Censored Quantile Regression for Duration Data with Time-varying Regressors and Endogeneity

Songnian Chen*
Hong Kong University of Science and Technology

April 3, 2019

Abstract

For duration data two fundamental features are censoring and time-varying regressors. The popular Cox proportional hazards model and other conventional duration models are highly restrictive in modelling how regressors affect the conditional duration distribution. Endogeneity such as selective compliance is also common in duration data, which cannot be accommodated by the Cox model and the associated partial likelihood approach. In this paper, we develop a quantile regression framework that allows for censoring, time-varying regressors and endogeneity, and we propose an easy-to-implement two-step quantile regression estimator. We present large sample results. Monte Carlo experiments indicate that our estimator performs well in finite samples.

1 Introduction

For duration data two fundamental features are censoring and time-varying regressors. Censoring occurs when not all spells are completed at the time of observation or follow-up; for example unemployment spells are often censored at 26 weeks when unemployment benefits run out. Time-varying regressors are very common in duration data and many important economic variables such as the nature of policy intervention, weekly unemployment benefit levels and local unemployment rates, among others, may change over time. In addition, endogeneity is also prevalent in duration analysis; for job training programs aimed at reducing unemployment duration, the treatment variable is likely to be endogenous due to, for example, selective compliance. In the context of time-invariant regressors, Powell (1991), Koenker and Geling (2001) and Fitzenberger and Wilke (2005), among others, argued that quantile regression model (QR), which is particularly well-equipped to deal with censoring, provides a flexible and yet comprehensive semiparametric approach to modeling the entire conditional duration distribution, with different regions of the conditional duration distribution characterized by different quantile regression coefficients; in particular, short-term and long-term unemployment phenomena can be modelled by lower and upper quantile regressions separately, without being unduly influenced by global features of the model specification. As a result,

*I would like to Xiaohong Chen, Andrew Chesher, Hide Ichimura, Roger Koenker, Whitney Newey, Ed Vytlacil and participants of various econometrics conferences for their helpful comments.

the quantile regression framework allows researchers to fit parsimonious models to an entire conditional distribution. In the existing literature, however, there is no effective method to conduct quantile regression analysis in the presence of censoring, time-varying regressors, and endogeneity. Our paper fills this important gap.

In economic duration analysis, conventional methods, both parametric and semiparametric, while being able to accommodate a broader class of covariates, typically impose stringent conditions on how the covariates are permitted to influence the conditional duration distribution. Consider, for example, the Cox proportional hazards model. By focusing on the conditional hazard function, the Cox proportional hazards model offers a natural way to deal with both censoring and time-varying regressors. Specifically, the Cox model specifies the conditional hazard function as

$$\lambda(t|x) = \lambda(t)\exp(\beta' x(t)),$$

where $\lambda(t)$ denotes the baseline hazard function and $x(t)$ includes both time-invariant or time-varying regressors. Accordingly, the well known Cox partial likelihood estimation approach, which involves a concave maximization problem, offers a straightforward and reliable estimation and inference mechanism for the regression coefficients and baseline hazard. In particular, censoring does not lead to much complication for the partial likelihood method. However, the Cox proportional hazards model suffers from some serious drawbacks.

As pointed out by Koenker and Geling (1994) and Koenker and Balias (2001), the proportional hazards model imposes rather drastic constraints on the way that covariates are permitted to influence the duration distribution. For example, for the case with time-invariant regressors, the implied conditional duration quantile function takes the form

$$Q_T(\tau|x) = S_0^{-1}\left((1-\tau)^{1/\gamma(x)}\right)$$

where $\gamma(x) = e^{-x'\beta}$ and $S_0(t)$ denotes the baseline survival function. Consequently, the marginal quantile effects are of the form

$$\frac{\partial Q_T(\tau|x)}{\partial x_j} = c(x,\tau)\beta_j$$

where $c(x,\tau) = \frac{(1-\tau)\log(1-\tau)\gamma(x)}{S_0'(Q_T(\tau|x))}$. So in the proportional hazards model the quantile marginal effects of the various covariates, viewed as functions of τ, are all identical up to the scalar factors determined by the components of the same global vector, β. In particular, the implicit quantile treatment effects for the Cox model must have the same sign as β_j for all τ, which effectively rules out any form of quantile treatment effect that would lead to crossings of survival functions for different settings of the covariates. Furthermore, the ratio of the quantile marginal effects for two different components remain constant for all τ as $\frac{\partial Q_T(\tau|)}{\partial x_i}/\frac{\partial Q_T(\tau|)}{\partial x_j} = \beta_i/\beta_j$, which is highly implausible in typical empirical settings. Furthermore, these restrictive features of the Cox proportional hazards model carry over to mixed proportional hazards model (MPH) and other common duration models.

The Cox proportional hazards model also suffers from another serious drawback; the partial likelihood approach does not allow the presence of endogeneity[1]. Endogeneity arises frequently in duration analysis. For example, a randomized experiment may suffer from selective compliance, which is likely to be correlated with outcome variables and thus gives rise to endogeneity. For the Illinois unemployment bonus experiment, which has been studied by Woodbury and Spiegelman (1987), Meyer (1996), Bijwaard and Ridder (2005), about 15% of the claimant bonus and 35% of the employer bonus group refused participation. Bijwaard and Ridder (2005) noted that there is evidence of selective compliance and endogeneity.

The censored quantile regression framework proposed in this paper overcomes the shortcomings associated with the Cox proportional hazards model in particular, and other conventional duration models in general. Rather than making global assumptions about how covariates influence different regions of conditional duration distribution, quantile regression allows us to focus on the estimation of particular local features of the conditional duration distribution. Thus we may explore the effect of covariates on just the upper or lower tails, or the middle regions, of the conditional distribution without being distorted by modeling assumptions about the rest of the conditional distribution. For example, in the analysis of unemployment duration, comparison of the quantile regressions for lower and upper tails of the duration distribution provides important insights on how different determinants affect short or long-term unemployment. Quantile regression constitutes a natural and flexible framework for the analysis of duration data.

For the case with time-invariant regressors, censored quantile regression with endogeneity has been a very active area of research. Blundell and Powell (2007) and Chernozhukov et. al (2015) proposed two-step control function based estimators. The control function-based approach, however, requires fully parametric specification of the joint distribution the endogenous variables and the outcome duration variable, and thus impose very strong structural restrictions on the underlying model.[2] Typically, the econometrician does not have a good understanding of the nature of endogeneity to propose a reasonably accurate model. As a result, the control function based approach is likely to lead to inconsistent estimates and misleading inference when the exact nature of endogeneity is misspecified. Furthermore, the control function approach requires the endogenous variables to be continuously distributed, thus ruling out censored and discrete endogenous variables. Indeed, the control function approach is not applicable to most program evaluation studies where the leading case typically involves binary, or multi-valued but discrete, endogenous treatment variables. In addition, it is not clear how to extend the control function based approach to the case with time-varying regressors. On the other hand, Hong and Tamer (2003) and Khan and Tamer (2010) considered moment inequalities based approaches. However, their estimation methods and

[1] While it is possible to accommodate some continuous endogenous variable through a control function approach in the context of the mixed proportional hazard model, it requires complete parametric specification for the joint generating process for the endogenous variable and the outcome variable; in addition, the control function approach does not apply to discrete endogenous variable such as the binary treatment variable. Furthermore, it is not clear how to accommodate time-varying regressors for the control function approach.

[2] As it was pointed out by Honoré and Hu (2004) that in a linear model, a reduced form in the first stage can be thought of as a linear projection, and as such it is essentially always well-defined and consistently estimated by the OLS estimator. This is not the case in a nonlinear model where it is typically assumed that the first stage is a conditional expectation and that the error is independent of the instruments. Other control function based approaches also require similar strong restrictions.

related inference procedures are quite complicated and difficult to implement. Furthermore, it is also difficult to incorporate time-varying regressors in constructing unconditional moment inequalities. More recently, Chen (2018) proposed a sequential instrumental variable censored quantile regression estimation procedure for the structural quantile regression. Chen's (2018) approach, however, only deals with the time-invariant regressors.

When all the regressors are exogenous, by combining the insights behind the standard quantile regression techniques (Koenker and Bassett, 1978) and the accelerated life time (AFT) model with time-varying regressors (Cox and Oaks, 1984), recently Chen (2019) proposed a quantile regression framework that accommodates time-varying regressors in a natural way. Chen (2019) further proposed a censored quantile regression estimator with time-varying regressors. Unlike the standard QR problem, which can be implemented through efficient linear programming, Chen's (2019) estimator, however, involves nonconvex and nonlinear optimization, which can be difficult to implement. In addition, Chen (2019) only considered the case where all regressors are exogenous.

In this paper, we develop a quantile regression framework that can accommodate censoring, time-varying regressors and endogeneity and further propose a quantile regressor estimator[3]. When all regressors are exogenous, our model reduces to the Cox-Oaks AFT model when all the quantile coefficients are parallel, and reduces the Koenker-Bassett QR model when all the regressors are time-invariant. When endogenous regressors are also present, our model reduces to the AFT model with endogeneity of Bijwarrd and Ridder (2005) when the structural quantile coefficients are all parallel. On the other hand, our model reduces to the structural quantile regression model of Chernozhukov and Hansen (2006, 2008) when all regressors are time-invariant.

An important insight behind our estimation method is the recognition that an appropriate transformation of the duration time leads to a linear quantile regression framework in terms of the time-invariant regressors, whose quantile coefficients can thus be estimated by the standard QR. Consequently, we propose a two-step method by profiling over the quantile regression coefficients for the time-varying regressors; the computational burden of our estimator essentially depends on the dimension of the endogenous variables and time-varying regressors. In addition, we deal with the complication caused by censoring based on the insight behind the sequential quantile regression approach by Chen (2018). In typical empirical settings, our estimators are easy to implement.

The paper is organized as follows. In Section 2 we discuss the model and our two-step censored quantile estimator with exogenous time-varying regressors. In Section 3 we extend the model and our estimator to allow for endogeneity. Section 4 contains some simulation results. Section 5 concludes. All the proofs are in the appendix.

[3]Bijwaard and Ridder (2005) proposed a two-stage estimator in a generalized AFT setting. However, their approach requires full compliance for the control group; in addition, like the Cox proportional hazards model, their model is also very restrictive in how covariates are permitted to influence the duration distribution. It is interesting to note that even though Bijwaard and Ridder (2005) does not consider a quantile regression framework, they nevertheless interpret their regression coefficients estimates in terms of the covariate effects on the quantiles of the distribution of the transformed duration relative to the reference individual.

2 Censored Quantile Regression with Time-Varying Regressors

To fix ideas, we consider the following model

$$\int_0^{T^*} \exp(-X'(s)\beta(U))ds = 1 \tag{2.1}$$

where T^* is the duration time, $X(s)'\beta(U) = X_1'(s)\beta_1(U) + X_2'\beta_2(U)$, U has a uniform $U(0,1)$ distribution, $X = (X_1, X_2)$ with two types of regressors X_1 and X_2, where $X_1 = \{X_1(t), t > 0\}$ is k_1-dimensional time-varying regressors and X_2 denotes a k_2-dimensional time-invariant regressors, respectively. The above model corresponds to the following quantile representation

$$\int_0^{Q_{T^*}(\tau|X)} \exp(-X'(s)\beta(\tau))ds = 1 \tag{2.2}$$

where $Q_{T^*}(\tau|X)$ denotes the τth conditional quantile function of T^* conditional on X.

When $\beta(U) = (\alpha(U), \tilde{\beta})$, model (2.1) reduces to the AFT model with time-varying regressors of Cox-Oaks (1984),

$$\int_0^{T^*} \exp(-\tilde{X}'(s)\tilde{\beta} - \alpha(U))ds = 1 \tag{2.3}$$

where with $X(s) = (1, \tilde{X}'(s))'$. The corresponding conditional quantile function satisfies

$$\int_0^{Q_{T^*}(\tau|X)} \exp(-\tilde{X}'(s)\tilde{\beta} - \alpha(\tau))ds = 1; \tag{2.4}$$

in other words, in the Cox-Oaks AFT model all the quantile coefficients are identical other than a pure location shift, which is highly restrictive in allowing how covariates affect the conditional duration distribution.

When all the regressors are time-invariant, on the other hand, model (2.1) reduces to the standard QR model,

$$\int_0^{T^*} \exp(-X'\beta(U))ds = 1 \tag{2.5}$$

or equivalently,

$$\ln T^* = X'\beta(U).$$

In particular, the conditional quantile function satisfies

$$\int_0^{Q_{T^*}(\tau|X)} \exp(-X'\beta(\tau))ds = 1 \tag{2.6}$$

or equivalently,

$$Q_{\ln T^*}(\tau|X) = X'\beta(\tau).$$

The Cox-Oaks AFT model provides a natural mechanism to accommodate time-varying regressors, whereas the standard QR model (Koenker and Bassett, 1978) provides a flexible and yet comprehensive approach to modelling the conditional duration distribution. By combining the attractive features of these two models, our model (2.1) and (2.2) offers a natural framework to conduct quantile regression for duration data with time-varying regressors

To gain some further insight into the model (2.1). First consider the standard AFT model with time-invariant regressors:

$$\ln T^* = X'\beta - \varepsilon,$$

in which case the conditional survival function satisfies

$$S(t|x) = \Pr(T^* < t|X = x) = S\left(t \exp(x'\beta)\right).$$

Consider some reference point x_0 and note that

$$S(t|x) = S\left(t \exp(x'\beta)\right) = S(t \exp((x - x_0)'\beta)|x_0).$$

Therefore, the duration times corresponding to x and x_0 differ by an accelerating factor $\exp((x - x_0)'\beta)$ for the standard AFT model with time-invariant regressors.

