Integrated Quantile Threshold Regression and Distributional Threshold Effects*

Ping Yu
University of Auckland

Started: July 2009
First Version: January 2013
This Version: July 2013

Abstract

This paper studies the estimation and specification testing in quantile threshold regression. First, we put forward a new estimator of the threshold point, the integrated quantile threshold regression estimator, derive its asymptotic distribution in both the fixed-threshold-effect framework and the shrinking-threshold-effect framework. This new estimator integrates much of the quantile difference information between the two regimes, so is more efficient than the existing estimators such as the least squares estimator and the least absolute deviation estimator. It is actually comparable to the maximum likelihood estimator, so can serve as a better starting point in the adaptive estimation of the threshold point. Inference methods on the threshold point in both frameworks are also discussed. Second, based on the usual conditional quantile threshold process, we define and estimate the marginal distributional threshold process and the marginal quantile threshold process, and provide both the asymptotic and resampling inference methods for these processes. Third, we propose a new score-type test in testing the existence of any quantile threshold effect. This type of test is more powerful than the conventional tests based solely on the least squares estimator or the least absolute deviation estimator. Comparing with the usual Wald-type test, it is computationally less intensive, and its critical values are easier to obtain by the simulation method. Simulation studies confirm the theoretical analysis, and the new estimation and testing procedures are applied to an economic growth model.

KEYWORDS: integrated quantile threshold regression estimator, efficiency, conditional quantile threshold process, marginal distributional threshold process, marginal quantile threshold process, Durbin problem, uniformity, score test, power, resampling method, simulation method
JEL-CLASSIFICATION: C12, C13, C14, C21.

*Email: p.yu@auckland.ac.nz.

1 Introduction

Since the pioneering work by Tong (1978, 1983), threshold models get much popularity in current applied statistical and econometric practice. An encyclopedic survey is available in Tong (1990) and a selective review of the history is given by Tong (2011); see also Hansen (2011) for a summary of applications especially in economics. The usual setup of the threshold regression model is

$$y = \begin{cases} (1, x', q)\,\beta_1 + e_1 \equiv \mathbf{x}'\beta_1 + e_1 & q \leq \gamma; \\ (1, x', q)\,\beta_2 + e_2 \equiv \mathbf{x}'\beta_1 + e_2 & q > \gamma; \end{cases} \tag{1}$$

where q is the threshold variable used to split the sample, $x \in \mathbb{R}^{d-2}$, and $\beta = (\beta_1', \beta_2')' \in \mathbb{R}^{2d}$ are identified by some conditional moment restrictions, e.g., $E\left[e_\ell | x, q\right] = 0$, $\ell = 1, 2$, identifies β as the conditional mean parameters and correspondingly, γ is identified by the least squares estimator (LSE). In practice, quantile threshold effects are also of interest; see, e.g., Section 2 of Oka and Qu (2011) for some empirical examples. Quantile threshold regression explores a *set* of moment conditions: $Q_\tau[e_\ell | x, q] = 0$, $\tau \in \mathcal{T} = [\underline{\tau}, \overline{\tau}] \subset (0, 1)$ with $\underline{\tau} > 0$ and $\overline{\tau} < 1$, where $Q_\tau\left[\cdot\right]$ is the τth quantile of the argument. Under such moment restrictions, it is more convenient to rewrite (1) as

$$y = \begin{cases} \mathbf{x}'\beta_1\left(\tau\right) + e_{1\tau}, & q \leq \gamma; \\ \mathbf{x}'\beta_2\left(\tau\right) + e_{2\tau}, & q > \gamma; \end{cases} \tag{2}$$

to allow both β_ℓ and e_ℓ to depend on the quantile index τ. The parameters of interest are $\left\{\gamma, \beta\left(\tau\right)\right\}_{\tau \in \mathcal{T}}$ with $\beta\left(\tau\right) = \left(\beta_1\left(\tau\right)', \beta_2\left(\tau\right)'\right)'$.

