epsilon,\tau;\theta\right)\right| \geq \frac{1}{4}\underline{\phi} t_n^2\right) \\ &= \xi + O\left(\frac{C_\xi^{1/2}}{t}\right).\end{aligned}$$

The latter can be made arbitrarily small by choosing ξ arbitrarily small and t large enough. Hence, recalling that $t_n = \left(t\log^{3/4} n\right)/n^{1/4}$, we have,

$$\sup_{(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\times\Theta} \left|\left|\widehat{\mathcal{E}}(\tau;\theta)\right|\right| = O_{\mathbb{P}}\left(\frac{\log^{3/4} n}{n^{1/4}}\right).$$

This proves Proposition 2-(i). Finally, note that $O_{\mathbb{P}}\left(\frac{\log^{3/4} n}{n^{1/4}}\right) = \left(\frac{\log^{3/4} n}{n^{1/4}}\right) O_{\mathbb{P}}(1) = o(1)O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)$. □

Proof of Proposition 2-(ii). Setting $Z_i(\theta) = Y_i(\theta) - X_i'(\theta)\beta(\tau;\theta)$,

$$\begin{aligned} \widehat{S}(\tau;\theta) - \widehat{S}(\tau;\theta_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{s}_i(\tau;\theta), \text{ where} \\ \widetilde{s}_i(\tau;\theta) &= [X_i(\theta)\{\mathbb{I}(Z_i(\theta)\le 0)-\tau\} - X_i(\theta_0)\{\mathbb{I}(Z_i(\theta_0)\le 0)-\tau\}] \\ &\le 2(X_i(\theta)+X_i(\theta_0)) \end{aligned}$$

This implies that

$$\left|\left|\frac{\widetilde{s}_i(\tau;\theta)}{\sqrt{n}}\right|\right| \le \frac{C}{\sqrt{n}} \asymp n^{-1/2} \equiv \overline{\nu}'''$$

By Assumption 2, for $C^{(1)} < \infty$ such that $\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta'}[Z_i(\theta)]\right\| \le C^{(1)}$, Taylor inequality gives $|Z_i(\theta) - Z_i(\theta_0)| \le C^{(1)}\,||\theta-\theta_0||$. Then, under Assumptions 1 and 3, we have

$$\begin{aligned} & Var\left(\frac{\widetilde{s}_i(\tau;\theta)}{\sqrt{n}}\right) \\ &= \frac{1}{n}\mathbb{E}\left[\left((X(\theta_0)\tau - X(\theta)\tau) + (X(\theta)\mathbb{I}[Z(\theta)\le 0] - X(\theta_0)\mathbb{I}[Z(\theta_0)\le 0])\right)^2\right] \\ &\le \frac{2}{n}\mathbb{E}\left[(X(\theta_0)\tau - X(\theta)\tau)^2 + (X(\theta)(\mathbb{I}[Z(\theta)\le 0] - \mathbb{I}[Z(\theta_0)\le 0])\right. \\ &\quad \left. +\mathbb{I}[Z(\theta_0)\le 0](X(\theta)-X(\theta_0)))^2\right] \\ &\le \frac{C}{n}\|\theta-\theta_0\|^2 + \frac{C}{n}\mathbb{E}\left[\mathbb{I}\left(-\frac{C}{\sqrt{n}} \le Z(\theta_0) \le \frac{C}{\sqrt{n}}\right)\right] \le \frac{C}{n}\|\theta-\theta_0\| = O\left(n^{-3/2}\right). \end{aligned}$$

Hence, the standard deviation of $\widetilde{s}_i(\tau;\theta)/\sqrt{n}$ is $\overline{\sigma}''' \asymp n^{-3/4}$. Then arguing as in Steps 2-3 of Lemma 3 (see Appendix 2),

$$\begin{aligned} \mathbb{E}\left[\sup_{\tau\in[\underline{\tau},\overline{\tau}],||\theta-\theta_0||\le C/\sqrt{n}}\left|\left|\widehat{S}(\tau;\theta) - \widehat{S}(\tau;\theta_0)\right|\right|\right] &= O\left(n^{1/2}\overline{\sigma}'''\log^{1/2}n + (\overline{\sigma}''' + \overline{\nu}''')\log n\right) \\ &= O\left(\frac{\log^{1/2}n}{n^{1/4}} + \frac{\log n}{n^{3/4}} + \frac{\log n}{n^{1/2}}\right) \\ &= O\left(\frac{\log^{1/2}n}{n^{1/4}}\left(1 + \frac{\log^{1/2}n}{n^{1/2}} + \frac{\log^{1/2}n}{n^{1/4}}\right)\right) \\ &= O\left(\frac{\log^{1/2}n}{n^{1/4}}\right). \end{aligned}$$

Note that by Lemma 2 we have

$$\sup_{(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\times\mathcal{B}\left(\theta_0,Cn^{-1/2}\right)} \left\|H^{-1}(\theta;\tau) - H^{-1}(\theta_0;\tau)\right\| = O\left(n^{-1/2}\right);$$
$$\sup_{(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\times\mathcal{B}\left(\theta_0,Cn^{-1/2}\right)} \left\|H^{-1}(\theta;\tau)\right\| \leq C.$$

Markov inequality and Lemma 5, then, explain the order in (3.5). □

Appendix 2. Proof of intermediary Lemmas for Proposition 2

Proof of Lemma 3: Bound for $\mathbb{R}_n^1(\gamma,\epsilon,\tau;\theta)$. In what follows, we treat quantities varying with i as random variables. The proof of Lemma 3 proceeds in steps as follows.

Step 1. Variance of $R(\gamma,\epsilon,\tau;\theta)$. Note that $\rho_a(b) = (a - \mathbb{I}(b<0))b = \int_0^b (a - \mathbb{I}(t<0))dt$. Denoting

$$\delta(\gamma;\theta) = X(\theta)'\gamma/\sqrt{n}, \text{ and } Z(\tau;\theta) = Y(\theta) - X(\theta)'\beta(\tau;\theta), \tag{A2.1}$$

and using definitions in (A1.6) and (A1.7), for a given $\theta \in \Theta$ we have

$$\begin{aligned} R(\gamma,\epsilon,\tau;\theta) &= \rho_\tau\left(Z(\tau;\theta) - \delta(\gamma+\epsilon;\theta)\right) \\ &\quad - \rho_\tau\left(Z(\tau;\theta) - \delta(\gamma;\theta)\right) - \delta(\epsilon;\theta)\left(\mathbb{I}\left(Z(\tau;\theta)\leq 0\right) - \tau\right) \\ &= \int_{\delta(\gamma;\theta)}^{\delta(\gamma;\theta)+\delta(\epsilon;\theta)} \left(\mathbb{I}\left(Z(\tau;\theta)\leq t\right) - \mathbb{I}\left(Z(\tau;\theta)\leq 0\right)\right)dt. \end{aligned} \tag{A2.2}$$

Using Cauchy-Schwarz inequality,

$$\begin{aligned} R(\gamma,\epsilon,\tau;\theta)^2 &\leq |\delta(\epsilon;\theta)| \left|\int_{\delta(\gamma;\theta)}^{\delta(\gamma;\theta)+\delta(\epsilon;\theta)} \left(\mathbb{I}\left(Z(\tau;\theta)\leq t\right) - \mathbb{I}\left(Z(\tau;\theta)\leq 0\right)\right)^2 dt\right| \\ &\leq |\delta(\epsilon;\theta)| \left|\int_{\delta(\gamma;\theta)}^{\delta(\gamma;\theta)+\delta(\epsilon;\theta)} \mathbb{I}\left(|Z(\tau;\theta)|\leq |t|\right)dt\right|. \end{aligned}$$