Within the framework of the Cox-Oaks AFT model with time-varying regressors (2.3-2.4), we now consider the impact of sustained exposure to different air qualities or other environmental hazards on life expectancy. Recently there have been several influential studies on the impact of sustained exposure to air pollution on life expectancy in China. Several decades of rapid economic growth has also brought widespread deterioration of air quality and general environment in China. For example, Ebenstein et al. (2015) suggests that China's modest growth in life-expectancy for the period 1991-2012 is mainly due to the country's severe problems with air pollution; even though China's income growth has improved health outcomes, but failed to do so for pollution-sensitive causes of death. In China, some of the most polluted cities are located in the north, especially in regions surrounding Beijing, and the cities in the south are generally less polluted, with those on the Hainan Island having some of the best air quality. Consider a representative individual who has lived in Beijing, Shanghai, and Sanya (a city in Hainan), with air qualities represented by x_{bj}, x_{sh} and x_{sy} respectively. Suppose the life expectancy for this individual follows (2.3-2.4) with a common slope coefficient β, and further assume that the corresponding relative accelerating factors satisfy $\frac{\exp(x_{sy}\beta)}{\exp(x_{sh}\beta)} = \frac{1}{0.9}$ and $\frac{\exp(x_{sy}\beta)}{\exp(x_{bj}\beta)} = \frac{1}{0.8}$; in other words, if this person were to live to 80 years in Beijing, she could have lived to 90 years in Shanghai or 100 years in Sanya. On the other hand, if this person lives to 90 years, with 30 years in each of these three cities, then she would have lived to $30 + 30 * \frac{8}{9} + 30 * \frac{8}{10} = 80.67$ years in Beijing, $30 * \frac{9}{8} + 30 + 30 * \frac{9}{10} = 90.75$ in Shanghai or $30 * \frac{10}{8} + 30 * \frac{10}{9} + 30 = 100.83$ years in Sanya if she were to live in one of these cities exclusively. Therefore, one major drawback of model (2.3-2.4) is that the model implies a uniform accelerating factor, ruling out the fact that individual with different health conditions would have experienced differently. Indeed, people more sensitive to air quality are much more likely to be affected adversely by poor air quality. But the Cox-Oaks AFT model cannot handle this type of heterogeneity.

On the other hand, the quantile regression model with time-varying regressor (2.1-2.2) provides a natural framework to accommodate various types of heterogeneity. Let U denote the unobservable measure of an individual's general health condition normalized to uniform distribution $U(0,1)$ in the population and let $\beta(u)$ denote the uth quantile coefficients for individuals with health condition at the uth quantile in the general population. Then for any given u, the corresponding accelerating factors would be

$$\pi_{bj,sh} = \frac{\exp(x'_{bj}\beta(u))}{\exp(x'_{sh}\beta(u))},\ \pi_{bj,sy} = \frac{\exp(x'_{bj}\beta(u))}{\exp(x'_{sy}\beta(u))} \text{ and } \pi_{sh,sy} = \frac{\exp(x'_{sh}\beta(u))}{\exp(x'_{sy}\beta(u))}$$

between Beijing and Shanghai, Beijing and Sanya, and Shanghai and Sanya respectively, and the above three terms can be thought of relative quantile accelerating factors, which can be different for people with different health condition. Clearly, one major advantage of the quantile regression model (2.1-2.2) allows for different $\beta(u)$ for different u, and indeed, $\beta(u)$ is expected to decrease as u increases if poor air quality affects those with poor health disproportionately. More generally, consider an individual with health condition u and exposed to air quality $x(t)$ at time t; suppose her life span is $Q(x,u)$ and has lived in a city during the period $[t_k, t_{k+1})$ with air quality $x(t_k)$, where $0 = t_0 < t_1 < t_2 ... < t_K = Q(x,u)$. Now consider the counterfactual question: how long would be this person's life expectancy if she were to live in a city entirely with air quality x_0? We answer this question from two angles. According to the Koenker and Bassett (1978) quantile regression framework, this person would have her lifespan equal to $\exp(x'_0\beta(u))$; on the other hand, based on the relative accelerating factors for individual with health condition u, the answer would be

$$\sum_{k=0}^{K} \Delta t_k \frac{\exp(x'_0\beta(u))}{\exp(x'(t_k)\beta(u))}$$

where $\Delta t_k = t_{k+1} - t_k$, with $t_K = Q(x,u)$. Taking limits yields

$$\lim_{\Delta t_k \to 0} \sum_{k=0}^{K} \Delta t_k \frac{\exp(x'_0\beta(u))}{\exp(x'(t_k)\beta(u))} = \int_0^{Q(x,u)} \frac{\exp(x'_0\beta(u))}{\exp(x'(s)\beta(u))} ds.$$

By equating the two answers, we obtain

$$\int_0^{Q(x,u)} \frac{\exp(x'_0\beta(u))}{\exp(x'(s)\beta(u))} ds = \exp(x'_0\beta(u))$$

and eliminating the common factor $\exp(x'_0\beta(u))$ yields model (2.1-2.2).

We now provide some additional remarks on model (2.1-2.2).

Remark 1: Marginal effects play a fundamental role in understanding an econometric model. Chen (2019) has shown that for a small policy change of Δx_k for the kth policy variable during the time period $[t_0, t_1]$, the corresponding marginal effect takes the form

$$ME_k = c(x, t_0, t_1) \times \exp(x'(t^*)\beta) \times \beta_k(\tau)$$

where $t^* = Q_{T^*}(\tau|x)$,

$$c(x, t_0, t_1) = \int_0^{Q_{T^*}(\tau|x)} \exp(x'(s)\beta(\tau)) 1\{t_1 > s > t_0\}) ds.$$

Here the magnitude of the proportionality factor $c(x, t_0, t_1)$ depends on the duration of the policy change. For the special case of a permanent policy change, namely, $t_0 = 0$ and $t_1 = \infty$, then $c(x, t_0, t_1) = 1$, which implies

$$ME_k = \exp(x'(t^*)\beta) \times \beta_k(\tau)$$

which is clearly a natural extension of the case with time-invariant regressors. On the other hand, when the policy change that lasts a short time period, namely, $[t_0, t_1]$ is a small time interval, then $c(x, t_0, t_1)$ takes a small value and consequently the impact of a short-term policy change is minimal.

Remark 2: In quantile regression models, an important property is quantile monotonicity (no crossing). For the quantile regression with time-invariant regressors, quantile monotonicity requires that

$$x'\beta(\tau_2) > x'\beta(\tau_1) \quad \text{for } \tau_2 > \tau_1. \tag{2.7}$$

For the quantile regression model with time-varying regressors (2.1-2.2), it is straightforward to demonstrate that quantile monotonicity would require

$$\int_0^t \exp(-x'(s)\beta(\tau_1)) ds > \int_0^t \exp(-x'(s)\beta(\tau_2)) ds \quad \text{for any } t, \tag{2.8}$$

which is similar to the first order stochastic dominance condition, or equivalently,

$$\frac{1}{t}\int_0^t \exp(-x'(s)\beta(\tau_1)) ds > \frac{1}{t}\int_0^t \exp(-x'(s)\beta(\tau_2)) ds \quad \text{for any } t$$

which suggests that condition (2.8) holds in an average sense. Note that a sufficient condition for (2.8) is that

$$x'(s)\beta(\tau_2) > x'(s)\beta(\tau_1) \quad \text{for } \tau_2 > \tau_1 \text{ and any } s.$$

We now turn to the estimation of the quantile regression model (2.1-2.2). In duration models, censoring is a common phenomenon, where the duration time T^* is subject to censoring. We consider the fixed censoring case when we observe $T = \min\{C, T^*\}$, and for simplicity we assume that C is a known constant. Extensions to the random censoring case is straightforward. For a random sample $\{X_i, T_i\}$, $i = 1, 2, ..., n$, Chen (2019) recently proposed an integrated maximum score estimator, a special case of which solves

$$\min \frac{1}{n} \sum_{i=1}^{n} \rho_\tau (\ln T_i - \min\{\ln T(X_i, b), \ln C_i\}) \tag{2.9}$$

where $\rho_\tau(u) = (\tau - 1\{u < 0\})u$, $T(X_i, b)$ satisfies

$$\int_0^{T(x,b)} \exp(-x_1'(s)b_1 - x_2'b_2)ds = 1, \tag{2.10}$$

and in particular

$$T(x, \beta(\tau)) = Q_{T^*}(\tau|x).$$

One major drawback, however, associated with Chen's (2018) estimator is that it requires solving a nonlinear nonconvex minimization problem, which can be very demanding computationally. In this paper, we propose a computationally attractive alternative. In particular, the computational difficulty of our new estimator essentially depends on the dimension of the time-varying regressors.

To motivate our new approach, define the following transformation of the duration time

$$T_1^*(b_1) = \int_0^{T^*} \exp(-x_1'(s)b_1)ds. \tag{2.11}$$

With some algebra we can show that

$$Q_{\ln T_1^*(\beta_1(\tau))}(\tau|X) = X_2'\beta_2(\tau) \tag{2.12}$$

In other words, if $\beta_1(\tau)$ were known and there is no censoring, then $\beta_2(\tau)$ can be estimated by standard quantile regression of $\ln T_1^*(\beta_1(\tau))$ on X_2.

(2.11) and (2.12) suggest the following two-step method when there is no censoring. In the first step, for any given b_1, let $\hat{b}_{2\tau}(b_1)$ be a solution to the minimization problem

$$\min_{b_2} \sum_{i=1}^{n} \rho_\tau(\ln T_{i1}^*(b_1) - X_{2i}'b_2).$$

Then in the second step, $\beta_1(\tau)$ can be estimated by $\hat{\beta}_{1\tau}$, which solves

$$\min_{b_1} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(\ln T_i^* - \ln T(X_i, b_1, \hat{b}_{2\tau}(b_1)))$$

and then we estimate $\beta(\tau)$ by $\hat{\beta}(\tau) = (\hat{\beta}_{1\tau}, \hat{\beta}_{2\tau})$, where $\hat{\beta}_{2\tau} = \hat{b}_{2\tau}\left(\hat{b}_{1\tau}\right)$.

To extend the above two-step method to the censored case, we exploit the equivariance property of the conditional quantile function under monotone transformation. Specifically, we have

$$Q_{\ln T_1(\beta_1(\tau))}(\tau|X) = \min\left\{C_1(\beta_1(\tau)), X_2'\beta_2(\tau)\right\} \tag{2.13}$$

and

$$Q_{\ln T}(\tau|X) = \min\left\{\ln T(X_1, \beta_1(\tau)), \ln C\right\}, \tag{2.14}$$

where

$$T_1(b_1) = \min\left\{T_1^*(b_1), C_1(b_1)\right\} = \int_0^T \exp(-X_1'(s)b_1)ds$$

with

$$C_1(b_1) = \int_0^C \exp(-X_1'(s)b_1)ds.$$

To deal with the problem caused by censoring, we define the subsample selector $d(X, \beta(\tau)) = 1\left\{T(X, \beta(\tau)) < C\right\}$, and note that $d(X, \beta(\tau)) = 1$ if and only if

$$\ln \int_0^C \exp(-X_1'(s)\beta_1(\tau))ds > X_2'\beta_2(\tau),$$

or equivalently,

$$\int_0^C \exp(-X_1'(s)\beta_1(\tau) - X_2'\beta_2(\tau))ds > 1$$

Therefore, when $d(X, \beta(\tau)) = 1$, we have

$$Q_{\ln T_{1i}(\beta_{1\tau})}(\tau|X) = Q_{\ln T_{1i}^*(\beta_{1\tau})}(\tau|X) = X_2'\beta_2(\tau)$$

and

$$Q_{\ln T_i}(\tau|X) = Q_{\ln T_i^*}(\tau|X) = \ln T(X_i, \beta(\tau)).$$

Consequently, we can design an infeasible two-step estimation procedure based on the subsample for which $d_i = d(X_i, \beta(\tau)) = 1$. In the first step, for any given b_1, let $\hat{b}_{2\tau}(b_1)$ be a solution to the minimization problem

$$\min_{b_2} \frac{1}{n}\sum_{i=1}^{n} d_i\rho_\tau(\ln T_{i1}(b_1) - X_{2i}'b_2)$$

and then in the second step, we estimate $\beta_1(\tau)$ by $\hat{b}_{1\tau}$, which solves

$$\min_{b_1} \frac{1}{n}\sum_{i=1}^{n} d_i\rho_\tau(\ln T_i - \ln T(X_i, b_1, \hat{b}_{2\tau}(b_1)))$$

and we estimate $\beta(\tau)$ by $\hat{\beta}(\tau) = (\hat{\beta}_{1\tau}, \hat{\beta}_{2\tau})$, where $\hat{\beta}_{2\tau} = \hat{b}_{2\tau}\left(\hat{b}_{1\tau}\right)$. However, this procedure is clearly not feasible as the subsample selector itself depends on the unknown parameters.

Based on the insights behind the above two-step method and sequential estimation method for censored quantile regression of Chen (2018), we are now ready to develop a feasible, easy-to-implement estimation procedure for the entire family of quantile regression coefficients for censored duration data with time-varying regressors. First, define a grid of τ-values, $S_{L_n} = \{\tau_0 < \tau_1 < \cdots > \tau_{L_n} = \tau_u\}$ and we set $\tau_0 = 0.01$ and τ_u is set to be highest quantile for which there is adequate sample information for reasonably precise estimation of the corresponding quantile coefficients.

Note that when data are censored from below, the amount of information at the top of the conditional duration distribution or at the right tail, is reduced; however, in a typical setting the information at the left tail is not affected. Therefore, if we pick a bottom quantile $\tau_0 = 0.01$, typically it is reasonable to assume that censoring does not affect the 0.01th quantile regression. This assumption is satisfied if censoring level does not exceed 99% for any demographic group; of course, if necessary, we can restrict quantile regression estimation by removing demographic groups for which censoring level exceeds 99%.

We now describe the details of our sequential estimation procedure. For $\tau = \tau_0$, for any given b_1, let $\hat{b}_{2\tau_0}(b_1)$ be a solution to the minimization problem

$$\min_{b_2} \sum_{i=1}^{n} \rho_{\tau_0}(\ln T_{i1}(b_1) - X_{2i}' b_2)$$

then in the second step, we estimate $\beta_1(\tau_0)$ by $\hat{\beta}_{1\tau_0} = \hat{b}_{1\tau_0}$, which solves

$$\min_{b_1} \frac{1}{n} \sum_{i=1}^{n} \rho_{\tau_0}(\ln T_i - \ln T(X_i, b_1, \hat{b}_{2\tau_0}(b_1)))$$

and we estimate $\beta(\tau_0)$ by $\hat{\beta}(\tau_0) = (\hat{\beta}_{1\tau_0}, \hat{\beta}_{2\tau_0})$, where $\hat{\beta}_{2\tau_0} = \hat{b}_{2\tau_0}\left(\hat{b}_{1\tau_0}\right)$.