A key observation in model (2) is that although $\beta\left(\tau\right)$ depends on τ, the threshold parameter γ is invariant to τ. In other words, there is a shift in the conditional distribution of y at $q = \gamma$, so all quantile differences between these two regimes can be integrated to identify γ or test whether there are threshold effects. This of course will improve the efficiency of the usual quantile threshold estimator and the power of the usual specification testing based on only one quantile difference. Also, the set of β differences, $\left\{\beta_2\left(\tau\right) - \beta_1\left(\tau\right)\right\}_{\tau \in \mathcal{T}}$, will provide a more complete picture about the threshold effects in the conditional distribution of y than a single β difference resulting from the LSE.

The contributions of this paper are threefold. First, we put forward a new estimator of γ, called the integrated quantile threshold regression estimator (IQTRE). We derive its asymptotic distribution in both the fixed-threshold-effect framework of Chan (1993) and the shrinking-threshold-effect framework of Hansen (2000). We also discuss the inference methods in both frameworks. To motivate our new estimator, consider the following simple example. Suppose $y = \sigma_1 e 1(q \leq \gamma) + \sigma_2 e 1(q > \gamma)$, where $1(\cdot)$ is the indicator function, $\sigma_1 \neq \sigma_2$, e is independent of q, and $E[e] = 0$. In this simple model, there is no threshold effect in the conditional mean of y given that $E[y|q] = 0$ for any q, so the least squares estimator (LSE) cannot identify γ. If we further assume that e is symmetric about zero, then the least absolute deviation estimator (LADE) cannot identify γ either since $Q_{0.5}[y|q] = 0$ for any q. In a more general model like $y = e_1 1(q \leq \gamma) + e_2 1(q > \gamma)$ with $Q_\tau[e_1 | q = \gamma-] = Q_\tau[e_2 | q = \gamma+]$, even the τth quantile regression estimator (QRE) cannot identify γ. In other words, any single characteristic of the conditional distribution of y, such as the conditional mean or conditional quantile, cannot guarantee the identification of γ without *ex ante* knowledge on the conditional distribution of y. However, the only information available in practice is usually that there are some differences between the two regimes in the conditional distribution of y. Based on such little information, we can integrate enough many quantile differences to identify γ, which does not require any *ex ante* knowledge on the existence of the threshold effect at a specific τ. Even if γ can be identified by a

single characteristic of the conditional distribution of y, we expect the IQTRE to be more efficient since it uses more information to identify γ. Note that the collected conditional quantiles are the inverse function of the conditional distribution, and they contain the same information, so integrating all quantile differences to identify γ is equivalent to identify γ based on the conditional distribution of y. In other words, the IQTRE would have similar efficiency as the maximum likelihood estimator (MLE) in Yu (2012). This result is quite surprising since we do not specify the conditional distribution of y parametrically while obtain similar efficiency as a parametric estimator. In summary, the IQTRE is expected to have more identification power and be more efficient than the existing estimators. Furthermore, we treat the IQTRE as an intermediate estimator, and use it as the starting point of the adaptive estimator of γ, the semiparametric empirical Bayes estimator (SEBE) proposed in Yu (2008). Given that the IQTRE is more efficient than the LSE and the LADE, we expect the new SEBE performs better in finite samples. As to the inference of γ, we suggest the nonparametric posterior interval (NPI) of Yu (2008) in Chan's framework. In Hansen's framework, we suggest the likelihood-ratio-based confidence interval (LR-CI) of Hansen (2000), with the only difference being that the likelihood ratio is based on the objective function of the IQTRE rather than the LSE.

Second, based on the conditional quantile (CQ) threshold process, we define the marginal distributional (MD) threshold process and the associated marginal quantile (MQ) threshold process, provide estimators for them and derive the corresponding weak limits. We also discuss both the asymptotic and resampling inference methods for these stochastic processes. Third, we propose a new score-type test in testing whether there is a threshold effect at some $\tau \in \mathcal{T}$. This type of test is more powerful than the conventional tests based solely on the LSE or the LADE. Comparing to the usual Wald-type test, this type of test is computationally less intensive, and its critical values are easier to obtain by the simulation method of Hansen (1996).