Under Assumption 3,

$$\mathbb{E}[R^2(\gamma,\epsilon,\tau;\theta)|X(\theta)] \leq |\delta(\epsilon;\theta)| \left|\int_{\delta(\gamma;\theta)}^{\delta(\gamma;\theta)+\delta(\epsilon;\theta)} \left\{\int \mathbb{I}\left(|y - X(\theta)'\beta(\tau;\theta)|\leq |t|\right) f(y|X,\theta)dy\right\}dt\right|,$$

$$\leq ||f(\cdot|\cdot,\cdot)||_\infty \, |\delta(\epsilon;\theta)| \left| \int_{\delta(\gamma;\theta)}^{\delta(\gamma;\theta)+\delta(\epsilon;\theta)} \left\{ \int \mathbb{I}\left(|y - X(\theta)'\beta(\tau;\theta)| \leq |t|\right) dy \right\} dt \right|,$$

$$\leq ||f(\cdot|\cdot,\cdot)||_\infty \, |\delta(\epsilon;\theta)| \left| 2 \int_{\delta(\gamma;\theta)}^{\delta(\gamma;\theta)+\delta(\epsilon;\theta)} |t| dt \right|,$$

$$= ||f(\cdot|\cdot,\cdot)||_\infty \, |\delta(\epsilon;\theta)| \left| 2 \int_{0}^{\delta(\epsilon;\theta)} |\delta(\gamma;\theta) + u| du \right| \text{(change of variable } t{=}u{+}\delta(\gamma;\theta))$$

$$\leq ||f(\cdot|\cdot,\cdot)||_\infty \, |\delta(\epsilon;\theta)| \left| 2 \int_{0}^{|\delta(\epsilon;\theta)|} (|\delta(\gamma;\theta)| + |u|) du \right|,$$

$$\leq C||f(\cdot|\cdot,\cdot)||_\infty \, |\delta(\epsilon;\theta)|^2 \left(|\delta(\gamma;\theta)| + |\delta(\epsilon;\theta)|\right) \leq \frac{C||X(\theta)||^3}{n^{3/2}} ||\epsilon||^2 (||\gamma|| + ||\epsilon||).$$

Under Assumption 3,

$$\begin{aligned}
\operatorname{Var}(R(\gamma,\epsilon,\tau;\theta)) &\leq \mathbb{E}[R^2(\gamma,\epsilon,\tau;\theta)] = \mathbb{E}[\mathbb{E}[R^2(\gamma,\epsilon,\tau;\theta)|X(\theta)]] \\
&\leq \mathbb{E}\left[\frac{C||X(\theta)||^3}{n^{3/2}} ||\epsilon||^2 (||\gamma|| + ||\epsilon||)\right] \\
&= \frac{C||\epsilon||^2(||\gamma|| + ||\epsilon||)}{n^{3/2}} \int ||x||^3 f_X(x|\theta) dx \leq \frac{C||\epsilon||^2(||\gamma|| + ||\epsilon||)}{n^{3/2}}.
\end{aligned} \tag{A2.3}$$

Step 2. Brackets of $\{R(\gamma,\epsilon,\tau;\theta)\}$. Let $\mathcal{F} = \{R(\gamma,\epsilon,\tau;\theta); (\gamma,\epsilon,\tau;\theta) \in \mathcal{B}(0,t_\gamma) \times \mathcal{B}(0,t_\epsilon) \times [\underline{\tau},\overline{\tau}] \times \Theta\}$. This step finds coverings of $\mathcal{F}$ with brackets $[\underline{R},\overline{R}]$, where the bracket $[\underline{R},\overline{R}]$ is the set of all R_j such that $\underline{R} \leq R_j \leq \overline{R}$ almost surely. Define for γ in $\mathbb{R}^P$

$$\widetilde{R}(\gamma,\tau;\theta) = \int_0^{\delta(\gamma;\theta)} \left(\mathbb{I}\left(Z(\tau;\theta) \leq t\right) - \mathbb{I}\left(Z(\tau;\theta) \leq 0\right)\right) dt,$$

which is such that, from (A2.2),

$$R(\gamma,\epsilon,\tau;\theta) = \widetilde{R}(\gamma+\epsilon,\tau;\theta) - \widetilde{R}(\gamma,\tau;\theta) \tag{A2.4}$$

Let $\operatorname{sgn}(t) = \mathbb{I}(t \geq 0) - \mathbb{I}(t < 0)$, such that with a change of variable $u = t/\operatorname{sgn}(\delta(\gamma;\theta))$, we have

$$\widetilde{R}(\gamma,\tau;\theta) = \int_0^{|\delta(\gamma;\theta)|} (\mathbb{I}(Z(\tau;\theta) \leq \operatorname{sgn}(\delta(\gamma;\theta))u) - \mathbb{I}(Z(\tau;\theta) \leq 0)) \operatorname{sgn}(\delta(\gamma;\theta)) du$$

$$
\begin{aligned}
&= \int_0^{|\delta(\gamma;\theta)|} |\mathbb{I}(Z(\tau;\theta) \le \text{sgn}(\delta(\gamma;\theta))u) - \mathbb{I}(Z(\tau;\theta) \le 0)| du \\
&= |\delta(\gamma;\theta)| \int_0^1 |\mathbb{I}(Z(\tau;\theta) \le \delta(\gamma;\theta)v) - \mathbb{I}(Z(\tau;\theta) \le 0)| dv, \\
&= |\delta(\gamma;\theta)| \int_0^1 |\mathbb{I}(Z(\tau;\theta) \text{ lies between 0 and } \delta(\gamma;\theta)v)| dv, \qquad \text{(A2.5)}
\end{aligned}
$$

where the second last line is obatined using change of variable $v = u/|\delta(\gamma;\theta)|$. Hence, $0 \le \widetilde{R}(\gamma,\tau;\theta) \le |\delta(\gamma;\theta)|$. Then, using the definition of $\delta(\gamma;\theta)$ in (A2.1), we get for all $\gamma \in \mathcal{B}(0, t_\gamma + t_\epsilon)$,

$$
|\widetilde{R}(\gamma,\tau;\theta)| \le ||X(\theta)|| \frac{||\gamma||}{\sqrt{n}} \le \frac{\overline{\nu}}{4}, \text{ where } \overline{\nu} \asymp \frac{t_\gamma + t_\epsilon}{\sqrt{n}}. \qquad \text{(A2.6)}
$$

It follows from (A2.4) and the result of Step 1 given in (A2.3) that

$$
\begin{aligned}
\mathbb{E}\left[|R(\gamma,\epsilon,\tau;\theta) - \mathbb{E}[R(\gamma,\epsilon,\tau;\theta)]|^k\right] &= \mathbb{E}\Big[\Big|\widetilde{R}(\gamma+\epsilon,\tau;\theta) - \mathbb{E}\left[\widetilde{R}(\gamma+\epsilon,\tau;\theta)\right] - \\
&\quad \left\{\widetilde{R}(\gamma,\tau;\theta) - \mathbb{E}\left[\widetilde{R}(\gamma,\tau;\theta)\right]\right\}\Big|^{k\text{-}2} |R(\gamma,\epsilon,\tau;\theta) - \mathbb{E}\left[R(\gamma,\epsilon,\tau;\theta)\right]|^2\Big] \qquad \text{(A2.7)} \\
&\le \left(4 \times \frac{\overline{\nu}}{4}\right)^{k-2} \text{Var}(R(\gamma,\epsilon,\tau;\theta)) \le \frac{k!}{2} \overline{\nu}^{k-2} \overline{\sigma}^2, \text{ where } \overline{\sigma}^2 \asymp \frac{t_\epsilon^2 (t_\epsilon + t_\gamma)}{n^{3/2}}.
\end{aligned}
$$