Once $\hat{\beta}(\tau_0)$ is available, we turn to the estimation of $\beta(\tau_1)$, and in particular we make use of $\hat{\beta}(\tau_0)$ to construct the subsample for the τ_1th quantile regression as $\beta(\tau)$ changes gradually with τ, and thus $\beta(\tau_1) \approx \beta(\tau_0)$ when τ_0 and τ_1 are close to each other. Specifically, define

$$\begin{aligned} \hat{d}_{i\tau_1} &= 1\left\{\ln C_{1i}(\hat{\beta}_{1\tau_0}) - X_{2i}'\hat{\beta}_{2\tau_0} > \delta_n\right\} \\ &= 1\left\{\ln \int_0^C \exp(-X_{1i}'(s)\hat{\beta}_{1\tau_0} - X_{2i}'\hat{\beta}_{2\tau_0}) ds > \delta_n\right\}, \end{aligned}$$

where δ_n is chosen to go to zero slowly as n increases. In particular, when the sample size increases, with large probability, $\left\{\ln C_{1i}(\hat{\beta}_{1\tau_0}) - X_{2i}'\hat{\beta}_{2\tau_0} > \delta_n\right\}$ implies $\left\{\ln C_{1i}(\beta_{1\tau_1}) - X_{2i}'\beta_{2\tau_1} > 0\right\}$ when $\tau_1 - \tau_0 = o(\delta_n)$. Once we have selected the subsample with $\hat{d}_{i\tau_1} = 1$, for any given b_1, we define $\hat{b}_{2\tau}(b_1)$ as a solution to

$$\min_{b_2} \sum_{i=1}^{n} \hat{d}_{i\tau_1} \rho_{\tau_1}(\ln T_{i1}(b_1) - X_{2i}' b_2)$$

and then we proceed to estimate $\beta_1(\tau_1)$ by $\hat{\beta}_{1\tau_1}$, which solves

$$\min_{b_1} \frac{1}{n} \sum_{i=1}^{n} \hat{d}_{i\tau_1} \rho_{\tau_1} (\ln T_i - \ln T(X_i, b_1, \hat{b}_{2\tau_1}(b_1)))$$

and we estimate $\beta(\tau)$ by $\hat{\beta}(\tau) = (\hat{\beta}_{1\tau_1}, \hat{\beta}_{2\tau_1})$, where $\hat{\beta}_{2\tau_1} = \hat{b}_{2\tau_1}\left(\hat{\beta}_{1\tau_1}\right)$.

As part of our sequential estimation procedure, for $j = 1, ..., L_n - 1$, given $\hat{\beta}(\tau_j)$, we define

$$\hat{d}_{i\tau_{j+1}} = 1\left\{\ln C_{1i}(\hat{\beta}_{1\tau_j}) - X'_{2i}\hat{\beta}_{2\tau_j} > \delta_n\right\}$$

and we can estimate $\beta(\tau_{j+1})$ using the above two-step method based on the subsample with $\hat{d}_{i\tau_{j+1}} = 1$. Finally, once we have obtained estimates for quantile regression coefficients on the grid, then for any $\tau \in (\tau_j, \tau_{j+1})$, for $j = 1, ..., L_n$, we can estimate $\beta(\tau)$ with the two-step method based on the subsample with $\hat{d}_{i\tau} = 1$ where $\hat{d}_{i\tau} = 1\left\{\ln C_1(\hat{\beta}_{1\tau_j}) - X'_2\hat{\beta}_{2\tau_j} > \delta_n\right\}$.

We now describe the large sample properties of our estimator. We make the following assumptions.

Assumption 1: $\{T_i^*, X_i, U_i$: $i = 1, 2, ...n\}$ is a random sample generated from model (2.1) where $U_i \sim U(0, 1)$, independent of X.

Assumption 2: The duration time T^* is continuously distributed with its conditional density function $f_{T^*}(\cdot|x)$ uniformly bounded away from 0 in the neighborhood of $Q_{T^*}(\tau|X = x)$, uniform in $\tau \in [\tau_0, \tau_u]$. In addition, $ET^{*2} < \infty$ and $E\sup_t |X(\cdot, t)|^2 < \infty$.

Assumption 3: The parameter space $B \in R^k$, with $k = k_1 + k_2$, is a compact set with $\beta_\tau = \beta(\tau)$ an interior point for $\tau \in [\tau_0, \tau_u]$.

Assumption 4: For any $b \in B$, there exist $c_1, c_2 > 0$ such that

$$c_2 \|b - \beta_\tau\| \geq \|T(\cdot, b_1) - T(\cdot, \beta_\tau)\| \geq c_1 \|b - \beta_\tau\|$$

for any $\tau \in [\tau_0, \tau_u]$, where

$$\|T(\cdot, b) - T(\cdot, \beta_\tau)\| = \left(E[T(X_i, b) - T(X_i, \beta_\tau)]^2 1\{T(X_i, \beta_\tau) < C - \delta_0\}\right)^{1/2}$$

for some $\delta_0 > 0$.

Assumption 5: $\beta(\tau)$ is Lipschitz in $\tau \in [\tau_0, \tau_u]$, with $|\beta(\tau') - \beta(\tau'')| < K|\tau' - \tau''|$ for some constant K.

Assumption 6: $L_n \to \infty$, $L_n = o(n^{1/2})$ and $L_n\delta_n \to \infty$ as $n \to \infty$.

Assumption 7: For $\tau \in [\tau_0, \tau_u]$, the matrices

$$\Omega_\tau = E\left[f_{T^*}(T(X,\beta(\tau))|X)\frac{\partial \ln T(X,\beta(\tau))}{\partial b}\frac{\partial \ln T(X_i,\beta(\tau))}{\partial b'}1\{T(X,\beta(\tau)) < C\}\right]$$

are uniformly positive definite in that

$$\inf_{\tau\in[\tau_0,\tau_u]} \text{mineig}[\Omega_\tau] \geq \lambda_0 > 0$$

for a positive constant λ_0, where mineig$(\cdot)$ denotes the minimum eigenvalue of a matrix, and

$$\lim_{\varepsilon\to 0}\sup_{\tau\in[\tau_0,\tau_u]} \Pr(|T(X,\beta(\tau)) - C| < \varepsilon) \to 0.$$

Assumption 1 describes the data generating mechanism. Assumption 2 contains the continuity assumption on the conditional distribution of T^* on X, and bounded second moments. Assumption 3 is a standard assumption in the literature. Assumption 4 is a global identification condition, which rules out the possibility that $\|T(\cdot, b_n) - T(\cdot,\beta)\| \to 0$ but $\|b_n - \beta\| \geq c_0$ for some positive c_0. Assumption 5 implies that the conditional quantile coefficients evolve slowly, which is a reasonable assumption when the conditional distribution of T^* given X changes continuously; as a result, the quantile regression coefficients change gradually across the entire quantile family. Assumption 6 requires that the sequence δ_n goes to zero slowly whereas the number of grid points increases to infinity but at a faster rate than $1/\delta_n$. Assumption 7 is a local identification condition, similar to that in Powell (1984, 1986, 1991), except that it is a uniform version over the quantile range $[\tau_0, \tau_u]$. From Assumption 7, we can easily deduce that there exists some $\delta_0 > 0$ such that

$$\text{mineig}E\left[f_{T^*}(T(X,\beta(\tau))|X)\frac{\partial \ln T(X,\beta(\tau))}{\partial b}\frac{\partial \ln T(X_i,\beta(\tau))}{\partial b'}1\{T(X,\beta(\tau)) < C - \delta_0\}\right] \geq \lambda_0/2$$

uniformly in τ. The following theorem provides the uniform rate of convergence of our estimator over the grid.

Theorem 1: If Assumptions 1-7 hold, then

$$\max_{j=1,2,\ldots L_n} |\hat\beta(\tau_j) - \beta(\tau_j)| = O\left(n^{-1/2}\ln\ln n\right)$$

almost surely.

Theorem 2: If Assumptions 1-7 hold, then

$$\max_{\tau\in[\tau_l,\tau_u]} |\hat\beta(\tau) - \beta(\tau)| = O\left(n^{-1/2}\ln\ln n\right)$$

almost surely and

$$\sqrt{n}\left(\hat\beta(\tau) - \beta(\tau)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{\tau i} + o_p(1)$$

uniformly in $\tau \in [\tau_0, \tau_u]$, where $\phi_{\tau i}$ is defined in the appendix, and $\sqrt{n}\left(\hat{\beta}(\cdot) - \beta(\cdot)\right)$ converges to a mean zero Gaussian process $G(\cdot)$ for $\tau \in [\tau_0, \tau_u]$ with covariance function

$$EG(\tau)G(\tau')' = E\left[\phi_{\tau i}\phi'_{\tau i}\right].$$

In order to conduct large sample statistical inference, it is important to have consistent estimators for the asymptotic covariance matrices. Given the complex nature of our estimators and the fact that nonparametric kernel estimation is typically required in a quantile regression setting Here we adopt resampling methods. Similar to Chen et al. (2003), Chernozhukov et al (2015) and Chen (2018), we consider the multiplier bootstrap. Specifically, let $\{\xi_i\}_1^n$ be i.i.d. draws of positive random variables with $E\xi = \text{Var}(\xi) = 1$, independent of the data. For a fixed τ, define

$$\hat{d}_{i\tau} = 1\left\{\ln \int_0^{C_i} \exp(-X'_{1i}(s)\hat{\beta}_1(\tau) - X'_{2i}\hat{\beta}_2(\tau))ds > \delta_n\right\}$$

where $\hat{\beta}(\tau)$ is our estimator for $\beta(\tau)$ based on a given sample, which is fixed in the resampling process. We also follow the two-step approach in the resampling stage. For a given b_1, $\hat{b}^*_{2\tau}(b_1)$ solves

$$\min_{b_2} \sum_{i=1}^{n} \hat{d}_{i\tau}\xi_i\rho_\tau(\ln T_{i1}(b_1) - X'_{2i}b_2)$$

then we define $\hat{\beta}^*_{1\tau}$ as a solution to the following minimization problem,

$$\min_{b_1} \frac{1}{n}\sum_{i=1}^{n} \hat{d}_{i\tau}\xi_i\rho_\tau(\ln T_i - \ln T(X_i, b_1, \hat{b}^*_{2\tau}(b_1)))$$

and define $\hat{\beta}^*_{2\tau} = \hat{b}^*_{2\tau}\left(\hat{\beta}^*_{1\tau}\right)$.

Note that we define the subsample selection indicator in terms of the estimator $\hat{\beta}(\tau)$ for the original data, thus fixed in the resampling process. Therefore, estimation in the resampling stage does not involve a sequential process. We will show that the asymptotic distribution of $\sqrt{n}\left(\hat{\beta}(\cdot) - \beta(\cdot)\right)$ can be approximated by the limiting distribution of $\sqrt{n}\left(\hat{\beta}^*(\cdot) - \hat{\beta}(\cdot)\right)$. We make the following additional assumption.

Assumption 8: The weights $\{\xi_i\}_1^n$ are i.i.d. draws from a positive random variable ξ with $E\xi = \text{Var}(\xi) = 1$ and it possesses $2 + c_0$ moment for some $c_0 > 0$ that lives in a probability space $(\Omega_\xi, F_\xi, P_\xi)$, independent of the data $\{T_i^*, X_i\}$.

Theorem 3: If Assumptions 1-8 hold, then conditional on the data, $\sqrt{n}\left(\hat{\beta}^*(\cdot) - \hat{\beta}(\cdot)\right)$ converges to a mean zero Gaussian process $G(\cdot)$ for $\tau \in [\tau_0, \tau_u]$ with covariance function

$$EG(\tau)G(\tau')' = E\left[\phi_{\tau i}\phi'_{\tau i}\right]$$

Following the practice in the resampling literature, the distribution of $\sqrt{n}\left(\hat{\beta}^*(\cdot)-\hat{\beta}(\cdot)\right)$ conditional on the data can be approximated through numerical simulation. For $m=1,2,...,M$, define $\hat{\beta}^*_{m1\tau}$ as the solution to

$$\min_{b_1}\frac{1}{n}\sum_{i=1}^{n}\xi_{mi}\hat{d}_{i\tau}\rho_\tau(\ln T_i-\ln T(X_i,b_1,\hat{b}^*_{m2\tau}(b_1)))$$

where $\hat{b}^*_{m2\tau}(b_1)$ solves

$$\min_{b_2}\sum_{i=1}^{n}\xi_{mi}\hat{d}_{i\tau}\rho_\tau(\ln T_{i1}(b_1)-X'_{2i}b_2)$$

and we define $\hat{\beta}^*_{m2\tau}=\hat{b}^*_{m2\tau}\left(\hat{\beta}^*_{m1\tau}\right)$, where $(\xi_{m1},\xi_{m2}...,\xi_{mn})$ forms a random sample of size n for the mth replication, drawn from a distribution satisfying Assumption 8, independent of the data. Then the conditional distribution of $\sqrt{n}\left(\hat{\beta}^*(\cdot)-\hat{\beta}(\cdot)\right)$ can be approximated by the the empirical distribution of $\sqrt{n}\left(\hat{\beta}^*_m(\cdot)-\hat{\beta}(\cdot)\right)$ for a large M.

3 Instrumental Variables Censored Quantile Regression with Time-Varying Regressors

In this section we extend the quantile regression model in the previous section by allowing for endogenous regressors. For simplicity we focus on the case with a single binary endogenous regressor $D(\cdot)$ with an instrument Z, which can easily be extended to the general case with multiple endogenous regressors, as in Chen (2018). The data is generated from the model

$$\int_0^{T^*}\exp(-D(s)\gamma(U)-X'_1(s)\beta_1(U)-X'_2\beta_2(U))ds=1 \tag{3.1}$$

where T^* is the duration time, X_1 and X_2 are k_1 and k_2 dimensional time-varying and time-invariant exogenous regressors, U has a uniform $U(0,1)$ distribution, independent of (X,Z).