There exists some literature on the estimation of γ and conditional quantile threshold effects. In the framework of Hansen (2000), Caner (2002) derives the asymptotic distribution of the LADE and argues that the LADE of γ is more efficient than the LSE when the error term has a heavy tail just as in the regular parameter case. Cai and Stander (2008) consider the quantile self-exciting threshold autoregressive (Q-SETAR) model within the Bayesian framework, but they are only interested in regular parameters $\{\beta(\tau)\}_{\tau\in\mathcal{T}}$ and do not derive the asymptotic properties of their estimators. Cai (2010) considers the forecasting problem in the framework of Cai and Stander (2008). Galvao et al. (2011) discuss also the Q-SETAR model, but their theoretical analyses focus mainly on the estimation and inference of $\{\beta(\tau)\}_{\tau\in\mathcal{T}}$; for γ, only consistency is proved. In the related structural change literature, Bai (1995) discusses the asymptotic distribution of the LADE in both frameworks with one break point, and Bai (1998) extends to the LADE with multiple (possibly infinite many) break points in the framework of Hansen (2000). Chen (2008) extends Bai's work to a single QRE, and Oka and Qu (2011) extend further to the estimation based on multiple quantile changes in Hansen's framework and for repeated-sections. As to the inference of γ, Caner (2002) suggests the LR-CI based on the objective function of the LADE. In the structural change literature, all CIs of the break dates are Wald-type; this type of CIs invert the t-statistic and are tractable in Hansen's framework. There is also some literature concentrating on the specification testing of threshold models. Su and Xiao (2008) consider the sup-Wald test and Qu (2008) considers also the subgradient-based test for structural changes in regression quantiles. Kato (2009) extends the scope of convexity arguments to the case where estimators are obtained as stochastic processes and applies this technique to test the existence of median threshold effects using the sup-Wald statistic.

The rest of this paper is organized as follows. Section 2 uses two examples to justify our specification of (2) and define all new estimators in this paper. Section 3 derives the asymptotic distribution of the IQTRE and discuss the inference of γ in both frameworks. Section 4 contains the weak limits for the CQ, MD and MQ threshold processes, and also both the asymptotic and resampling inference methods for these three stochastic

processes. Section 5 constructs the new score-type test and simulates its critical values. Section 6 and 7 include some Monte Carlo simulation results and an application in the economic growth model, respectively, and Section 8 concludes and proposes some future research plans. All proofs and lemmas are given in Appendix A and B, respectively. A word on notation: ℓ is always used for indicating the two regimes in (2), so is not written out explicitly as "$\ell = 1, 2$" throughout the paper. $\beta_\ell(\tau) = (\beta_{\ell 1}(\tau), \beta'_{\ell x}(\tau), \beta_{\ell q}(\tau))' \equiv \left(\beta_{\ell 1}(\tau), \underline{\beta}'_\ell(\tau)\right)'$ and $\delta(\tau) = \beta_1(\tau) - \beta_2(\tau)$. Parameters with superscript 0 (e.g., $\beta^0_\ell(\tau)$) or subscript 0 (e.g., $\sigma_{\ell 0}$, γ_0, δ_0) denote their true values. $\rightsquigarrow$ signifies the weak convergence over a compact metric space. $\ell^\infty(\mathcal{F})$ is the space of real-valued bounded functions defined on the index set equipped with the supremum norm $\|\cdot\|_{\ell^\infty(\mathcal{F})}$. $\|x\|_1 = \sum_i |x_i|$ is the ℓ_1 norm of a vector x, $\|x\|_2 = \sqrt{\sum_i x_i^2}$ is the Euclidean norm, and $\|x\|_\infty = \max_i |x_i|$ is the sup-norm. $\|\cdot\|$ without subscript means the Euclidean norm.