In order to find covering for $\mathcal{F}$, we first define $\widetilde{\mathcal{F}}_t = \{\widetilde{R}(\gamma,\tau;\theta); (\gamma,\tau,\theta) \in \mathcal{B}(0,t) \times [\underline{\tau}, \overline{\tau}] \times \Theta\}$ and show that it is sufficient to find covering of $\widetilde{\mathcal{F}}_t$, with set of brackets $\{[\underline{R}_j, \overline{R}_j], 1 \le j \le e^{h(t_b;t)}\}$, where $t_b \in (0,1)$ denotes length of a bracket, satisfying,

$$
\mathbb{E}\left[\left|\overline{R}_j - \underline{R}_j\right|^k\right] \le \frac{k!}{8} \left(\frac{\overline{\nu}}{2}\right)^{k-2} t_b^2, \qquad \text{(A2.8)}
$$

$$
h(t_b;t) \le C \log\left(\frac{nt}{t_b}\right). \qquad \text{(A2.9)}
$$

Consider the following two coverings of $\widetilde{\mathcal{F}}_{t_\gamma}$ and $\widetilde{\mathcal{F}}_{t_\gamma + t_\epsilon}$

$$
\widetilde{\mathcal{F}}_{t_\gamma} \subset \bigcup_{1 \le j \le e^{h(t_b;t_\gamma)}} \left[\underline{R}_j^1, \overline{R}_j^1\right], \quad \widetilde{\mathcal{F}}_{t_\gamma + t_\epsilon} \subset \bigcup_{1 \le j \le e^{h(t_b;t_\gamma + t_\epsilon)}} \left[\underline{R}_j^2, \overline{R}_j^2\right]
$$

If such coverings of $\widetilde{\mathcal{F}}_{t_\gamma}$ and $\widetilde{\mathcal{F}}_{t_\gamma + t_\epsilon}$ exist, then for every $(\gamma,\epsilon,\tau;\theta)$, $\widetilde{R}(\gamma,\tau;\theta) \in \left[\underline{R}_{j_1}^1, \overline{R}_{j_1}^1\right]$, $\widetilde{R}(\gamma+\epsilon,\tau;\theta) \in \left[\underline{R}_{j_2}^2, \overline{R}_{j_2}^2\right]$, for some j_1 and j_2, and from (A2.4), we have $R(\gamma,\epsilon,\tau;\theta) \in \left[\underline{R}_{j_2}^2 - \overline{R}_{j_1}^1, \overline{R}_{j_2}^2 - \underline{R}_{j_1}^1\right]$. Hence, $\mathcal{F}$ can be covered by $e^{h'(t_b;t)}$ brackets such that, using (A2.8) and (A2.9),

$$h'(t_b;t) = h(t_b;t_\gamma) + h(t_b;t_\gamma+t_\epsilon) \leq C\log\left(\frac{n(t_\gamma+t_\epsilon)}{t_b}\right), \text{ and}$$

$$\begin{aligned}\mathbb{E}\left[\left|\overline{R}^2_{j_2} - \underline{R}^1_{j_1} - \left(\underline{R}^2_{j_2} - \overline{R}^1_{j_1}\right)\right|^k\right] &= \mathbb{E}\left[\left|\left(\overline{R}^2_{j_2} - \underline{R}^2_{j_2}\right) + \left(\overline{R}^1_{j_1} - \underline{R}^1_{j_1}\right)\right|^k\right] \\ &\leq 2^{k-1}\left(\mathbb{E}\left[\left|\overline{R}^2_{j_2} - \underline{R}^2_{j_2}\right|^k\right] + \mathbb{E}\left[\left|\overline{R}^1_{j_1} - \underline{R}^1_{j_1}\right|^k\right]\right) \\ &\leq 2^{k-1}\frac{k!}{8}\left(\frac{\overline{\nu}}{2}\right)^{k-2} t_b^2 = \frac{k!}{2}\overline{\nu}^{k-2}t_b^2.,\end{aligned}$$

where the inequality in the second line of the above equation follows because, for $a > 0$, $b > 0$, $(a+b)^k \leq 2^{k-1}(a^k + b^k)$. We now construct covering for $\widetilde{\mathcal{F}}_t$. Lemma 2 proves that $\beta(\tau;\theta)$ is continuously differentiable in $\mu = (\tau,\theta)$ over $[\underline{\tau},\overline{\tau}] \times \Theta$ with bounded derivative. Then from from Taylor's inequality we get, for all μ_1, μ_2 in $[\underline{\tau},\overline{\tau}] \times \Theta$,

$$|x(\theta)'\beta(\mu_1) - x(\theta)'\beta(\mu_2)| \leq C\,||\mu_1 - \mu_2||\,. \tag{A2.10}$$

Also, given $\theta \in \Theta$, for all γ_1, γ_2 in $\mathbb{R}^P$, we have

$$|\delta(\gamma_1;\theta_1) - \delta(\gamma_2;\theta_2)| \leq \frac{C}{\sqrt{n}}\,||\gamma_1 - \gamma_2||\,. \tag{A2.11}$$

Define $r(q,\delta) = \int_0^1 \rho(q,\delta v)dv$, where

$$\rho(q,\delta) = |\mathbb{I}(q \leq \delta) - \mathbb{I}(q \leq 0)| = \mathbb{I}\left(q \in (0,\delta]\right)\mathbb{I}\left(\delta \geq 0\right) + \mathbb{I}\left(q \in [\delta,0)\right)\mathbb{I}\left(\delta < 0\right).$$

So, from (A2.5), we see

$$\widetilde{R}(\gamma,\tau;\theta) = |\delta(\gamma;\theta)|\, r\left(Z(\tau;\theta), \delta(\gamma;\theta)\right).$$

Note that $\rho(q,\delta)$ is a step function which is 1 for q between 0 and δ, and 0 elsewhere, for a given δ. Let $\underline{\rho}(q,\delta)$ and $\overline{\rho}(q,\delta)$ be smooth approximations of $\rho(q,\delta)$, constructed using Friedrichs mollifier of the form

$$\Phi(x) = C\begin{cases} e^{-1/\left(1-|x|^2\right)}, & \text{if } |x| < 1 \\ 0, & \text{if } |x| \geq 1 \end{cases},$$

where $C > 0$ and chosen so that $\int_{-1}^1 \Phi(x)dx = 1$ (see Stroock (2011), chapter 6 for details). As such, for $\eta > 0$, the convolution procedure yields that there exist smooth approximation functions $\underline{\rho}(q,\delta)$, $\overline{\rho}(q,\delta)$, and an open set $D_\eta \subset \mathbb{R}^2$ such that:

(i) $0 \le \underline{\rho}(q,\delta) \le \rho(q,\delta) \le \overline{\rho}(q,\delta) \le 1$ for all $(q,\delta) \in D_\eta$, with $\underline{\rho}(q,\delta) = \rho(q,\delta) = \overline{\rho}(q,\delta)$ if $(q,\delta) \in \mathbb{R}^2 \backslash D_\eta$,

(ii) $\sup_{(q,\delta)\in D_\eta} \left(\left|\frac{\partial \underline{\rho}(q,\delta)}{\partial q}\right| + \left|\frac{\partial \underline{\rho}(q,\delta)}{\partial \delta}\right| + \left|\frac{\partial \overline{\rho}(q,\delta)}{\partial q}\right| + \left|\frac{\partial \overline{\rho}(q,\delta)}{\partial \delta}\right| \right) \le C\eta^{-1/2}$, and, $\frac{\partial \underline{\rho}(q,\delta)}{\partial q} = \frac{\partial \underline{\rho}(q,\delta)}{\partial \delta} = \frac{\partial \overline{\rho}(q,\delta)}{\partial q} = \frac{\partial \overline{\rho}(q,\delta)}{\partial \delta} = \frac{\partial \rho(q,\delta)}{\partial q} = \frac{\partial \rho(q,\delta)}{\partial \delta} = 0$, when $(q,\delta) \in \mathbb{R}^2 \backslash D_\eta$,

(iii) $D_\eta \subset D'_\eta = \left\{(q,\delta) \in \mathbb{R}^2; |q| \le C\eta^{1/2} \text{ or } |q-\delta| \le C\eta^{1/2}\right\}$