Note our model reduces to the structural quantile regression model of Chernozhukov and Hansen (2006, 2008) when all regressors are time-invariant. Following Chernozhukov and Hansen (2006, 2008), we consider the potential outcome framework such that conditional on $X=x$, the potential outcome satisfies $T^*_d=q(d,x,U_d)$, where U_d represents ranking of the unobserved individual characteristic with the same observed characteristics x and treatment d, and the conditional structural τ-quantile function $q(d,x,\tau)$, which is strictly increasing in τ for any give (d,x), satisfies

$$\int_0^{q(d,x,\tau)}\exp(-d(s)\gamma(\tau)-x'_1(s)\beta_1(\tau)-x'_2\beta_2(\tau))ds=1. \tag{3.2}$$

Similar to Chernozhukov and Hansen (2006, 2008), we make the following assumption:

Assumption 1':

A1. Potential Outcome. Given $X = x$, for each d, $T_d^* = q(d, x, U_d)$, where $U_d \sim U(0,1)$ and $q(d, x, \tau)$ is strictly increasing in τ.

A2. Independence. Given $X = x$, $\{U_d\}$ is independent of Z.

A3. Selection. Given $X = x$, $Z = z$, for some unknown function δ and random vector v, $D = \delta(z, x, v)$.

A4. Rank Similarity. For each d and d', given (v, X, Z), $U_d \sim U_{d'}$.

A5. Observed variables consist of $T = \min\{T^*, C\} = \min\{q(D, X, U_D\}, C\}$, $D = \delta(Z, X, v)$, X, Z.

Note that Assumption 1' largely follows Assumption 1 in Chernozhukov and Hansen (2006) except that we allow for censoring here. Then, similar to Chernozhukov and Hansen (2006, 2008), we adopt a linear quantile specification (3.1) and it is straightforward to show that

$$\Pr(T^* < q(D, X, \tau))|X, Z) = \tau \tag{3.3}$$

for $\tau \in (0, 1)$, which serves as the basis for our estimation method. Similar to the exogenous case considered in the previous section, we assume that there is a bottom quantile τ_0, say, $\tau_0 = 0.01$, such that $q(D, X, \tau_0) < C$ almost surely.

We focus on the estimation of the quantile regression coefficient process $\theta(\tau) = (\alpha(\tau), \beta'(\tau))'$ for $\tau \in [\tau_0, \tau_u]$. Again, similar to the exogenous case, we adopt a sequential approach, and define a grid of τ-values, $S_{L_n} = \{\tau_0 < \tau_1 < \cdots < \tau_{L_n} = \tau_u\}$.

We first consider the estimation of $\theta(\tau_0)$. Similar to Chernozhukov and Hansen (2006, 2008) and Chen (2018), as well as the estimation approach for the exogenous case in the previous section, we adopt a two-step method for each given quantile. For given $\theta_1 = (b_1, \gamma)$, we define $\hat{b}_{2\tau_0}(\theta_1)$ be a solution to the minimization problem

$$\min_{b_2} \sum_{i=1}^{n} \rho_{\tau_0}(\ln T_{i1}(\theta_1) - X_{2i}' b_2)$$

where, with a slight abuse of notation,

$$T_1(\theta_1) = \min\{T_1^*(\theta_1), C_1(\theta_1)\} = \int_0^T \exp(-X_1'(s)b_1 - D(s)\gamma)ds$$

with

$$T_1^*(\theta_1) = \int_0^{T^*} \exp(-X_1'(s)b_1 - D(s)\gamma)ds$$

and

$$C_1(\theta_1) = \int_0^C \exp(-X_1'(s)b_1 - D(s)\gamma)ds.$$

Similar to (2.12), we have

$$Q_{\ln T_1(\theta_1(\tau_0))}(\tau_0|X) = \min\left\{C_1(\theta_1(\tau_0)), X_2'\beta_2(\tau_0)\right\} = X_2'\beta_2(\tau_0). \tag{3.4}$$

Then in the second step, we estimate $\theta_1(\tau_0)$ by $\hat{\theta}_{1\tau_0}$, which solves

$$\min_{\theta_1} ||\frac{1}{n}\sum_{i=1}^{n}\varphi_{\tau_0}(\ln T_1(\theta_1) - X_i'\hat{\beta}_{2\tau_0}(\theta_1))(\bar{X}_{1i}', X_{2i}, Z_i)'||$$

where $\varphi_\tau(u) = (1\{u<0\} - \tau)$, $\left(\bar{X}_{1i}', X_{2i}', Z_i\right)'$ are appropriate instruments. Consequently we estimate $\theta(\tau_0)$ by $\left(\hat{\theta}_{1\tau_0}', \hat{\beta}_{2\tau_0}'\right)'$, where $\hat{\beta}_{2\tau_0} = \hat{b}_{2\tau_0}\left(\hat{\theta}_{1\tau_0}\right)$.

Next, we consider the estimation of $\theta(\tau_1) = (\theta_1'(\tau_1), \beta_2'(\tau_1))'$. Define

$$d(X,\theta,\delta) = 1\{\ln\int_0^C \exp(-\max\{\gamma,0\} - X_1'(s)\beta_1 - X_2'\beta_2)ds > \delta\}.$$

Then from (3.2) we can infer that $d(X,\theta(\tau),0) = 1$ implies that $C > Q(D,X,\tau)$; in other words, for observation i, with $d(X_i,\theta(\tau),0) = 1$, the τth structural quantile function is not affected by censoring. Consequently, for the subsample with $d(X,\theta(\tau),0) = 1$, we can ignore the presence of censoring for the purpose of conducting τth quantile regression. Therefore, given $\hat{\theta}(\tau_0)$, we define the subsample selector $\hat{d}_i(\tau_0) = d(X_i,\hat{\theta}(\tau_0),\delta_n)$ and estimation of $\theta(\tau_1)$ is based on the subsample with $\hat{d}_i(\tau_0) = 1$. Specifically, for a given θ_1, define $\hat{b}_{2\tau_1}(\theta_1)$ as a solution to the following minimization problem

$$\min_{b_2}\sum_{i=1}^{n}\hat{d}_i(\tau_0)\rho_{\tau_1}(\ln T_{i1}(\theta_1) - X_{2i}'b_2)$$

and then define our estimator for $\theta_1(\tau_1)$ by $\hat{\theta}_1(\tau_1)$, which solves

$$\min_{\theta_1}||\frac{1}{n}\sum_{i=1}^{n}\hat{d}_i(\tau_0)\varphi_{\tau_1}(\ln T_{i1}(\theta_1) - X_{2i}'\hat{b}_2(\theta_1,\tau_1))(\bar{X}_{1i}', X_{2i}, Z_i)'||$$

and we define $\hat{\theta}(\tau_1) = \left(\hat{\theta}_1(\tau_1), \hat{\beta}_2'(\tau_1)\right)'$ as the estimator for $\theta(\tau_1)$, where $\hat{\beta}_2(\tau_1) = \hat{b}_{2\tau_1}(\hat{\theta}_1(\tau_1))$.

Next, given $\hat{\theta}(\tau_j)$ for any $j = 1,2,...,L_n - 1$, define $\hat{b}_{2\tau_{j+1}}(\theta_1)$ as a solution to the minimization problem

$$\min_{b_2}\sum_{i=1}^{n}\hat{d}_i(\tau_j)\rho_{\tau_{j+1}}(\ln T_{i1}(\theta_1) - X_{2i}'b_2)$$

where $\hat{d}_i(\tau_j) = d(X_i,\hat{\theta}(\tau_j),\delta_n)$, and define our estimator for $\theta_1(\tau_{j+1})$ by $\hat{\theta}_1(\tau_{j+1})$, which solves

$$\min_{\theta_1}||\frac{1}{n}\sum_{i=1}^{n}\hat{d}_i(\tau_j)\varphi_{\tau_{j+1}}(\ln T_{i1}(\theta_1) - X_{2i}'\hat{b}_2(\theta_1,\tau_{j+1}))(\bar{X}_{1i}', X_{2i}, Z_i)'||$$

and define $\hat{\theta}(\tau_{j+1}) = \left(\hat{\theta}_1'(\tau_{j1}), \hat{\beta}_2'(\tau_{j+1})\right)'$ as the estimator for $\theta(\tau_{j+1})$, where $\hat{\beta}_2(\tau_{j+1}) = \hat{b}_{2\tau_{j+1}}(\hat{\theta}_1(\tau_{j+1}))$.

Finally, for any $\tau \in (\tau_j, \tau_{j+1})$, for $j = 1, ..., L_n$, define $\hat{b}_{2\tau}(\theta_1)$ as a solution to the minimization problem

$$\min_{\beta \in B} \frac{1}{n} \sum_{i=1}^{n} \hat{d}_i(\tau_j) \rho_\tau (\ln T_{i1}(\theta_1) - X_{2i}' b_2)$$

and then $\hat{\theta}_1(\tau)$ solves

$$\min_{\alpha \in A} || \frac{1}{n} \sum_{i=1}^{n} \hat{d}_i(\tau_j) \varphi_\tau (\ln T_{i1}(\theta_1) - X_{2i}' \hat{b}_2(\theta_1, \tau_{j+1}))(\bar{X}_{1i}', X_{2i}', Z_i))' ||$$

and define $\hat{\theta}(\tau) = \left(\hat{\theta}_1'(\tau), \hat{\beta}_2'(\tau)\right)'$ as the estimator for $\theta(\tau)$, where $\hat{\beta}_2(\tau) = \hat{b}_{2\tau}(\hat{\theta}_1(\tau))$.

We make the following additional assumptions.

Assumption 2': The duration time T^* is continuously distributed with its conditional density function $f_{T^*}(\cdot|x, d)$ uniformly bounded away from 0 in the neighborhood of $Q(d, \tau, x)$, uniform in $\tau \in [\tau_0, \tau_u]$. In addition, $E\,|T^*|^2 < \infty$ and $E \sup_t |X(\cdot, t)|^2 < \infty$.

Assumption 3': The parameter space $\Theta \in R^{k+1}$ is a compact set with $\theta\,(\tau)$ an interior point.

Let

$$U_c(\theta\,(\tau)\,, \tau) = E\left[d_0(X, \theta(\tau)) \varphi_\tau (\ln T_{i1}(\theta_1\,(\tau)) - X_2' \beta_2(\tau))(\bar{X}_1', X_2', Z)'\right]$$

where $d_0(X, \theta) = d(X, \theta, 0)$.

Assumption 4': For any $\varepsilon > 0$, $\inf_{\tau \in [\tau_0, \tau_u]} \inf_{||\theta - \theta(\tau)|| > \varepsilon} U_c(\theta\,(\tau)\,, \tau) > 0$.

Assumption 5': $\theta\,(\tau)$ is Lipschitz in $\tau \in [\tau_0, \tau_u]$, with $|\theta\,(\tau') - \theta\,(\tau'')\,| < K\,|\tau' - \tau''|$ for a constant K.

Assumption 6': $L_n \to \infty$, $L_n = o(n^{1/2})$ and $L_n \delta_n \to \infty$ as $n \to \infty$.

Assumption 7': The matrices $V\,(\tau)$ for $\tau \in [\tau_0, \tau_u]$ are uniformly nonsingular in that

$$\inf_{\tau \in [\tau_l, \tau_u]} \min \mathrm{eig}(V(\tau) V(\tau)') \geq \lambda_0 > 0$$

for a positive constant λ_0, and

$$\lim_{\varepsilon \to 0} \sup_{\tau \in [\tau_0, \tau_u]} \Pr\left(|q(1, X, \max\{\gamma(\tau), 0\}, \beta_1\,(\tau)\,, \beta_2\,(\tau)\,, \tau) - C| < \varepsilon\right) \to 0.$$

where

$$V\,(\tau) = E\left[d_0(X, \theta(\tau)) f_{T^*}(q(D, X, \theta\,(\tau))|X, Z, D) \begin{pmatrix} \bar{X}_1' & X_2 & Z \end{pmatrix}' \frac{\partial q(D, X, \theta, \tau)}{\partial \theta'}\right].$$

with $q(d, x, \theta, \tau)$ satisfying

$$\int_0^{q(d,x,\theta,\tau)} \exp(-d(s)'\gamma - x_1'(s)b_1 - x_2'b_2)ds = 1.$$

Assumptions 1'-7' are similar to Assumptions 1-7 in the previous section. The global and local identification conditions are similar to those in Chernozhukov and Hansen (2006, 2008) and Chen (2018).

Theorem 4: If Assumptions 1'-7' hold, then

$$\max_{j=1,2,\ldots,L_n} |\hat{\theta}(\tau_j) - \theta(\tau_j)| = O\left(n^{-1/2}\ln\ln n\right)$$

almost surely.

The next theorem describes the uniform consistency and weak convergence of the quantile regression coefficient process over $\tau \in [\tau_0, \tau_u]$.