2 The Setup and Estimators

We first consider two examples where the moment conditions $Q_\tau[e_{\ell\tau}|x, q] = 0$, $\tau \in \mathcal{T}$, are satisfied. Suppose first there is not heteroskedasticity in each regime, that is, $e_\ell = \sigma_\ell e$ with e being independent of $(x', q)'$. Such a specification is considered in the original quantile regression literature such as Koenker and Bassett (1978). For model identifiability, we assume that e has median zero and $E[e^2] = 1$. In this case, $\underline{\beta}_\ell(\tau) = \underline{\beta}_\ell$ invariant of τ, $\beta_{\ell 1}(\tau) = \beta_{\ell 1} + \sigma_\ell \xi_\tau$, and $e_{\ell\tau} = \sigma_\ell(e - \xi_\tau)$, where $\beta_{\ell 1}$ is the intercept in the LAD estimation, and ξ_τ is the τth quantile of e. If there is heteroskedasticity in each regime, assume $e_\ell = (\mathbf{x}'\lambda_\ell)\, e$ with e similarly specified as in the first example. Such a specification is considered in, e.g., Koenker and Bassett (1982), Gutenbrunner and Jurečková (1992), and Koenker and Xiao (2002). In this case, $\beta_\ell(\tau) = \beta_\ell + \lambda_\ell \xi_\tau$ and $e_{\ell\tau} = \mathbf{x}'_\ell \lambda_\ell (e - \xi_\tau)$ if $\mathbf{x}'\lambda_\ell \geq 0$ for all $\mathbf{x}$ in regime ℓ, $\beta_\ell(\tau) = \beta_\ell + \lambda_\ell \xi_{1-\tau}$ and $e_{\ell\tau} = \mathbf{x}'\lambda_\ell(e - \xi_{1-\tau})$ if $\mathbf{x}'\lambda_\ell < 0$ for all $\mathbf{x}$ in regime ℓ, where β_ℓ is the parameter in the median regression. Note that it is possible for $\mathbf{x}'\lambda_\ell$ to have the same sign in regime ℓ no matter $(x', q)'$ is bounded or not.[1] Note also that if e has a continuous distribution in both examples, then $\delta(\tau)$ is a continuous function of τ. Although the quantile threshold specification (2) seems restrictive, it is actually far from it. For example, a subset of $\beta_\ell(\tau)$ may be restricted to be constant over the two regimes to allow for partial structural changes; $\mathbf{x}$ may include functions (e.g., polynomials or B-splines) of the original covariates so that the conditional quantile of y may be a nonlinear function of the original $(x', q)'$.

We next define the IQTRE of γ. For this purpose, we first define the QRE of γ. Suppose a random sample $\{w_i\}_{i=1}^n$ is observed, where $w_i = (y_i, x'_i, q_i)'$, the QRE of γ is usually defined by a profiled procedure.

$$\widehat{\gamma}_\tau = \arg\min_\gamma Q_{\tau n}(\gamma),$$

where

$$Q_{\tau n}(\gamma) = \min_{\beta_1, \beta_2} \frac{1}{n} \sum_{i=1}^n \rho_\tau\left(y_i - \mathbf{x}'_i \beta_1 1(q_i \leq \gamma) - \mathbf{x}'_i \beta_2 1(q_i > \gamma)\right),$$

and

$$\rho_\tau(z) = z(\tau - 1(z \leq 0))$$

is the check function of quantile regression. Usually, there is an interval of γ minimizing this objective function. Most literature in threshold regression takes the left endpoint of the interval as the minimizer

[1] If $(x', q)'$ is bounded, $\mathbf{x}'\lambda_\ell$ is bounded, so it is possible for $\mathbf{x}'\lambda_\ell$ to maintain the same sign for all $\mathbf{x}$ in regime ℓ as long as λ_ℓ is suitably defined. Even if $(x', q)'$ is unbounded, this assumption can still hold. For example, suppose $\mathbf{x} = x \in \mathbb{R}_+ \equiv \{z |\ z \in \mathbb{R} \text{ and } z \geq 0\}$; then $\mathbf{x}'\lambda_\ell$ will have the same sign for all $\mathbf{x}$ no matter λ_ℓ is positive or negative. It is not hard to check that this assumption may still hold when $\mathbf{x}$ includes other covariates.