Define $\underline{r}(q,\delta) = \int_0^1 \underline{\rho}(q, v\delta)dv$, $\overline{r}(q,\delta) = \int_0^1 \overline{\rho}(q, v\delta)dv$ and

$$\underline{R}(\gamma,\tau;\theta) = |\delta(\gamma;\theta)|\, \underline{r}\left(Z(\tau;\theta), \delta(\gamma;\theta)\right), \quad \overline{R}(\gamma,\tau;\theta) = |\delta(\gamma;\theta)|\, \overline{r}\left(Z(\tau;\theta), \delta(\gamma;\theta)\right)$$

such that condition (i) implies

$$\underline{R}(\gamma,\tau;\theta) \le \widetilde{R}(\gamma,\tau;\theta) \le \overline{R}(\gamma,\tau;\theta). \tag{A2.12}$$

We now bound $\underline{R}(\gamma_1,\mu_1) - \underline{R}(\gamma_2,\mu_2)$ and $\overline{R}(\gamma_1,\mu_1) - \overline{R}(\gamma_2,\mu_2)$.

$$\begin{aligned}
|\underline{R}(\gamma_1,\mu_1) - \underline{R}(\gamma_2,\mu_2)| &= ||\delta(\gamma_1;\theta_1)|\, \underline{r}\left(Z(\mu_1), \delta(\gamma_1;\theta_1)\right) - |\delta(\gamma_2;\theta_2)|\, \underline{r}\left(Z(\mu_2), \delta(\gamma_2;\theta_2)\right)| \\
&= ||\delta(\gamma_1;\theta_1)|\, \underline{r}\left(Z(\mu_1), \delta(\gamma_1;\theta_1)\right) - |\delta(\gamma_2;\theta_2)|\, \underline{r}\left(Z(\mu_2), \delta(\gamma_2;\theta_2)\right) \\
&\quad + |\delta(\gamma_2;\theta_2)|\, \underline{r}\left(Z(\mu_1), \delta(\gamma_1;\theta_1)\right) - |\delta(\gamma_2;\theta_2)|\, \underline{r}\left(Z(\mu_1), \delta(\gamma_1;\theta_1)\right)|
\end{aligned}$$

Hence,

$$\begin{aligned}
|\underline{R}(\gamma_1,\mu_1) - \underline{R}(\gamma_2,\mu_2)| \le{} & |\,|\delta(\gamma_1;\theta_1) - \delta(\gamma_2;\theta_2)|\, \underline{r}\left(Z(\mu_1), \delta(\gamma_1;\theta_1)\right) \\
& + |\delta(\gamma_2;\theta_2)|\, |\underline{r}\left(Z(\mu_1), \delta(\gamma_1;\theta_1)\right) - \underline{r}\left(Z(\mu_2), \delta(\gamma_2;\theta_2)\right)|\,|.
\end{aligned}$$

Using the definitions of $Z(\tau;\theta)$ and $\delta(\gamma;\theta)$ given in (A2.1), the bounds on increments of $x(\theta)'\beta(\tau;\theta)$ and $\delta(\gamma;\theta)$ obtained in (A2.10) and (A2.11), respectively, conditions (i, ii) and Taylor's inequality, we have, for all (γ_1,μ_1), (γ_2,μ_2) in $\mathcal{B}(0,t)\times[\underline{\tau},\overline{\tau}]\times\Theta$, where $t = t_\gamma + t_\epsilon \ge 1$,

$$\begin{aligned}
|\underline{R}(\gamma_1,\mu_1) - \underline{R}(\gamma_2,\mu_2)| &\le \frac{C\,||\gamma_1-\gamma_2||}{\sqrt{n}} + C\frac{t\eta^{-1/2}}{\sqrt{n}}\left(||\mu_1-\mu_2|| + \frac{||\gamma_1-\gamma_2||}{\sqrt{n}}\right) \\
&\le \frac{C}{\sqrt{n}}\left(1 + t\eta^{-1/2}\right)\left(||\mu_1-\mu_2|| + ||\gamma_1-\gamma_2||\right).
\end{aligned}$$

Arguing similarly gives

$$\left|\overline{R}(\gamma_1,\mu_1) - \overline{R}(\gamma_2,\mu_2)\right| \le \frac{C}{\sqrt{n}}\left(1 + t\eta^{-1/2}\right)\left(||\mu_1-\mu_2|| + ||\gamma_1-\gamma_2||\right).$$

From van de Geer (2000) there exists a covering of $\mathcal{B}(0,t)\times[\underline{\tau},\overline{\tau}]\times\Theta$ by L balls $\mathcal{B}((\gamma_j,\mu_j),\eta)$ with centre (γ_j,μ_j) and radius η such that

$$L \le \max\left(1, \frac{Ct^P}{\eta^{P+d+1}}\right), \text{ where } \gamma\in\mathbb{R}^P, \mu=(\tau;\theta)\in[\underline{\tau},\overline{\tau}]\times\mathbb{R}^d. \tag{A2.13}$$

Note that for a ball of radius η with centre (γ_j,μ_j) and (γ_2,μ_2) inside this ball,

$$|\underline{R}(\gamma_j,\mu_j)-\underline{R}(\gamma_2,\mu_2)| \le \frac{C}{\sqrt{n}}\left(1+t\eta^{-1/2}\right)\eta, \quad \left|\overline{R}(\gamma_j,\mu_j)-\overline{R}(\gamma_2,\mu_2)\right| \le \frac{C}{\sqrt{n}}\left(1+t\eta^{-1/2}\right)\eta$$

Define

$$\underline{R}'_j = \underline{R}(\gamma_j,\mu_j) - \frac{C}{\sqrt{n}}\left(1+t\eta^{-1/2}\right)\eta, \quad \overline{R}'_j = \overline{R}(\gamma_j,\mu_j) + \frac{C}{\sqrt{n}}\left(1+t\eta^{-1/2}\right)\eta,$$

$$\underline{R}_j = \max(0,\underline{R}'_j), \quad \overline{R}_j = \min\left(\frac{\overline{\nu}}{2},\overline{R}'_j\right). \tag{A2.14}$$

Then, from (A2.12), for (γ,θ) in $\mathcal{B}((\gamma_j,\mu_j),\eta)$, we have

$$\underline{R}'_j \le \underline{R}_j \le \widetilde{R}(\gamma,\theta) \le \overline{R}_j \le \overline{R}'_j \tag{A2.15}$$

This implies that $\{[\underline{R}_j,\overline{R}_j], j=1,\cdots,L\}$ is a covering of $\widetilde{\mathcal{F}}_t$, with,

$$\left|\underline{R}_j-\overline{R}_j\right| \le \frac{\overline{\nu}}{2} \asymp \frac{t}{\sqrt{n}}, \tag{A2.16}$$

since $0\le\underline{R}_j\le\overline{R}_j\le\overline{\nu}/2$. We now bound $\mathbb{E}\left[\left(\overline{R}_j-\underline{R}_j\right)^2\right]$ and $\mathbb{E}\left[\left|\underline{R}_j-\overline{R}_j\right|^k\right]$. The definitions of $\delta(\gamma;\theta)$, $Z(\tau;\theta)$ in (A2.1), conditions (i, iii), Assumption 3, (A2.15) and the inequality $(a+b)^2\le 2(a^2+b^2)$ yield