Theorem 5: If Assumptions 1'-7' hold, then

$$\max_{\tau\in[\tau_0,\tau_u]} |\hat{\theta}(\tau) - \theta(\tau)| = O\left(n^{-1/2}\ln\ln n\right)$$

almost surely and

$$\sqrt{n}\left(\hat{\theta}(\tau) - \theta(\tau)\right) = J_\theta(\tau)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi_{\theta i}(\tau) + o_p(1)$$

uniformly in $\tau \in [\tau_0, \tau_u]$, where

$$J_\theta(\tau) = E\left[d_i(\tau)f_{T^*}(q(D_i, X_i, \theta(\tau))|X_i, D_i, Z_i)(\bar{X}_{1i}', X_{2i}', Z_i'))'\frac{\partial q(D_i, X_i, \theta(\tau))}{\partial\theta'}\right]$$

and

$$\Phi_{\theta i}(\tau) = d(X_i, \theta(\tau))\varphi_\tau(T_i - q(D_i, X_i, \theta(\tau))(\bar{X}_{1i}', X_{2i}', Z_i'))'$$

and thus $\sqrt{n}\left(\hat{\theta}(\tau) - \theta(\tau)\right)$ converges to a mean zero Gaussian process $G_\theta(\cdot)$ for $\tau \in [\tau_l, \tau_u]$ with covariance function

$$EG_\theta(\tau)G_\theta(\tau')' = J_\theta(\tau)^{-1}E\left[\Phi_{\theta i}(\tau)\Phi_{\theta i}(\tau')'\right]J_\theta(\tau')^{-1\prime}.$$

As for the exogenous case, we use the resampling method for the purpose of making statistical inference. Again, let $\{\xi_i\}_1^n$ be i.i.d. random draws with $E\xi = \text{Var}(\xi) = 1$, independent of the data. For any given τ, we define $\hat{b}_{2\tau}^*(\theta_1)$ as a solution to the minimization problem

$$\min_{\beta\in B}\frac{1}{n}\sum_{i=1}^{n}\xi_i\hat{d}_i(\tau)\rho_\tau(\ln T_{i1}(\theta_1) - X_{2i}'b_2)$$

and then $\hat{\theta}_1^*(\tau)$ solves

$$\min_{\alpha\in A} ||\frac{1}{n}\sum_{i=1}^{n}\xi_i\hat{d}_i(\tau)\varphi_\tau(\ln T_{i1}(\theta_1)-X_{2i}'\hat{b}_2(\theta_1,\tau_{j+1}))(\bar{X}_{1i}',X_{2i}',Z_i))'||$$

and define $\hat{\theta}^*(\tau) = \left(\hat{\theta}_1^*(\tau),\hat{\beta}_2^{*\prime}(\tau)\right)'$, where $\hat{\beta}_2^*(\tau) = \hat{b}_{2\tau}^*(\hat{\theta}_1^*(\tau))$. As in the exogenous case, the subsample selector is based on the original estimator, thus fixed in the resampling process. Then the asymptotic distribution of $\sqrt{n}\left(\hat{\theta}(\cdot)-\theta(\cdot)\right)$ can be approximated by the limiting distribution of $\sqrt{n}\left(\hat{\theta}^*(\cdot)-\hat{\theta}(\cdot)\right)$. We make the following additional assumption.

Assumption 8': The weights $\{\xi_i\}_1^n$ are i.i.d. draws from a positive random variable ξ with $E\xi$ =Var$(\xi)=1$ and it possesses $2+c_0$ moment for some $c_0>0$ that lives in a probability space (Ω_ξ,F_ξ,P_ξ), independent of the data $\{Y_i,X_i,D_i,Z_i\}$.

Theorem 6: If Assumptions 1'-8' hold, then conditional on the data,$\sqrt{n}\left(\hat{\theta}^*(\cdot)-\hat{\theta}(\cdot)\right)$ converges to a mean zero Gaussian process $G_\theta(\cdot)$ for $\tau\in[\tau_0,\tau_u]$ with covariance function

$$EG_\theta(\tau)G_\theta(\tau')' = J_\theta(\tau)^{-1}E\left[\Phi_{\theta i}(\tau)\Phi_{\theta i}(\tau')'\right]J_\theta(\tau')^{-1\prime}.$$

Our instrumental variable censored quantile regression estimator with time-varying regressors and the instrumental variables censored quantile regression estimator in Chen (2012) have very similar structure. Both are sequential estimation procedures, and at the each stage of the sequential estimation process, the first step of both estimators involves solving some linear quantile regression, whereas both of the second step estimators are based on solving quantile moment equations. As a result, the proofs of Theorems 4-6 are very similar to those of Theorems 4-6 in Chen (2018), and therefore the details are omitted here.

4 Monte Carlo Experiments

In this subsection, we report the results of a set of Monte Carlo experiments to illustrate the finite sample performance of our estimators. We consider both homogenous and heterogenous designs with data subject to fixed censoring.

First, we consider the case where all regressors are exogenous. The duration time T^* is generated by

$$\int_0^{T^*} e^{-\beta_0(U)-\beta_1(U)^T\tilde{X}-\beta_2(U)X(t)}dt = 1$$

where $U\sim U[0,1]$, and

$$X(t)=\begin{cases} X_0 \text{ for } t\le t_1\\ X_1 \text{ for } t>t_1\end{cases}$$

where $X_0 \sim 1 + 2U_0$, $X_1 \sim X_0 + U_1, \tilde{X} = (U_2, E)$; here U_0, U_1, U_2 and U are *iid* $U[0,1]$, and E is *iid* exponential with scale parameter 1. For the homogeneous design, we set $t_1 = 7$, $\beta_0(U) = U, \beta_1(U) = (1,0), \beta_2(U) = 0.5$, thus the quantile coefficients are parallel. For the heterogeneous design, we set $t_1 = 9$, $\beta_0(U) = U, \beta_1(U) = (1,0), \beta_2(U) = U$. By varying the censoring constant, we consider cases with 15% and 30% censoring.

We consider the estimation of the quantile regression coefficients for $\tau = 0.3$, 0.5 and 0.7 respectively. For each design we report Bias, and standard deviation (SD), estimated standard deviation (est.SD) and the coverage probabilities of the 95% (CP95) confidence intervals using the resampling method proposed in the paper. Sample sizes were chosen to be 200 and 800, respectively, with 1000 replications. The resampling size is set 500. Table 1 reports the results for the homogeneous designs. Note that our estimator performs well for all three quantiles and both censoring levels; in fact there is little bias even for $\tau = 0.7$ with 30% censoring when $n = 200$. Estimated standard deviations are quite close to the true standard deviations and the estimated confidence intervals, in general, have very good coverage properties. When the sample size increases to 800, the standard deviations are roughly cut by half. Table 2 reports the results for the heterogeneous designs with fixed censoring. Compared with the homogenous designs, while the estimator still performs satisfactorily, we observe significant increase in biases and standard deviations, for the $\tau = 0.7$ and the censoring level equal to 30%. Except for this particular combination, the estimated confidence intervals continue to have desirable coverage properties. In general, the situation improves when the sample size increases to 800.

For the case when an endogenous regressor is present, the duration time T^* is generated according to the model

$$\int_0^T e^{-\beta_0(U) - \beta_1^T X_1 - \beta_2(U) D(t)} dt = 1$$

where

$$D(t) = \begin{cases} D \text{ for } t \le t_1 \\ 0 \text{ for } t > t_1 \end{cases}$$

with $D = (Z + U + W > 0), Z = 2U_0 - 1, X_1 = (U_1, U_2)$; here U_0, U_1, U_2 and U are *iid* $U[0,1]$, independent of W, which is drawn from the standard normal distribution. As in the exogenous case, we consider both homogenous and heterogenous designs separately. For the homogeneous design, we set $t_1 = 7$, $\beta_0(U) = U$, $\beta_1 = (1,1)^T, \beta_2(U) = 0.5$, and for the heterogeneous design, we set $t_1 = 5$, $\beta_0(U) = U$, $\beta_1 = (1,1)^T, \beta_2(U) = U - 0.5$. The censoring constant C is chosen so that the designs we have adopted have about 15% and 30% censoring respectively.

Table 3 reports the results for the homogenous design. Similar to the exogenous case, our estimator performs well for all combination of τ and censoring levels, even for $n = 200$. Results for the heterogenous design is reported in Table 4. While our estimator performs well generally in terms of bias, coverage probabilities, for the case with $\tau = 0.7$, 30% censoring and $n = 200$, the differences between the estimated standard deviations and the true standard deviations can differ up to about 20%, leading to some undercoverage. Once n is increased to 800, our estimator performs very well across the board.

5 Conclusion

In economic duration analysis, popular econometric models such as the Cox proportional hazards model, the mixed proportional hazards model and various extensions are highly restrictive in allowing how regressors affect the conditional duration distribution. In addition, endogeneity such as selective compliance is also difficult to accommodate. Censoring and time-varying regressors are fundamental features of duration data. Endogeneity is common in applied duration analysis. Fitzenberger and Wilke (2005) and Koenker and Geling (2001), among others, argued that quantile regression model is particularly well-equipped to deal with censoring and provides a flexible semiparametric approach to model the conditional duration distribution. In this paper, we develop a quantile regression framework that allows for censoring, time-varying regressors and endogeneity, and we propose an easy-to-implement two-step quantile regression estimator. Monte Carlo experiments indicate that our estimator performs well in finite samples.

Appendix

We first present lemma A1, which is useful in analyzing the asymptotic properties of our two-step estimator. For any τ, let $\hat{b}_{2\tau}(\bar{b}, b_1, \tau, \delta)$ denote a minimizer of

$$\min_{b_2} \sum_{i=1}^{n} d_i(\bar{b}, \delta)\rho_\tau(\ln T_{i1}(b_1) - X_{2i}' b_2)$$

where $d_i(\bar{b}, \delta) = 1\left\{T(X_i, \bar{b}) < C - \delta\right\}$, $T_1(b_1)$ is defined in the main text.

Lemma A1: For any $b_1 = \beta_{1\tau} + r_n$, where $r_n = o\left(\delta_n^2\right)$ and $\bar{b} = \beta_\tau + o\left(\delta_n\right)$. Under Assumptions 1-5,

$$\hat{b}_{2\tau}(\bar{b}, b, \tau, \delta) - \beta_{2\tau} = \min\left\{r_n^{1/2} + \left(n^{-1/2}\ln\ln n\right)^{1/2}\right\}$$

almost surely, and furthermore

$$(\hat{b}_{2\tau}(b_1) - \beta_{2\tau}) = \Gamma_{\tau 22}^{-1} S_{n2}(\beta_\tau, \beta_\tau, \tau, \delta_n) + \Gamma_{\tau 22}^{-1}\Gamma_{\tau 21}(b_1 - \beta_{1\tau}) + o_p\left(n^{-1/2}\right)$$

uniformly in $b_1 = \beta_{1\tau} + r_n$, $\bar{b} = \beta_\tau + o\left(\delta_n\right)$ and $\tau \in [\tau_0, \tau_u]$, where

$$\Gamma_{\tau 21} = \frac{\partial}{\partial b_1'} S_2(\beta_\tau, \beta_{1\tau}, \beta_{2\tau}, 0) = E\left[f_{T^*}\left(T(X, \beta_\tau)|X\right) X_2 \frac{\partial}{\partial b_1'} T(X, \beta_\tau) 1\left\{T(X, \beta_\tau) < C\right\}\right]$$

and

$$\Gamma_{\tau 22} = \frac{\partial}{\partial b_2'} S_2(\beta_\tau, \beta_{1\tau}, \beta_{2\tau}, 0) = E\left[f_{T^*}\left(T(X, \beta_\tau)|X\right) X_2 X_2' 1\left\{T(X, \beta_\tau) < C\right\}\right].$$

with

$$S_2(\bar{b}, b, \tau, \delta) = E\left[S_{n2}(\bar{b}, b, \tau, \delta)\right],$$

and

$$S_{n2}(b, \bar{b}, \tau, \delta_n) = \frac{1}{n}\sum_{i=1}^{n} d_i(\bar{b}, \delta_n)\left(1\left\{T_i < T(X_i, b)\right\} - \tau\right) X_{2i}.$$

Proof: First note that by Assumptions 1-3, we have

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n} d_i(\bar{b}, \delta_n)\rho_\tau(\ln T_{i1}(b_1) - X_{2i}' b_2) - \rho_\tau(\ln T_{i1}(b_1) - X_{2i}'\beta_{2\tau}) \\
= \;& O\left(r_n\right) + \frac{1}{n}\sum_{i=1}^{n} d_i(\bar{b}, \delta_n)\rho_\tau(\ln T_{i1}(\beta_{1\tau}) - X_{2i}' b_2) - \rho_\tau(\ln T_{i1}(\beta_{1\tau}) - X_{2i}'\beta_{2\tau}) \\
= \;& O\left(r_n\right) + \frac{1}{n}\sum_{i=1}^{n} d_i(\bar{b}, \delta_n)\rho_\tau(\varepsilon_{1\tau} - X_{2i}'(b_2 - \beta_{2\tau})) - \rho_\tau(\varepsilon_{1\tau})
\end{aligned}$$

uniformly in $b_1 = \beta_{1\tau} + o(r_n)$, $\bar{b} = \beta_\tau + o(\delta_n)$, and $\tau \in [\tau_0, \tau_u]$, where $\varepsilon_{1\tau} = \ln T_{i1}(\beta_{1\tau}) - X_{2i}'\beta_{2\tau}$. Then, similar to the argument in the proof of Theorem 1 below, we can show that

$$\hat{b}_{2\tau}(\bar{b}, b_1, \tau, \delta) - \beta_{2\tau} = O_p\left(r_n^{1/2} + n^{-1/2}\ln n\right)$$

almost surely, uniformly in $b_1 = \beta_{1\tau} + o(r_n)$ and $\tau \in [\tau_0, \tau_u]$ almost surely. Then, similar to Powell (1991), Honoré (1992) and Chen (2018), the estimating equation for $\hat{b}_{2\tau}(\bar{b}, b_1, \tau, \delta_n)$ satisfies

$$\begin{aligned} & S_{n2}(b_1, \hat{b}_{2\tau}(\bar{b}, b_1, \tau, \delta_n), \bar{b}, \tau, \delta_n) \\ = & \frac{1}{n}\sum_{i=1}^{n} d_i(\bar{b}, \delta_n)\left(1\left\{T_i < T(X_i, b_1, \hat{b}_{2\tau}(\bar{b}, b_1, \tau, \delta_n))\right\} - \tau\right) X_{2i} \\ = & o_p\left(n^{-1/2}\right). \end{aligned}$$

Furthermore, it is easy to verify that the class of functions

$$\{1\{T(X, \bar{b}) < C - \delta\}\left(1\{T < T(X, b)\} - \tau\right) X_2 \colon \bar{b} \in R^k, b \in R^k, \tau \in (0,1), \delta \in R\}$$

is Euclidean (Pakes and Pollard 1989) with a square integrable envelope, thus we

$$\begin{aligned} o_p\left(n^{-1/2}\right) &= S_{n2}(b_1, \hat{b}_{2\tau}(\bar{b}, b_1, \tau, \delta_n), \bar{b}, \tau, \delta_n) \\ &= \left[S_2(b_1, \hat{b}_{2\tau}(\bar{b}, b_1, \tau, \delta_n), \beta_\tau, \tau, \delta_n) - S_2(\beta_\tau, \beta_{1\tau}, \beta_{2\tau}, \tau, \delta_n)\right] \\ &\quad + S_{n2}(\beta_\tau, \beta_{1\tau}, \beta_{2\tau}, \tau, \delta_n) + o_p\left(n^{-1/2}\right) \\ &= \Gamma_{\tau 21}(b_1 - \beta_{1\tau}) + \Gamma_{\tau 22}(\hat{b}_{2\tau}(\bar{b}, b_1, \tau, \delta_n) - \beta_{2\tau}) + o_p\left(n^{-1/2}\right) \\ &\quad + S_{n2}(\beta_\tau, \beta_{1\tau}, \beta_{2\tau}, \tau, \delta_n) \end{aligned}$$

uniformly in $b_1 = \beta_{1\tau} + o(\delta_n)$, $\bar{b} = \beta_\tau + o(\delta_n)$, and $\tau \in [\tau_0, \tau_u]$. Therefore,

$$(\hat{b}_{2\tau}(\bar{b}, b_1, \tau, \delta_n) - \beta_{2\tau}) = \Gamma_{\tau 22}^{-1} S_{n2}(\beta_\tau, \beta_{1\tau}, \beta_{2\tau}, \tau, \delta_n) + \Gamma_{\tau 22}^{-1}\Gamma_{\tau 21}(b_1 - \beta_{1\tau}) + o_p\left(n^{-1/2}\right),$$

uniformly in $b_1 = \beta_{1\tau} + o(\delta_n)$, $\bar{b} = \beta_\tau + o(\delta_n)$, and $\tau \in [\tau_0, \tau_u]$.