$$\begin{aligned}
&\mathbb{E}\left[\left(\overline{R}_j-\underline{R}_j\right)^2\right] \le \mathbb{E}\left[\left(\overline{R}'_j-\underline{R}'_j\right)^2\right] = \mathbb{E}\left[\left(\left(\overline{R}(\gamma_j,\mu_j)-\underline{R}(\gamma_j,\mu_j)\right)+\frac{2C}{\sqrt{n}}\left(1+t\eta^{-1/2}\right)\eta\right)^2\right]\\
&\le 2\mathbb{E}\left[\left(\overline{R}(\gamma_j,\mu_j)-\underline{R}(\gamma_j,\mu_j)\right)^2\right] + \frac{C}{n}\left(1+t\eta^{-1/2}\right)^2\eta^2\\
&\le 2\mathbb{E}\left[\left(\overline{R}(\gamma_j,\mu_j)-\underline{R}(\gamma_j,\mu_j)\right)^2\right] + \frac{C(1+t)^2(\eta+\eta^2)}{n}\\
&= 2\mathbb{E}\left[\delta^2(\gamma_j;\theta_j)\left(\overline{r}(Z(\mu_j),\delta(\gamma_j;\theta_j))-\underline{r}(Z(\mu_j),\delta(\gamma_j;\theta_j))\right)^2\right] + \frac{C(1+t)^2(\eta+\eta^2)}{n}
\end{aligned}$$

$$\leq \frac{2\,||\gamma_j||^2}{n}\int ||x||^2 \left\{\int\left\{\int_0^1 \mathbb{I}\left((Z(\mu_j),\delta(\gamma_j;\theta_j)v)\in D_\eta\right)dv\right\}f(y|x,\theta)dy\right\}f_X(x|\theta)dx$$
$$+\frac{C(1+t)^2(\eta+\eta^2)}{n}$$
$$\leq C\frac{(1+t)^2}{n}(\eta+\eta^2+\eta^{1/2}),$$

where the last inequality follows from Assumption 3 and condition (iii), since

$$\int\int_0^1 \mathbb{I}\left((Z(\mu_j),\delta(\gamma_j;\theta_j)v)\in D_\eta\right)dv f(y|x,\theta)dy \leq \int \mathbb{I}\left(y\in D_\eta + x(\theta)'\beta(\mu_j)\right)f(y|x,\theta)dy$$
$$\leq C\int \mathbb{I}\left(y\in D_\eta + x(\theta)'\beta(\mu_j)\right)dy = C(\text{length of } D_\eta)\leq C\eta^{1/2}.$$

The above bound, together with (A2.16), gives for any integer $k\geq 2$,

$$\mathbb{E}\left[\left|\overline{R}_j-\underline{R}_j\right|^k\right]=\mathbb{E}\left[\left|\overline{R}_j-\underline{R}_j\right|^2\left|\overline{R}_j-\underline{R}_j\right|^{k-2}\right]\leq\left(\frac{\overline{\nu}}{2}\right)^{k-2}\mathbb{E}\left[\left(\overline{R}_j-\underline{R}_j\right)^2\right]$$
$$\leq\frac{k!}{8}\left(\frac{\overline{\nu}}{2}\right)^{k-2}C\frac{(1+t)^2}{n}(\eta+\eta^2+\eta^{1/2}).$$

Hence, (A2.8) holds if

$$\eta=\frac{1}{3C}\min\left(\left(\frac{n}{(1+t)^2}\right)^{1/2}t_b,\left(\frac{n}{(1+t)^2}\right)t_b^2,\left(\frac{n}{(1+t)^2}\right)^2t_b^4\right).$$

Recall that $t\geq 1$ and $t_b\in(0,1)$. Then it follows from (A2.13)

$$L=e^{h(t_b;t)}\leq\max\left(1,\frac{Ct^P}{\min\left(\left(\frac{n}{(1+t)^2}\right)^{1/2}t_b,\left(\frac{n}{(1+t)^2}\right)t_b^2,\left(\frac{n}{(1+t)^2}\right)^2t_b^4\right)^{P+d+1}}\right)$$
$$\leq\max\left(1,\frac{Cnt^5}{t_b^4}\right)^{P+d+1},$$

such that for large n,

$$h(t_b;t)\leq\max\left(0,(P+d+1)\log\left(\frac{Cnt^5}{t_b^4}\right)\right)=C(\log n+5\log t-4\log t_b)$$
$$\leq 4C(\log n+\log t-\log t_b)+C\log t\leq 4C\log\left(\frac{nt}{t_b}\right)+C\log\left(\frac{nt}{t_b}\right)\leq C\log\left(\frac{nt}{t_b}\right),$$

which proves (A2.9). This completes our task of constructing covering for $\widetilde{\mathcal{F}}_t$.

Step 3. Bound for $\mathbb{E}\left(\sup_{(\gamma,\epsilon,\tau;\theta)} |\mathbb{R}_n^1(\gamma,\epsilon,\tau;\theta)|\right)$.

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{(\gamma,\epsilon,\tau;\theta)\in\mathcal{B}(0,t_\gamma)\times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta}\left|\mathbb{R}_n^1(\gamma,\epsilon,\tau;\theta)\right|\right] \\
&=\mathbb{E}\left[\sup_{(\gamma,\epsilon,\tau;\theta)\in\mathcal{B}(0,t_\gamma)\times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta}\left|\sum_{i=1}^n\left(R_i(\gamma,\epsilon,\tau;\theta)-\mathbb{E}[R_i(\gamma,\epsilon,\tau;\theta)|X(\theta)]\right)\right|\right] \\
&\leq\mathbb{E}\left[\sup_{(\gamma,\epsilon,\tau;\theta)\in\mathcal{B}(0,t_\gamma)\times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta}\left|\sum_{i=1}^n\left(R_i(\gamma,\epsilon,\tau;\theta)-\mathbb{E}[R_i(\gamma,\epsilon,\tau;\theta)]\right)\right|\right] \\
&+\mathbb{E}\left[\sup_{(\gamma,\epsilon,\tau;\theta)\in\mathcal{B}(0,t_\gamma)\times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta}\left|\mathbb{E}\left[\sum_{i=1}^n\left(R_i(\gamma,\epsilon,\tau;\theta)-\mathbb{E}[R_i(\gamma,\epsilon,\tau;\theta)]\right)|X(\theta)\right]\right|\right] \\
&\leq 2\mathbb{E}\left[\sup_{(\gamma,\epsilon,\tau;\theta)\in\mathcal{B}(0,t_\gamma)\times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta}\left|\sum_{i=1}^n\left(R_i(\gamma,\epsilon,\tau;\theta)-\mathbb{E}[R_i(\gamma,\epsilon,\tau;\theta)]\right)\right|\right]
\end{aligned}
$$

Let $\overline{\nu}$, $\overline{\sigma}$, and $h(\cdot;\cdot)$ be as defined in Step 2 by equations (A2.6), (A2.7) and (A2.9). Recall that $t=t_\gamma+t_\epsilon\geq 1$ and $\overline{\sigma}<1\leq n(t_\gamma+t_\epsilon)$. Let us use the notation $h(u;t)=h(u)$. Applying Theorem 6.8 of Massart (2007), we get

$$
\mathbb{E}\left[\sup_{\substack{(\gamma,\epsilon,\tau;\theta)\in\mathcal{B}(0,t_\gamma)\\ \times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta}}\left|\sum_{i=1}^n\left(R_i(\gamma,\epsilon,\tau;\theta)-\mathbb{E}[R_i(\gamma,\epsilon,\tau;\theta)]\right)\right|\right]\leq C\left(n^{1/2}\int_0^{\overline{\sigma}}h^{1/2}(u)du+\left(\overline{\nu}+\overline{\sigma}\right)h(\overline{\sigma})\right).
$$

From the discussion in Step 2 equation (A2.9), since $\overline{\sigma}<1$, for all $u\in(0,\overline{\sigma}]$, $h(u;t)=h(u)\leq C\log\left(n(t_\gamma+t_\epsilon)/u\right)$. Therefore, by Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
n^{1/2}\int_0^{\overline{\sigma}}h^{1/2}(u)du\leq(n\overline{\sigma})^{1/2}\left(\int_0^{\overline{\sigma}}h(u)du\right)^{1/2}&\leq C(n\overline{\sigma})^{1/2}\left(\int_0^{\overline{\sigma}}\log\left(\frac{n(t_\gamma+t_\epsilon)}{u}\right)du\right)^{1/2} \\
&=C(n\overline{\sigma})^{1/2}\left(\overline{\sigma}\left(\log\left(\frac{n(t_\gamma+t_\epsilon)}{\overline{\sigma}}\right)+1\right)\right)^{1/2} \\
&\leq Cn^{1/2}\overline{\sigma}\log^{1/2}\left(\frac{n(t_\gamma+t_\epsilon)}{\overline{\sigma}}\right).
\end{aligned}
$$

With the assumptions on the order of t_γ and t_ϵ as stated in the statement of Lemma 3 and the order of $\overline{\sigma}$ obtained in (A2.7), it follows

$$\log\left(\frac{n(t_\gamma+t_\epsilon)}{\overline{\sigma}}\right) \leq C\log\left(\frac{n^{7/4}(t_\gamma+t_\epsilon)^{1/2}}{t_\epsilon}\right) \leq C\log\left(\frac{n^{7/4}n^{1/4}}{\log^{1/2} n}\right) \leq C\log n.$$

Hence, on substituting, we get

$$\mathbb{E}\left[\sup_{(\gamma,\epsilon,\tau;\theta)\in\mathcal{B}(0,t_\gamma)\times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta}\left|\mathbb{R}_n^1(\gamma,\epsilon,\tau;\theta)\right|\right] \leq C\left(n^{1/2}\overline{\sigma}\log^{1/2} n + (\overline{\nu}+\overline{\sigma})\log n\right)$$
$$\leq C\frac{t_\epsilon\left(t_\epsilon+t_\gamma\right)^{1/2}\log^{1/2} n}{n^{1/4}}\left(1+\log^{1/2} n\left(\frac{1}{n^{1/2}}+\frac{(t_\epsilon+t_\gamma)^{1/2}}{t_\epsilon n^{1/4}}\right)\right) \leq C\frac{\log^{1/2} n}{n^{1/4}}t_\epsilon\left(t_\epsilon+t_\gamma\right)^{1/2},$$

which proves Lemma 3. □

Proof of Lemma 4. The proof of Lemma 4 follows the same steps as in Lemma 3 and, hence, a sketch of the proof is provided here. Treating quantities varying with i as random variables, the expressions for $R(\gamma,\epsilon,\tau;\theta)$ given in (A2.2), $\mathbb{R}^2(\gamma,\epsilon,\tau;\theta)$ from (A1.9) and $H(\tau;\theta)$ gives

$$\begin{aligned}
&\mathbb{R}^2(\gamma,\epsilon,\tau;\theta)\\
&= \int_{\delta(\gamma;\theta)}^{\delta(\gamma;\theta)+\delta(\epsilon;\theta)} \left(F\left(X(\theta)'\beta(\tau;\theta)+t|X,\theta\right) - F\left(X(\theta)'\beta(\tau;\theta)|X,\theta\right)\right)dt - \frac{1}{2}\epsilon' H(\tau;\theta)(\epsilon+2\gamma)\\
&= \int_{\delta(\gamma;\theta)}^{\delta(\gamma;\theta)+\delta(\epsilon;\theta)} \left(F\left(X(\theta)'\beta(\tau;\theta)+t|X,\theta\right) - F\left(X(\theta)'\beta(\tau;\theta)|X,\theta\right) - tf\left(X(\theta)'\beta(\tau;\theta)|X,\theta\right)\right)dt\\
&= \int_{\delta(\gamma;\theta)}^{\delta(\gamma;\theta)+\delta(\epsilon;\theta)} t\left\{\int_0^1 \left(f\left(X(\theta)'\beta(\tau;\theta)+vt|X,\theta\right) - f\left(X(\theta)'\beta(\tau;\theta)|X,\theta\right)\right)dv\right\}dt.
\end{aligned}$$

Define

$$r(\gamma,\tau;\theta) = \int_0^{\delta(\gamma;\theta)} t\left\{\int_0^1 \left(f\left(X(\theta)'\beta(\tau;\theta)+vt|X,\theta\right) - f\left(X(\theta)'\beta(\tau;\theta)|X,\theta\right)\right)dv\right\}dt$$

which implies that $\mathbb{R}^2(\gamma,\epsilon,\tau;\theta) = r(\gamma+\epsilon,\tau;\theta) - r(\gamma,\tau;\theta)$. Using the definition of $\delta(\gamma;\theta)$ in (A2.1) and because under Assumption 3 we have $n_0 > 0$ such that $|f(a+b|x,\theta) - f(a|x,\theta)| \leq$

$n_0 |b|$, from Lemma 2, we have

$$\left|\mathbb{R}^2(\gamma,\epsilon,\tau;\theta)\right| \le \frac{n_0}{2}\left|\int_{\delta(\gamma;\theta)}^{\delta(\gamma;\theta)+\delta(\epsilon;\theta)} t^2 dt\right| = C\left|\delta(\epsilon;\theta)\left(3\delta(\gamma;\theta)^2+3\delta(\gamma;\theta)\delta(\epsilon;\theta)+\delta(\epsilon;\theta)^2\right)\right|$$
$$\le C\left|\delta(\epsilon;\theta)\right|\left(3\left|\delta(\gamma;\theta)\right|^2+3\left|\delta(\gamma;\theta)\right|\left|\delta(\epsilon;\theta)\right|+\left|\delta(\epsilon;\theta)\right|^2\right) \le C\left|\delta(\epsilon;\theta)\right|\left(\left|\delta(\gamma;\theta)\right|+\left|\delta(\epsilon;\theta)\right|\right)^2$$
$$\le C\frac{||X(\theta)||^3\,||\epsilon||\left(||\gamma||+||\epsilon||\right)^2}{n^{3/2}}.$$
$$|r(\gamma,\tau;\theta)| \le C\left|\delta(\gamma;\theta)\right|^3 \le C\frac{||X(\theta)||^3\,||\gamma||^3}{n^{3/2}} \tag{A2.17}$$

Thus, for all $\gamma \in \mathcal{B}(0, t_\gamma + t_\epsilon)$ and all $(\tau,\theta) \in [\underline{\tau},\overline{\tau}] \times \Theta$, we have

$$|r(\gamma,\tau;\theta)| \le \frac{\overline{\nu}'}{2};\quad \overline{\nu}' \asymp \frac{\left(t_\gamma+t_\epsilon\right)^3}{n^{3/2}}.$$

From (A2.17), the variance of $\mathbb{R}^2(\gamma,\epsilon,\tau;\theta)$ for all $(\gamma,\epsilon,\tau,\theta)$ in $\mathcal{B}(0,t_\gamma)\times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta$ is obtained as follows under Assumption 3,

$$\mathrm{Var}\left(\mathbb{R}^2(\gamma,\epsilon,\tau;\theta)\right) \le \mathbb{E}\left[\mathbb{R}^2(\gamma,\epsilon,\tau;\theta)^2\right] \le \left(C\frac{||\epsilon||\left(||\gamma||+||\epsilon||\right)^2}{n^{3/2}}\right)^2 \int ||x(\theta)||^3 f_X(x|\theta)dx$$
$$\le C\frac{||\epsilon||^2\left(||\gamma||+||\epsilon||\right)^4}{n^3} \le \left(\overline{\sigma}'\right)^2;\quad \overline{\sigma}' \asymp \frac{t_\epsilon\left(t_\gamma+t_\epsilon\right)^2}{n^{3/2}}.$$

Then arguing as in step 2 of Lemma 3 to construct brackets, we have

$$\mathbb{E}\left[\sup_{(\gamma,\epsilon,\tau,\theta)\in\mathcal{B}(0,t_\gamma)\times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta}\left|\mathbb{R}_n^2(\gamma,\epsilon,\tau;\theta)-\mathbb{E}\left[\mathbb{R}_n^2(\gamma,\epsilon,\tau;\theta)\right]\right|\right]$$
$$\le Cn^{1/2}\overline{\sigma}'\log^{1/2}\left(\frac{n(t_\gamma+t_\epsilon)}{\overline{\sigma}'}\right)+\left(\overline{\sigma}'+\overline{\nu}'\right)\log\left(\frac{n(t_\gamma+t_\epsilon)}{\overline{\sigma}'}\right)$$