Proof of Theorem 1: Similar to the proof of Theorem 1 in Chen (2018), we proceed in two steps.

Step 1. In this step we prove the uniform consistency of $\hat{\beta}(\tau_j)$ and its rate of convergence, for $j = 1, 2, ..., L_n$, given $\hat{\beta}(\tau_0) - \beta(\tau_0) = O\left(n^{-1/2}\ln\ln n\right)$ almost surely.

First, we establish some rate of convergence results for some empirical processes. Define a class of functions

$$\{g(X, T, \bar{b}, b, \tau, \delta) \colon \bar{b} \in R^k, b \in R^k, \tau \in (0,1), \delta \in R\}$$

where

$$g(X,T,\bar{b},b,\tau,\delta)=\rho_\tau(T-T(X,b))1\left\{T(X,\bar{b})<C-\delta\right\}$$

Then, similar to the arguments in the proof of Theorem 3 in Chen et al. (2003), we can show that the above class of functions is Euclidean (Pakes and Pollard 1989) with a bounded envelope. Hence, by the law of the iterated logarithm for the Donsker class (Arcones, 1993), we have

$$G_n(\bar{b},b,\tau,\delta)-G(\bar{b},b,\tau,\delta)=O\left(n^{-1/2}\ln\ln n\right) \tag{R1}$$

almost surely, uniformly in $\left(\bar{b},b,\tau,\delta\right)$, where

$$G_n(\bar{b},b,\tau,\delta)=\frac{1}{n}\sum_{i=1}^{n}g(X_i,T_i,\bar{b},b,\tau,\delta)$$

and

$$G(\bar{b},b,\tau,\delta)=E\left[g(X_i,T_i,\bar{b},b,\tau,\delta)\right].$$

Next, define

$$S_n(\bar{b},b,\tau,\delta)=\frac{1}{n}\sum_{i=1}^{n}\varphi_\tau(T_i-T(X_i,b))\Delta(X,b)1\left\{T(x,\bar{b})<C-\delta\right\}$$

and

$$Q_n(\bar{b},b,\tau,\delta)=\frac{1}{n}\sum_{i=1}^{n}\frac{R(X_i,T_i,\bar{b},b,\tau,\delta)}{||b-\beta\left(\tau\right)||}$$

such that

$$\frac{\left|R(X,T,\bar{b},b,\tau,\delta)\right|}{||b-\beta\left(\tau\right)||}<\frac{|\Delta(X,b)|}{||b-\beta\left(\tau\right)||}1\left\{|\varepsilon_\tau|\leq|\Delta(X,b)|\right\}$$

where

$$\begin{aligned}
& R(X,T,\bar{b},b,\tau,\delta) \\
= & \left[\rho_\tau(T_i-T(X_i,b))-\rho_\tau(T_i-T(X_i,\beta\left(\tau\right)))-\varphi_\tau(T_i-T(X_i,\beta\left(\tau\right)))\Delta(X,b_2)\right] \\
& 1\left\{T(x,\bar{b})<C-\delta\right\} \\
= & \left[\rho_\tau(\varepsilon_{\tau i}-\Delta(X,b))-\rho_\tau(\varepsilon_{\tau_i})-\varphi_\tau(T_i-T(X_i,\beta\left(\tau\right)))\Delta(X,b)\right]1\left\{T(x,\bar{b})<C-\delta\right\}
\end{aligned}$$

with $\Delta(X,b)=\ln T(X,b))-\ln T(X,\beta\left(\tau\right))$ and $\varepsilon_{\tau i}=\ln T_i-\ln T(X_i,\beta\left(\tau\right)$.

From Knight's (1998) identity

$$\begin{aligned}
\rho_\tau(x-v)-\rho_\tau(x) &= v\left(1\left\{x\leq 0\right\}-\tau\right)+\int_0^v\left(1\left\{x\leq t\right\}-1\left\{x\leq 0\right\}\right)dt \\
&= v\left[\left(1\left\{x\leq 0\right\}-\tau\right)+\int_0^1\left(1\left\{x\leq vu\right\}-1\left\{x\leq 0\right\}\right)du\right]
\end{aligned}$$

we can show that

$$\begin{aligned}
& \frac{\rho_\tau(T-T(X,b))-\rho_\tau(T-T(X,\beta))}{T(X,b)-T(X,\beta)} \\
= & \left[(1\{T-T(X,\beta)\leq 0\}-\tau)+\int_0^1 (1\{\varepsilon_\tau \leq u\,(T(X,b)-T(X,\beta))\}-1\{\varepsilon_\tau\leq 0\})\,du\right]
\end{aligned}$$

where $\varepsilon_\tau = T - T(X,\beta)$. Then, similar to the arguments in the proof of Theorem 3 in Chen et al. (2003), we can show that the class of functions

$$F_1 = \{1\{\varepsilon_\tau \leq u\,(T(X,b)-T(X,\beta))\} - 1\{\varepsilon_\tau \leq 0\}\colon\ \tau\,(0,1)\,,\ b\in B,\ u\in[0,1]\}$$

is Euclidean (Pakes and Pollard 1989) with a bounded envelope; hence by Theorem 5.3 of Dudley (1987), the class of functions

$$\bar{\mathcal{F}}_1 = \left\{\int_0^1 (1\{\varepsilon_\tau \leq u\,(T(X,b)-T(X,\beta))\}-1\{\varepsilon_\tau\leq 0\})\,du\colon\ \tau\,(0,1)\,,\ b\in B,\ u\in[0,1]\right\}$$

also has finite entropy intergal with

$$\sup_Q D_2(\varepsilon,\bar{\mathcal{F}}_1,Q)\leq C_1\exp(C_2 e^{-\lambda})$$

for some constant terms C_1, C_2 and $\lambda < 2$ and any probability measure Q. In addition, also following the arguments in Chen et. al (2003), we can show that the classes of functions

$$\mathcal{F}_2 = \{1\{T(X,b) < C-\delta\}\colon\ \|b-\beta\,(\tau)\| < \delta/2, \tau\in(0,1)\}$$

and

$$\mathcal{F}_3 = \{\varphi_\tau(T-T(X,\beta\,(\tau))\colon\ \tau\in(0,1)\}$$

are Euclidean with bounded envelopes. Therefore, we can also establish that

$$\mathcal{F}_4 = \left\{\frac{|R(X,T,b_1,b_2,\tau,\delta)|}{T(X,b)-T(X,\beta)}\colon\ \|b-\beta\,(\tau)\|<\delta/2,\ \tau\in(0,1)\right\}$$

and

$$\begin{aligned}
\bar{\mathcal{F}}_4 &= \left\{\frac{T(X,b)-T(X,\beta)}{\|b-\beta\|}\frac{|R(X,T,b_1,b_2,\tau,\delta)|}{T(X,b)-T(X,\beta)}\colon\ \|b-\beta\,(\tau)\|<\delta/2,\ \tau\in(0,1)\right\} \\
&= \left\{\frac{|R(X,T,b_1,b_2,\tau,\delta)|}{\|b-\beta\|}\colon\ \|b-\beta\,(\tau)\|<\delta/2,\ \tau\in(0,1)\right\}
\end{aligned}$$

have finite entropy intergal with bounded envelope. Consequently, similar to (R1), we can show that

$$S_n(\bar{b},b,\tau,\delta)-S(\bar{b},b,\tau,\delta) = O\left(n^{-1/2}\ln\ln n\right) \tag{R2}$$

and

$$Q_n(\bar{b}, b, \tau, \delta) - Q(\bar{b}, b, \tau, \delta) = O\left(n^{-1/2} \ln \ln n\right) \tag{R3}$$

almost surely uniformly in $\left(\bar{b}, b, \tau, \delta\right)$, where

$$S(\bar{b}, b, \tau, \delta) = ES_n(\bar{b}, b, \tau, \delta)$$

and

$$Q(\bar{b}, b, \tau, \delta) = EQ_n(b, b, \tau, \delta).$$

With the above uniform convergence results, we are now ready to prove the uniform consistency of $\hat{\beta}\left(\tau_j\right)$ and its rate of convergence, for $j = 1, 2, ..., L_n$, through a sequential argument.

Given that $\hat{\beta}_{\tau_0} - \beta_{\tau_0} = O\left(n^{-1/2} \ln n\right)$ almost surely, following Chen (2018), we can show that

$$\hat{b}_{2\tau_1}(\beta_{1\tau_1}) - \beta_{2\tau_1} = O\left(n^{-1/2} \ln n\right)$$

for $\omega \in \Omega_0$ with $\Pr(\Omega_0) = 1$. From (R1) and the definition of $\hat{\beta}_1(\tau_1)$ we have

$$\begin{aligned} 0 &\geq G_n(\hat{\beta}_{\tau_0}, \hat{\beta}_{1\tau_1}, \hat{b}_{2\tau_1}(\hat{\beta}_{1\tau_1}), \tau_1, \delta_n) - G_n(\hat{\beta}_{\tau_0}, \beta_{1\tau_1}, \hat{b}_{2\tau_1}(\beta_{1\tau_1}), \tau_1, \delta_n) \\ &= G_0(\hat{\beta}_{\tau_0}, \hat{\beta}_{1\tau_1}, \hat{b}_{2\tau_1}(\hat{\beta}_{1\tau_1}), \tau_1, \delta_n) - G_0(\hat{\beta}_{\tau_0}, \beta_{1\tau_1}, \hat{b}_{2\tau_1}(\beta_{1\tau_1}), \tau_1, \delta_n) + O\left(n^{-1/2} \ln n\right) \\ &= G_0(\hat{\beta}_{\tau_0}, \hat{\beta}_{1\tau_1}, \hat{b}_{2\tau_1}(\hat{\beta}_{1\tau_1}), \tau_1, \delta_n) - G_0(\hat{\beta}_{\tau_0}, \beta_{1\tau_1}, \beta_{2\tau_1}, \tau_1, \delta_n) + O\left(n^{-1/2} \ln n\right). \end{aligned}$$

We now show that $\hat{\beta}_{\tau_1} - \beta_{\tau_1}$ converges to 0 for $\omega \in \Omega_0$; if this is not the case, then there exists a subsequence and a constant $c_1 \neq 0$, such that $\hat{\beta}_{\tau_1} - \beta_{\tau_1} - c_1 \to 0$. Then, following the arguments in the proof of Theorem1 in Chen (2018), we can show that, for any $\varepsilon_0 > 0$,

$$\varepsilon_0 \geq G(\beta_{\tau_1}, \beta_{\tau_1} + c, \tau_1, \delta_0) - G(\beta_{\tau_1}, \beta_{\tau_1}, \tau_1, \delta_0),$$

which contradicts Assumption 4.

Next we establish the almost sure rate of convergence for $\hat{\beta}_{\tau_1} - \beta_{\tau_1} \to 0$. Again, similar to the arguments in Chen (2018), we can show that for any given $\varepsilon \in \Omega_0$, N_1 can be chosen large enough so that for $n > N_1$,

$$0 \geq c_0 ||\hat{\beta}\left(\tau_1\right) - \beta\left(\tau_1\right)||^2 + O\left(n^{-\frac{1}{2}} \ln \ln n\right)\left(\hat{\beta}\left(\tau_1\right) - \beta\left(\tau_1\right)\right)$$

for some positive constant c_0. Hence

$$\hat{\beta}\left(\tau_1\right) - \beta\left(\tau_1\right) = O\left(n^{-\frac{1}{2}} \ln \ln n\right)$$

almost surely. Therefore, we can actually choose N_1 such that for $n > N_1$,

$$||\hat{\beta}\left(\tau_1\right) - \beta\left(\tau_1\right)|| < M_1 n^{-\frac{1}{2}} \ln \ln n.$$

Following the same logic, we can prove through a sequential argument similar to Chen (2018) that we can choose N large enough so that for $n > N$,

$$||\hat{\beta}(\tau_j) - \beta(\tau_j)|| < Mn^{-\frac{1}{2}} \ln \ln n$$

for $j = 1, 2, ..., L_n$. In other words,

$$\sup_{1 \leq j \leq L_n} ||\hat{\beta}(\tau_j) - \beta(\tau_j)|| = O\left(n^{-1/2} \ln \ln n\right)$$

almost surely.

Step 2. In this step, with similar arguments to those in Chen (2018), we can also show that $\hat{\beta}(\tau_0) - \beta(\tau_0) = O\left(n^{-1/2} \ln \ln n\right)$ almost surely.

Proof of Theorem 2: Similar to the proof of the uniform convergence of $\hat{\beta}(\tau_j)$ for $j = 0, 1, 2, ..., L_n$, we can show that

$$|\hat{\beta}(\tau) - \beta(\tau)| = O\left(n^{-1/2} \ln \ln n\right)$$

almost surely, uniformly in $\tau \in [\tau_0, \tau_u]$. We now establish the weak convergence of the quantile coefficient process.

Given the rate of convergence results in Theorem 1, following Sherman (1993) and Chen (2018),we can establish the following quadratic approximation,

$$\begin{aligned}
& G_n(\bar{b}, b, \tau, \delta_n) - G_n(\bar{b}, \beta_\tau, \tau, \delta_n) \\
= \; & G(\bar{b}, b, \tau, \delta_n) - G(\bar{b}, \beta_\tau, \tau, \delta_n) + \\
& G_n(\bar{b}, b, \tau, \delta_n) - G_n(\bar{b}, \beta_\tau, \tau, \delta_n) - \left[G(\bar{b}, b, \tau, \delta_n) - G(\bar{b}, \beta_\tau, \tau, \delta_n)\right] \\
& -S_n(\bar{b}, \beta_\tau, \tau, \delta_n)'(b - \beta_\tau) + o_p\left(n^{-1/2}(b - \beta_\tau)\right)) \\
= \; & \frac{1}{2}(b - \beta_\tau)' V_\tau (b - \beta_\tau) + o\left(\|b - \beta_\tau\|^2\right) \\
& +S_{n0}(\tau)'(b - \beta_\tau) + o_p\left(n^{-1/2}(b - \beta_\tau)\right))
\end{aligned}$$

uniformly in $b, \bar{b} = \beta_\tau + o(\delta_n)$ and $\tau \in [\tau_0, \tau_u]$, where

$$S_{n0}(\tau) = \frac{1}{n} \sum_{i=1}^{n} \varphi_\tau(T_i - T(X_i, \beta_\tau)) \left[\frac{\partial}{\partial b} \ln T(X_i, \beta_\tau)\right] 1\{T(X_i, \beta_\tau) < C\}$$

and

$$V_\tau = E\left[\frac{\partial^2}{\partial b \partial b'} \ln T(X_i, \beta_\tau) 1\{T(X_i, \beta_\tau) < C\}\right].$$

By Lemma A1, we have

$$\hat{b}_{2\tau}(\hat{\beta}_{1\tau}) - \beta_{2\tau} = \Gamma_{\tau 22}^{-1}\Gamma_{\tau 21}(\hat{\beta}_{1\tau} - \beta_{1\tau}) + \frac{1}{\sqrt{n}}W_{n21\tau} + o_p\left(n^{-1/2}\right)$$

where

$$W_{n21\tau} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{i21\tau}$$

with

$$\phi_{i21\tau} = \Gamma_{\tau 22}^{-1}\left(1\left\{\ln T_i < T(X_i, \beta_\tau)\right\} - \tau\right) X_{2i} 1\left\{T(X_i, \beta_\tau) < C\right\}.$$

Therefore, with some simple manipulation, we obtain

$$\begin{aligned}
& \left(\hat{\beta}_{1\tau} - \beta_{1\tau}\right)' V_\tau \left(\hat{\beta}_\tau - \beta_\tau\right) \\
= & \left(\left(\hat{\beta}_{1\tau} - \beta_{1\tau}\right)', \left(\frac{1}{\sqrt{n}}W_{n21\tau} + \Gamma_{\tau 21}(\hat{\beta}_{1\tau} - \beta_{1\tau})\right)'\right) V_\tau \begin{pmatrix} \hat{\beta}_{1\tau} - \beta_{1\tau} \\ \frac{1}{\sqrt{n}}W_{n21\tau} + \Gamma_{\tau 21}(\hat{\beta}_{1\tau} - \beta_{1\tau}) \end{pmatrix} \\
= & \left(\left(\hat{\beta}_{1\tau} - \beta_{1\tau}\right)', \left(\Gamma_{\tau 21}(\hat{\beta}_{1\tau} - \beta_{1\tau})\right)'\right) V_\tau \begin{pmatrix} \hat{\beta}_{1\tau} - \beta_{1\tau} \\ \Gamma_{\tau 21}(\hat{\beta}_{1\tau} - \beta_{1\tau}) \end{pmatrix} \\
& +2\left(0, \frac{1}{\sqrt{n}}W_{n21\tau}\right)\begin{pmatrix} V_{\tau 11} & V_{\tau 12} \\ V_{\tau 21} & V_{\tau 22} \end{pmatrix}\begin{pmatrix} \hat{\beta}_{1\tau} - \beta_{1\tau} \\ \Gamma_{\tau 21}(\hat{\beta}_{1\tau} - \beta_{1\tau}) \end{pmatrix} \\
& +\frac{1}{n}W_{n21\tau}' V_{\tau 22} W_{n21\tau} \\
= & (\hat{\beta}_{1\tau} - \beta_{1\tau})' V_{\tau 11}^0 (\hat{\beta}_{1\tau} - \beta_{1\tau}) + \frac{2}{\sqrt{n}}W_{n1\tau}'(\hat{\beta}_{1\tau} - \beta_{1\tau}) + \frac{1}{n}W_{n21\tau}' V_{\tau 22} W_{n21\tau}
\end{aligned}$$

where

$$W_{n1\tau} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(V_{\tau 21}' + \Gamma_{\tau 21}' V_{\tau 22})\phi_{21\tau i}$$

and

$$\begin{aligned}
V_{\tau 11}^0 &= (I, \Gamma_{\tau 12}')\begin{pmatrix} V_{\tau 11} & V_{\tau 12} \\ V_{\tau 21} & V_{\tau 22} \end{pmatrix}\begin{pmatrix} I \\ \Gamma_{\tau 12} \end{pmatrix} \\
&= V_{\tau 11} + \Gamma_{\tau 12}' V_{\tau 21} + V_{\tau 12}\Gamma_{\tau 12} + \Gamma_{\tau 12}' V_{\tau 22}\Gamma_{\tau 12}
\end{aligned}$$

In addition,

$$\begin{aligned}
S_{n0}'(\hat{\beta}_\tau - \beta_\tau) &= S_{n0}'\begin{pmatrix} \hat{\beta}_{1\tau} - \beta_{1\tau} \\ \frac{1}{\sqrt{n}}W_{n21\tau} + \Gamma_{\tau 22}^{-1}\Gamma_{\tau 21}(\hat{\beta}_{1\tau} - \beta_{1\tau}) \end{pmatrix} + o_p\left(n^{-1}\right) \\
&= S_{n0}'\begin{pmatrix} I \\ \Gamma_{\tau 22}^{-1}\Gamma_{\tau 21} \end{pmatrix}\left(\hat{\beta}_{1\tau} - \beta_{1\tau}\right) + S_{n0}'\begin{pmatrix} 0 \\ \frac{1}{\sqrt{n}}W_{n21\tau} \end{pmatrix} + o_p\left(n^{-1}\right).
\end{aligned}$$

Therefore,

$$
\begin{aligned}
& G_n(\hat{\beta}_{\tau_j}, \hat{\beta}_\tau, \tau, \delta_n) - G_n(\hat{\beta}_{\tau_j}, \hat{\beta}_\tau, \tau, \delta_n) \\
= \; & \frac{1}{2}(\hat{\beta}_\tau - \beta_\tau)' V_\tau (\hat{\beta}_\tau - \beta_\tau) + o\left(\left\|\hat{\beta}_\tau - \beta_\tau\right\|^2\right) \\
& + S_{n0}(\tau)'(\hat{\beta}_\tau - \beta_\tau) + o_p\left(n^{-1/2}(\hat{\beta}_\tau - \beta_\tau)\right) \\
= \; & \frac{1}{2}(\hat{\beta}_{1\tau} - \beta_{1\tau})' V_{11}^0 (\hat{\beta}_{1\tau} - \beta_{1\tau}) + \frac{1}{\sqrt{n}} W_{n1\tau}^{*\prime}(\hat{\beta}_{1\tau} - \beta_{1\tau}) + \frac{1}{2n} W_{n21\tau}' V_{\tau 22} W_{n21\tau} \\
& + S_{n0}' \begin{pmatrix} 0 \\ \frac{1}{\sqrt{n}} W_{n21\tau} \end{pmatrix} + o_p\left(n^{-1} + \left\|\hat{\beta}_\tau - \beta_\tau\right\|^2\right)
\end{aligned}
$$

where

$$W_{n1\tau}^* = W_{n1\tau} + \begin{pmatrix} I & \Gamma_{\tau 21}' \Gamma_{\tau 22}^{-1} \end{pmatrix} S_{n0}$$

Then using the arguments in Sherman (1993), we obtain

$$\sqrt{n}(\hat{\beta}_{1\tau} - \beta_{1\tau}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{1\tau i} + o_p(1)$$

where

$$\phi_{1\tau i} = \left(V_{\tau 11}^0\right)^{-1} \left[(V_{\tau 21}' + \Gamma_{\tau 12}' V_{\tau 22})\phi_{21\tau i} + \phi_{20\tau_i}\right]$$

with

$$\phi_{20\tau_i} = \varphi_\tau(T_i - T(X_i, \beta_\tau)) \left[\frac{\partial}{\partial b} \ln T(X_i, \beta_\tau)\right] 1\left\{T(X_i, \beta_\tau) < C\right\}$$

which also implies that

$$
\begin{aligned}
\hat{b}_{2\tau}(\hat{\beta}_{1\tau}) - \beta_{2\tau} &= \Gamma_{\tau 22}^{-1}\Gamma_{\tau 21}(\hat{\beta}_{1\tau} - \beta_{1\tau}) + \Gamma_{\tau 22}^{-1} S_{n2}(\beta_\tau, \beta_\tau, \tau, \delta_n) + o_p\left(n^{-1/2}\right) \\
&= \Gamma_{\tau 22}^{-1}\Gamma_{\tau 21}(\hat{\beta}_{1\tau} - \beta_{1\tau}) + \frac{1}{n}\sum_{i=1}^{n} \phi_{21\tau i} + o_p\left(n^{-1/2}\right). \\
&= \frac{1}{n}\sum_{i=1}^{n} \phi_{2\tau i} + o_p\left(n^{-1/2}\right)
\end{aligned}
$$

where

$$\phi_{2\tau i} = \phi_{21\tau i} + \Gamma_{\tau 22}^{-1}\Gamma_{\tau 21}\phi_{1\tau i}.$$

As a result, we have

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{\tau i} + o_p(1)$$

uniformly in $\tau \in [\tau_0, \tau_u]$. It is straightforward to show that the class of functions $\{\phi(T_i, X_i, \tau)\colon\ \tau \in [\tau_0, \tau_u]\}$ is Donsker, thus Theorem 2 follows from the standard functional central limit theorem.

Proof of Theorem 3: Following the proof of Theorem 2, we can show that uniform consistency of $\hat{\beta}^* (\tau)$ jointly in space $P = P \times P_\xi$,

$$\sup_{\tau \in [\tau_l, \tau_u]} ||\hat{\beta}^* (\tau) - \beta (\tau) || = O \left(n^{-1/2} \ln \ln n \right)$$

almost surely in P. Furthermore, we can also establish the asymptotic linear representations

$$\sqrt{n} \left(\hat{\beta} (\tau) - \beta (\tau) \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{i\tau} + o_{\mathbb{P}} (1)$$

and

$$\sqrt{n} \left(\hat{\beta}^* (\tau) - \beta (\tau) \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \phi_{i\tau} + o_{\mathbb{P}} (1)$$

and thus

$$\sqrt{n} \left(\hat{\beta}^* (\tau) - \hat{\beta} (\tau) \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\xi_i - 1) \phi_{i\tau} + o_{\mathbb{P}} (1)$$

uniformly in $\tau \in [\tau_0, \tau_u]$. Finally, from the Conditional Central Limit Theorem (Th. 2.9.6, van der Vaart and Wellner, 1996), $\sqrt{n} \left(\hat{\beta}^*(\tau) - \hat{\beta}(\tau) \right)$ converges to a mean zero Gaussian process $G(\cdot)$ for $\tau \in [\tau_0, \tau_u]$ with covariance function of the form

$$EG(\tau)G(\tau')' = E \left[\phi_{i\tau} \phi'_{i\tau} \right].$$

References

Arcones, M.A. (1993): The Law of the Iterated Logarithm for U-Processes, Journal of Multivariate Analysis, 47, 139-151.

Ebenstein A., M. Fan, M. Greenstone, G. He, and M. Zhou (2017): New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy, PNAS, 114 (39) 10384-10389.

Bijwaard, G. E. and G. Ridder (2005). Correcting for selective compliance in a re-employment bonus experiment. Journal of Econometrics 125, 77–111.

Buchinsky, M. (1994): Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression, Econometrica, 62, 405-458.

Buchinsky, M., and J. Hahn (1998): An Alternative Estimator for Censored Quantile Regression, Econometrica, 66, 653-671.

Blundell, R, and J.L. Powell (2007): Censored Regression Quantiles with Endogenous Regressors, Journal of Econometrics, 141, 65-83.

Chen, S. (2018): Sequential estimation of censored quantile regression models, Journal of Econometrics, 207, 30-52.

Chen, S. (2019): Quantile regression for duration models with time-varying regressors, Journal of Econometrics, 209,1-17.

Chen, X, O. Linton andI. van Keilegom (2003): Estimation of Semiparametric Models when the Criterion Function Is Not Smooth, Econometrica, 71, 1591–1608.

Chernozhukov, V., and H. Christian (2006): Instrumental Quantile Regression Inference for Structural and Treatment Effect Models, Journal of Econometrics,132, 491-525.

Chernozhukov, V., and H. Christian (2008): Instrumental variable quantile regression: A robust inference approach, Journal of Econometrics, 142, 379-398

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

Chernozhukov, V., and H. Hong (2002): Three-Step Quantile Regression and Extramarital Affairs, Journal of The American Statistical Association, 97, 872-882.