It follows from (A2.17) and Assumption 3 that for all $(\gamma,\epsilon,\tau,\theta)$ in $\mathcal{B}(0,t_\gamma)\times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta$,

$$\left|\mathbb{E}\left[\mathbb{R}_n^2(\gamma,\epsilon,\tau;\theta)\right]\right| = \left|n\mathbb{E}\left[\mathbb{R}_i^2(\gamma,\epsilon,\tau;\theta)\right]\right| \le n\mathbb{E}\left[\left|\mathbb{R}_i^2(\gamma,\epsilon,\tau;\theta)\right|\right]$$
$$\le \frac{C}{n^{1/2}}\mathbb{E}\left[||X(\theta)||^3\,||\epsilon||\left(||\gamma||+||\epsilon||\right)^2\right]$$
$$= \frac{C}{n^{1/2}}||\epsilon||\left(||\gamma||+||\epsilon||\right)^2\int ||x(\theta)||^3 f_X(x|\theta)dx \le \frac{C}{n^{1/2}}t_\epsilon\left(t_\gamma+t_\epsilon\right)^2,$$

and using the conditions on orders of t_γ and t_ϵ as specified in Lemma 3, such that $t_\gamma \geq 1$ and $t_\gamma/t_\epsilon = O\left(n/\log^{1/2} n\right)$, we have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\gamma,\epsilon,\tau,\theta)\in\mathcal{B}(0,t_\gamma)\times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta} |\mathbb{R}_n^2(\gamma,\epsilon,\tau;\theta)|\right] \\
&\leq \mathbb{E}\left[\sup_{(\gamma,\epsilon,\tau,\theta)\in\mathcal{B}(0,t_\gamma)\times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta} \left(\left|\mathbb{R}_n^2(\gamma,\epsilon,\tau;\theta) - \mathbb{E}\left[\mathbb{R}_n^2(\gamma,\epsilon,\tau;\theta)\right]\right| + \left|\mathbb{E}\left[\mathbb{R}_n^2(\gamma,\epsilon,\tau;\theta)\right]\right|\right)\right] \\
&\leq Cn^{1/2}\overline{\sigma}'\log^{1/2}\left(\frac{n(t_\gamma+t_\epsilon)}{\overline{\sigma}'}\right) + (\overline{\sigma}'+\overline{\nu}')\log\left(\frac{n(t_\gamma+t_\epsilon)}{\overline{\sigma}'}\right) + \frac{C}{n^{1/2}}t_\epsilon\left(t_\gamma+t_\epsilon\right)^2 \\
&\leq C\frac{t_\epsilon\left(t_\gamma+t_\epsilon\right)^2}{n}\log^{1/2}\left(\frac{n^{5/2}}{t_\epsilon\left(t_\gamma+t_\epsilon\right)}\right)\left(1+\frac{(t_\gamma+t_\epsilon)}{t_\epsilon n^{1/2}}\log^{1/2}\left(\frac{n^{5/2}}{t_\epsilon\left(t_\gamma+t_\epsilon\right)}\right)\right) + \frac{C}{n^{1/2}}t_\epsilon\left(t_\gamma+t_\epsilon\right)^2.
\end{aligned}$$

Recall that $t_\epsilon = t\log^{3/4} n/n^{1/4} = o(\log^{1/2} n)$ and $t_\gamma \asymp \log^{1/2} n$, such that $t_\gamma + t_\epsilon \asymp \log^{1/2} n$. It follows that, for large n, $n^{5/2}/\left(t_\epsilon\left(t_\gamma+t_\epsilon\right)\right) \leq (C/t)\left(n^{1/4}/\log^{5/4} n\right) \leq Cn^{1/4}/t$, such that $\log\left(n^{5/2}/\left(t_\epsilon\left(t_\gamma+t_\epsilon\right)\right)\right) \leq C\log n$. Similarly, $(t_\gamma+t_\epsilon)/\left(t_\epsilon n^{1/2}\right) \leq C/\left(n\log n\right)^{1/4}$. Thus, $\left((t_\gamma+t_\epsilon)/\left(t_\epsilon n^{1/2}\right)\right) \times \log^{1/2}\left(n^{5/2}/\left(t_\epsilon\left(t_\gamma+t_\epsilon\right)\right)\right) \leq C(\log n/n)^{1/4} = o(1)$ and $\left(1+(t_\gamma+t_\epsilon)/\left(t_\epsilon n^{1/2}\right)\times\log^{1/2}\left(n^{5/2}/\left(t_\epsilon\left(t_\gamma+t_\epsilon\right)\right)\right)\right) = 1+o(1) = 1$. Therefore, it follows,

$$\begin{aligned}
\mathbb{E}\left[\sup_{(\gamma,\epsilon,\tau,\theta)\in\mathcal{B}(0,t_\gamma)\times\mathcal{B}(0,t_\epsilon)\times[\underline{\tau},\overline{\tau}]\times\Theta} |\mathbb{R}_n^2(\gamma,\epsilon,\tau;\theta)|\right] &\leq C\frac{t_\epsilon\left(t_\gamma+t_\epsilon\right)^2}{n}\log^{1/2} n + \frac{C}{n^{1/2}}t_\epsilon\left(t_\gamma+t_\epsilon\right)^2 \\
&\leq C\frac{t_\epsilon\left(t_\gamma+t_\epsilon\right)^2}{n^{1/2}},
\end{aligned}$$

for large n, which proves Lemma 4. □

Proof of Lemma 5. The first order condition for $Q(\beta,\tau;\theta)$ gives

$$\mathbb{E}\left[X(\theta)\left\{F(X'(\theta)\beta(\tau;\theta)|X,\theta)-\tau\right\}\right] = 0.$$

Let $s_{\ell i}(\tau;\theta)$ denote the ℓ^{th} entry of the vector $\frac{s_i(\tau;\theta)}{\sqrt{n}}$ in (A1.3). Assumption 3 gives, uniformly in $(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\times\Theta$ for all i,

$$\begin{aligned}
&|s_{\ell i}(\tau;\theta)| \leq \overline{\nu}'', \text{ where } \overline{\nu}'' \asymp n^{-1/2} \\
&\mathrm{Var}(s_\ell(\tau;\theta)) \leq \mathbb{E}\left[\left(s_\ell(\tau;\theta)\right)^2\right] \leq \mathbb{E}\left[\frac{1}{n}X_\ell(\theta)^2\right] = \frac{1}{n}\int x^2 f_X(x|\theta)dx \leq \left(\overline{\sigma}''\right)^2,
\end{aligned}$$

where $\overline{\sigma}'' \asymp n^{-1/2}$.