Cox, D. R. and D. Oakes (1984). Analysis of Survival Data. London: Chapman and Hall

Dudley, R.M. (1987): Universal Donsker Classes and Metric Entropy, 15, Annals of Probability, 1306-1326.

Fitzenberger B (1996): A Guide to Censored Quantile Regressions, In C Rao, G Maddala(eds.), Handbook of Statistics, North-Holland: New York.

Fitzenberger B and P. Winker (2007): Improving the Computation of Censored Quantile Regressions, Computational Statistics and Data Anlaysis, 52, 88–108.

Honoré, B.E. and L. Hu (2004): Estimation of cross sectional and panel data censored regression models with endogeneity, Journal of Econometrics, 122, 293 – 316.

Khan, S. and J.L. Powell (2001): Two-step estimation of semiparametric censored regression models, Journal of Econometrics, 103,73-110.

Khan, S., and E. Tamer (2009): Inference on endogenously censored regression models using conditional moment inequalities, Journal of Econometrics, 152,104–119.

Knight, K. (1998). Limiting distributions for l1 regression estimators under general conditions. Ann. Statist. 26 755–770

Koenker, R. and G. Bassett (1978): "Regression Quantiles," Econometrica, 46, 33-50.

Koenker, R. and Y. Bilias (2001). Quantile regression for duration data: A reappraisal of the Pennsylvania reemployment bonus experiments. Empirical Economics, 26, 199–220.

Koenker, R. and O. Geling (2001). Reappraising medfly longevity: A quantile regression survival analysis. Journal of the American Statistical Association, 96, 458–468.

Meyer, B. D. (1996). What have we learned from the Illinois reemployment bonus experiment? Journal of Labor Economics 14, 26–51.

Pakes A. and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," Econometrica, 57, 1027-1057.

Powell, J.L. (1984): Least absolute deviations estimation for the censored regression model, Journal of Econometrics, 25, 303-325.

Powell, J.L. (1986): Censored regression quantiles, Journal of Econometrics, 32, 143-155

Powell, J.L. (1991): Estimation of monotonic regression models under quantile restrictions, Nonparametric and Semiparametric Methods in Econometrics and Statistics, Barnett W.A, J.L. Powell and G. Tauchen Ed. Cambridge University Press.

Sherman, R.P. (1993) The limiting distribution of the maximum rank correlation estimator. Econometrica 61, 123-13

van der Vaart, A.W. and J. Wellner (2006): Weak convergence and empirical processes. Springer.

Woodbury, S. A. and R. G. Spiegelman (1987). Bonuses to workers and employers to reduce unemployment: Randomized trials in Illinois. American Economic Review 77, 513–530.

Table 1: Homogeneous Design

		$n=200$				$n=800$			
τ	$\beta(\tau)$	Bias	SD	est. SD	CP95	Bias	SD	est. SD	CP95
15% censoring									
0.30	0.30	-0.0125	0.1346	0.1379	0.940	-0.0023	0.0678	0.0703	0.946
	1.00	0.0059	0.1190	0.1201	0.950	0.0026	0.0611	0.0612	0.958
	0.00	0.0032	0.0323	0.0338	0.954	0.0010	0.0163	0.0172	0.948
	0.50	0.0044	0.0562	0.0576	0.950	0.0011	0.0280	0.0293	0.944
0.50	0.50	-0.0103	0.1581	0.1600	0.946	-0.0045	0.0841	0.0840	0.954
	1.00	-0.0012	0.1386	0.1400	0.944	0.0021	0.0681	0.0714	0.952
	0.00	0.0010	0.0369	0.0372	0.936	-0.0002	0.0179	0.0192	0.952
	0.50	0.0041	0.0662	0.0655	0.946	0.0031	0.0337	0.0344	0.946
0.70	0.70	-0.0172	0.1651	0.1642	0.942	0.0003	0.0810	0.0827	0.948
	1.00	0.0044	0.1402	0.1435	0.942	0.0000	0.0703	0.0726	0.954
	0.00	-0.0020	0.0366	0.0366	0.940	-0.0006	0.0173	0.0185	0.956
	0.50	0.0055	0.0689	0.0667	0.934	0.0003	0.0325	0.0341	0.952
30% censoring									
0.30	0.30	-0.0084	0.1529	0.1539	0.936	-0.0036	0.0788	0.0802	0.946
	1.00	0.0024	0.1271	0.1297	0.944	0.0039	0.0691	0.0679	0.942
	0.00	0.0031	0.0335	0.0345	0.946	0.0011	0.0176	0.0180	0.952
	0.50	0.0031	0.0655	0.0639	0.920	0.0017	0.0327	0.0335	0.950
0.50	0.50	-0.0091	0.1840	0.1878	0.934	-0.0037	0.1014	0.0986	0.938
	1.00	0.0008	0.1631	0.1616	0.930	0.0000	0.0848	0.0832	0.928
	0.00	-0.0003	0.0390	0.0397	0.950	0.0001	0.0197	0.0206	0.942
	0.50	0.0036	0.0771	0.0773	0.942	0.0028	0.0411	0.0412	0.948
0.70	0.70	-0.0106	0.2148	0.2065	0.936	0.0031	0.1017	0.1038	0.944
	1.00	0.0044	0.1792	0.1864	0.950	-0.0038	0.0954	0.0927	0.936
	0.00	-0.0018	0.0413	0.0414	0.934	-0.0014	0.0203	0.0210	0.954
	0.50	-0.0001	0.0965	0.0884	0.932	-0.0006	0.0424	0.0453	0.956

Table 2: Heterogeneous Design

		$n = 200$				$n = 800$			
τ	$\beta(\tau)$	Bias	SD	est.SD	CP95	Bias	SD	est.SD	CP95
15% censoring									
0.30	0.30	-0.0253	0.3532	0.3768	0.956	-0.0034	0.1778	0.1863	0.944
	1.00	0.0185	0.3296	0.3363	0.942	0.0034	0.1679	0.1677	0.952
	0.00	0.0096	0.0985	0.0984	0.944	0.0031	0.0474	0.0489	0.946
	0.30	0.0070	0.1662	0.1742	0.948	0.0022	0.0818	0.0857	0.960
0.50	0.50	-0.0222	0.3798	0.4001	0.952	-0.0070	0.2006	0.2092	0.942
	1.00	-0.0077	0.3656	0.3724	0.948	-0.0012	0.1730	0.1866	0.962
	0.00	0.0030	0.1037	0.1061	0.952	0.0012	0.0512	0.0536	0.950
	0.50	0.0101	0.1788	0.1827	0.942	0.0062	0.0893	0.0947	0.960
0.70	0.70	0.0119	0.3821	0.3732	0.928	0.0156	0.1889	0.1896	0.952
	1.00	-0.0102	0.3518	0.3488	0.940	-0.0054	0.1754	0.1768	0.942
	0.00	-0.0024	0.1023	0.0995	0.948	-0.0030	0.0465	0.0493	0.958
	0.70	-0.0171	0.1765	0.1625	0.918	-0.0067	0.0818	0.0840	0.944
30% censoring									
0.30	0.30	-0.0316	0.3543	0.3709	0.944	0.0075	0.1191	0.1876	0.948
	1.00	0.0179	0.3271	0.3340	0.940	0.0034	0.1672	0.1684	0.958
	0.00	0.0110	0.0976	0.0967	0.940	0.0032	0.0470	0.0489	0.946
	0.30	0.0096	0.1676	0.1711	0.934	0.0023	0.0831	0.0863	0.962
0.50	0.50	0.0342	0.3859	0.3758	0.934	0.0119	0.2153	0.2047	0.938
	1.00	-0.0365	0.3566	0.3448	0.948	-0.0075	0.1767	0.1823	0.940
	0.00	0.0017	0.1030	0.0966	0.948	-0.0011	0.0510	0.0503	0.934
	0.50	-0.0190	0.1762	0.1631	0.940	-0.0030	0.0946	0.0907	0.930
0.70	0.70	0.1787	0.5120	0.4821	0.878	0.1687	0.2449	0.2680	0.868
	1.00	-0.0736	0.4013	0.4057	0.894	-0.0795	0.2175	0.2069	0.894
	0.00	-0.0034	0.1179	0.0993	0.894	-0.0046	0.0512	0.0493	0.934
	0.70	-0.1273	0.2429	0.2118	0.770	-0.1001	0.1129	0.1240	0.820

Table 1: Homogeneous Design

		$n = 200$				$n = 800$			
τ	$\beta(\tau)$	Bias	SD	est.SD	CP95	Bias	SD	est.SD	CP95
15% censoring									
0.3	0.3	0.0003	0.1526	0.1722	0.964	0.0016	0.0778	0.0808	0.946
	1.0	0.0033	0.1310	0.1531	0.968	0.0003	0.0636	0.0710	0.960
	1.0	0.0067	0.1283	0.1530	0.974	0.0044	0.0656	0.0710	0.940
	0.5	0.0048	0.1828	0.1898	0.958	-0.0021	0.0903	0.0933	0.972
0.5	0.5	-0.0098	0.2023	0.2171	0.958	0.0043	0.1022	0.1096	0.960
	1.0	0.0121	0.1703	0.1900	0.967	0.0015	0.0807	0.0904	0.944
	1.0	0.0178	0.1701	0.1903	0.968	0.0025	0.0874	0.0904	0.950
	0.5	0.0059	0.2222	0.2276	0.961	-0.0084	0.1122	0.1198	0.974
0.7	0.7	0.0103	0.2865	0.2760	0.952	-0.0056	0.1297	0.1432	0.968
	1.0	0.0159	0.2317	0.2459	0.955	0.0035	0.1005	0.1120	0.966
	1.0	0.0166	0.2313	0.2458	0.954	0.0064	0.1019	0.1134	0.972
	0.5	-0.0198	0.2956	0.2649	0.945	0.0033	0.1342	0.1468	0.970
30% censoring									
0.3	0.3	0.0059	0.1629	0.1725	0.963	0.0017	0.0761	0.0836	0.952
	1.0	0.0023	0.1518	0.1696	0.965	0.0008	0.0730	0.0817	0.972
	1.0	0.0026	0.1518	0.1713	0.965	0.0033	0.0737	0.0818	0.978
	0.5	0.0002	0.1957	0.1953	0.955	-0.0025	0.0922	0.0981	0.950
0.5	0.5	-0.0054	0.2124	0.2219	0.959	0.0033	0.1130	0.1138	0.948
	1.0	0.0171	0.2192	0.2299	0.963	0.0031	0.1076	0.1110	0.962
	1.0	0.0149	0.2124	0.2290	0.963	0.0009	0.1083	0.1109	0.956
	0.5	-0.0023	0.2368	0.2377	0.954	-0.0052	0.1242	0.1287	0.938
0.7	0.7	0.0084	0.3065	0.2971	0.9696	-0.0064	0.1416	0.1583	0.964
	1.0	0.0275	0.3934	0.3894	0.9372	0.0024	0.1487	0.1515	0.940
	1.0	0.0181	0.4008	0.3746	0.9372	0.0028	0.1460	0.1543	0.946
	0.5	-0.0211	0.3279	0.2812	0.9190	0.0052	0.1530	0.1714	0.972

Table 2: Heterogeneous Design

		$n = 200$				$n = 800$			
τ	$\beta(\tau)$	Bias	SD	est.SD	CP95	Bias	SD	est.SD	CP95
15% censoring									
0.3	0.3	0.0002	0.1590	0.1742	0.970	0.0001	0.0793	0.0832	0.946
	1.0	-0.0004	0.1635	0.1827	0.964	0.0023	0.0822	0.0865	0.958
	1.0	0.0083	0.1588	0.1839	0.969	0.0048	0.0790	0.0865	0.962
	-0.2	0.0098	0.2323	0.2254	0.932	-0.0009	0.1140	0.1199	0.942
0.5	0.5	-0.0112	0.2095	0.2167	0.952	0.0025	0.1076	0.1096	0.948
	1.0	0.0113	0.2093	0.2246	0.961	0.0023	0.0984	0.1037	0.956
	1.0	0.0173	0.2041	0.2222	0.969	0.0030	0.1031	0.1036	0.938
	0.0	0.0106	0.2699	0.2589	0.943	-0.0073	0.1388	0.1437	0.956
0.7	0.7	0.0124	0.2796	0.2920	0.970	-0.0045	0.1358	0.1470	0.974
	1.0	0.0097	0.3177	0.3230	0.945	-0.0007	0.1230	0.1355	0.966
	1.0	0.0122	0.3248	0.3250	0.947	0.0007	0.1237	0.1349	0.958
	0.2	-0.0213	0.3354	0.2890	0.934	0.0092	0.1644	0.1752	0.974
30% censoring									
0.3	0.3	0.0144	0.1847	0.1810	0.961	0.0014	0.0839	0.0887	0.944
	1.0	-0.0057	0.1897	0.1953	0.959	-0.0009	0.0854	0.0920	0.966
	1.0	-0.0048	0.1874	0.1956	0.959	0.0053	0.0871	0.0922	0.958
	-0.2	-0.0080	0.2370	0.2224	0.940	-0.0015	0.1195	0.1217	0.942
0.5	0.5	0.0355	0.2225	0.2352	0.957	0.0125	0.1116	0.1127	0.960
	1.0	-0.0172	0.2564	0.2741	0.946	-0.0099	0.1134	0.1178	0.952
	1.0	-0.0105	0.2620	0.2753	0.948	-0.0066	0.1204	0.1175	0.944
	0.0	-0.0439	0.2818	0.2601	0.929	-0.0141	0.1429	0.1444	0.954
0.7	0.7	0.0561	0.3727	0.3438	0.950	0.0124	0.1540	0.1620	0.956
	1.0	0.0203	0.7093	0.5846	0.898	-0.0209	0.1814	0.1931	0.946
	1.0	0.0117	0.7186	0.5630	0.895	-0.0209	0.1810	0.1940	0.942
	0.2	-0.0725	0.3758	0.2947	0.909	-0.0044	0.1846	0.1929	0.970