Hence, arguing as in Steps 2-3 of Lemma 3,

$$\mathbb{E}\left[\sup_{(\tau,\theta)\in[\underline{\tau},\overline{\tau}]\times\Theta}\left|\left|\widehat{S}(\tau;\theta)\right|\right|\right] = O\left(n^{1/2}\overline{\sigma}''\log^{1/2}n + (\overline{\sigma}'' + \overline{\nu}'')\log n\right)$$
$$\leq C\left(\log^{1/2}n + \left(\frac{\log n}{n}\right)^{1/2}\log^{1/2}n\right) = O\left(\log^{1/2}n\right).$$

Markov inequality, then, proves Lemma 5. □

Appendix 3. Proof of remark in Section 5

(i) In the expression for $C(\tau)$ in (4.2), $\left\{\int_0^\tau(\beta_0(t)+\beta_2(t)X_2)\,dt - \tau\left(\beta_0(\tau)+\beta_2(\tau)X_2\right)\right\}$ will have the form $p(\tau)+q(\tau)X_2$. Recall that $\widetilde{X}=[1,X_2]'$, $X=[1,X_1,X_2]'$ and $g(X)=\widetilde{X}\,[0,1,0]\,\mathbb{E}^{-1}\left[XX'\right]X$, then,

$$C(\tau)=\mathbb{E}\begin{bmatrix}\left([0,1,0]\,\mathbb{E}^{-1}\left[XX'\right]X\right)\left(p(\tau)+q(\tau)X_2\right)\\ \left([0,1,0]\,\mathbb{E}^{-1}\left[XX'\right]XX_2\right)\left(p(\tau)+q(\tau)X_2\right).\end{bmatrix}$$

If X_1 and X_2 are independent, elementary matrix algebra gives that

$$[0,1,0]\,\mathbb{E}^{-1}\left[XX'\right]X=\frac{1}{D}\left\{\left(\mathbb{E}[X_1]\mathbb{E}^2[X_2]-\mathbb{E}[X_1]\mathbb{E}[X_2^2]\right)+\left(\mathbb{E}[X_2^2]-\mathbb{E}^2[X_2]\right)X_1\right\},$$

where D is the determinant of the matrix $\mathbb{E}[XX']$. Plugging in this expression in $C(\tau)$ and simplifying using independence of X_1 and X_2 proves the result.

(ii) Given the result in (i), the increase in variance of the quantile estimates due to first step estimation, over standard quantile regression had the first step been known, is given by (4.2) as $H(\tau)^{-1}D(\tau)\mathcal{V}(\beta_1)D(\tau)'H(\tau)^{-1}$. Using $H(\tau)$ and $D(\tau)$ as given in (4.2), under independence of X_1 and X_2, the vector $H(\tau)^{-1}D(\tau)$ evaluates to $[-\mathbb{E}[X_1],0]'$. Therefore, the additional variance due to two-step estimation is given by

$$H(\tau)^{-1}D(\tau)\mathcal{V}(\beta_1)D(\tau)'H(\tau)^{-1}=\begin{bmatrix}\mathbb{E}^2[X_1]\mathcal{V}(\beta_1) & 0\\ 0 & 0\end{bmatrix}.$$

This proves (ii). □

References

Abrevaya, J. (2002). The effects of demographics and maternal behavior on the distribution of birth outcomes. In *Economic applications of quantile regression* (pp. 247–257). Springer.

Amemiya, T. (1974). The nonlinear two-stage least-squares estimator. *Journal of econometrics*, 2(2), 105–110.

Amemiya, T. & Powell, J. L. (1981). A comparison of the Box–Cox maximum likelihood estimator and the non-linear two-stage least squares estimator. *Journal of Econometrics*, 17(3), 351–381.

Andrews, D. W. & Buchinsky, M. (2000). A three-step method for choosing the number of bootstrap repetitions. *Econometrica*, 68(1), 23–51.

Arellano, M. & Bonhomme, S. (2017). Quantile selection models with an application to understanding changes in wage inequality. *Econometrica*, 85(1), 1–28.

Blundell, R. & Powell, J. L. (2003). Endogeneity in nonparametric and semiparametric regression models. In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress* (pp. 312–357). Cambridge University Press.

Blundell, R. & Powell, J. L. (2007). Censored regression quantiles with endogenous regressors. *Journal of Econometrics*, 141(1), 65–83.

Box, G. E. & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2), 211–252.

Buchinsky, M. (1994). Changes in the US wage structure 1963-1987: Application of quantile regression. *Econometrica*, (pp. 405–458).

Buchinsky, M. (1995). Quantile regression, Box–Cox transformation model, and the US wage structure, 1963–1987. *Journal of Econometrics*, 65(1), 109–154.

Chamberlain, G. (1994). Quantile regression, censoring, and the structure of wages. In C. A. Sims (Ed.), *Advances in Econometrics: Sixth World Congress*, Econometric Society Monographs (pp. 171—210). Cambridge University Press.

Chen, L., Galvao, A. F., & Song, S. (2018). Quantile regression with generated regressors. *Available at SSRN 3039602*.

Chernozhukov, V., Fernández-Val, I., & Kowalski, A. E. (2015). Quantile regression with censoring and endogeneity. *Journal of Econometrics*, 186(1), 201–221.

Chernozhukov, V., Fernández-Val, I., Newey, W., Stouli, S., & Vella, F. (2017). Semiparametric estimation of structural functions in nonseparable triangular models. *arXiv preprint arXiv:1711.02184*.

Chernozhukov, V. & Hansen, C. (2005). An IV model of quantile treatment effects. *Econometrica*, 73(1), 245–261.

Chernozhukov, V. & Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, 132(2), 491–525.

Chernozhukov, V. & Hansen, C. (2008). Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, 142(1), 379–398.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica*, 71(5), 1405–1441.

Fitzenberger, B., Wilke, R. A., & Zhang, X. (2009). Implementing Box–Cox quantile regression. *Econometric Reviews*, 29(2), 158–181.

Gimenes, N. (2017). Econometrics of ascending auctions by quantile regression. *Review of Economics and Statistics*, 99(5), 944–953.

Gimenes, N. & Guerre, E. (2016). Quantile methods for first-price auction: A signal approach. Technical report*, University of São Paulo (FEA-USP)*.

Hahn, J. & Ridder, G. (2013). Asymptotic variance of semiparametric estimators with generated regressors. *Econometrica*, 81(1), 315–340.

Haile, P. A., Hong, H., & Shum, M. (2003). Nonparametric tests for common values at first-price sealed-bid auctions. Technical report*, National Bureau of Economic Research*.

Hjort, N. L. & Pollard, D. (2011). Asymptotics for minimisers of convex processes. *arXiv preprint arXiv:1107.3806*.

Imbens, G. W. & Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5), 1481–1512.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

Koenker, R. (2017). Quantile regression: 40 years on. *Annual Review of Economics*, 9, 155–176.

Koenker, R. & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50.

Koenker, R. & Hallock, K. (2001). Quantile regression: An introduction. *Journal of Economic Perspectives*, 15(4), 43–56.

Koenker, R. & Ma, L. (2006). Quantile regression methods for recursive structural equation models. *Journal of Econometrics*, 134(2), 471–506.

Koenker, R. & Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, 94(448), 1296–1310.

Koenker, R. & Xiao, Z. (2002). Inference on the quantile regression process. *Econometrica*, 70(4), 1583–1612.

Lee, S. (2007). Endogeneity in quantile regression models: A control function approach. *Journal of Econometrics*, 141(2), 1131–1158.

Lu, J. & Perrigne, I. (2008). Estimating risk aversion from ascending and sealed-bid auctions: the case of timber auction data. *Journal of Applied Econometrics*, 23(7), 871–896.

Machado, J. A. & Mata, J. (2000). Box–Cox quantile regression and the distribution of firm sizes. *Journal of Applied Econometrics*, 15(3), 253–274.

Mammen, E., Rothe, C., & Schienle, M. (2012). Nonparametric regression with nonparametrically generated covariates. *The Annals of Statistics*, 40(2), 1132–1170.

Mammen, E., Rothe, C., & Schienle, M. (2016). Semiparametric estimation with generated covariates. *Econometric Theory*, 32(5), 1140–1177.

Massart, P. (2007). *Concentration inequalities and model selection*, volume 6. Springer.

Mu, Y. & He, X. (2007). Power transformation toward a linear regression quantile. *Journal of the American Statistical Association*, 102(477), 269–279.

Murphy, K. M. & Topel, R. H. (1985). Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics*, 3(4), 370–379.

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

Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, (pp. 221–247).

Powell, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. In W. A. Barnett, J. Powell, & G. Tauchen (Eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics* chapter 14, (pp. 357–384). Cambridge University Press.

Stroock, D. W. (2011). *Essentials of integration theory for analysis*, volume 262. Springer.

van de Geer, S. (2000). *Empirical Processes in M-Estimation*. Cambridge: University Press Cambridge.

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