and calls the estimator as the left-endpoint QRE (LQRE). Yu (2008a, 2012) shows in the least squares estimation that the middle point of the interval is more efficient than the left endpoint in most cases, so we will concentrate on the middle-point QRE (MQRE) in the following discussion. We now define the IQTRE of γ as

$$\widehat{\gamma} = \arg\min_{\gamma} Q_{Tn}(\gamma), \text{where } Q_{Tn}(\gamma) \equiv \sum_{t=1}^{T} Q_{\tau_t n}(\gamma), \text{ and } \tau_t \in \mathcal{T},\ t = 1 \cdots, T. \tag{3}$$

Due to a technical reason, T in the definition of IQTRE is finite. In practice, T can be chosen freely to capture the threshold effects at all possible quantiles. Also, if we have prior information on the magnitude of $\delta(\tau)$, $\tau \in \mathcal{T}$, then we can put more weights on the τ's with $\delta(\tau)$ being large, and the corresponding objective function changes to $\sum_{t=1}^{T} \omega(\tau_t) Q_{\tau_t n}(\gamma)$ with $\sum_{t=1}^{T} \omega(\tau_t) = 1$. The objective function in (3) corresponds to $\omega(\tau_t) = 1/T$ for all $t \in \{1 \cdots, T\}$.

We are now ready to estimate the MD threshold effects. For this purpose, we need first estimate the QR processes $\beta_\ell(\cdot)$. Let

$$\widehat{\beta}_1(\tau) = \arg\min_{\beta_1} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i'\beta_1)\, 1(q_i \le \widehat{\gamma}),$$
$$\widehat{\beta}_2(\tau) = \arg\min_{\beta_2} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i'\beta_2)\, 1(q_i > \widehat{\gamma});$$

then we can estimate the conditional distribution of y given $(x', q)'$ in each regime, $F_\ell(y|x,q)$, as

$$\widehat{F}_\ell(y|x,q) = \varepsilon + \int_{\varepsilon}^{1-\varepsilon} 1\left(\mathbf{x}'\widehat{\beta}_\ell(\tau) \le y\right) d\tau, (y, x', q)' \in \mathcal{Y}_\ell \mathcal{X}_\ell \mathcal{Q}_\ell,$$

where $\mathcal{T}$ is restricted as $[\varepsilon, 1-\varepsilon]$ for some small constant $\varepsilon > 0$, and $\mathcal{Y}_\ell \mathcal{X}_\ell \mathcal{Q}_\ell$ is the product space of $\mathcal{Y}_\ell$, $\mathcal{X}_\ell$, and $\mathcal{Q}_\ell$ with $\mathcal{Y}_\ell$, $\mathcal{X}_\ell$, and $\mathcal{Q}_\ell$ being the interested area of y, x and q. Usually, $\mathcal{Y}_1\mathcal{X}_1$ and $\mathcal{Y}_2\mathcal{X}_2$ are the same, say, the support of y and x, and $\mathcal{Q}_1 = [\underline{q}, \gamma_0]$, $\mathcal{Q}_2 = [\gamma_0, \overline{q}]$ with $[\underline{q}, \overline{q}]$ being the support of q. Then the MD threshold effect, which is defined as

$$\Delta_D(y) = F_1(y) - F_2(y),$$

can be estimated by[2]

$$\widehat{\Delta}_D(y) = \widehat{F}_1(y) - \widehat{F}_2(y),$$

where

$$F_\ell(y) \equiv \int_{\mathcal{X}_\ell \mathcal{Q}_\ell} F_\ell(y|x,q) dF_\ell(x,q)$$

is the marginal distribution of y on the interested area of $(x', q)'$, $F_\ell(x,q)$ is the joint distribution of $(x_i', q_i)'$ truncated on $\mathcal{X}_\ell \mathcal{Q}_\ell$,

$$\widehat{F}_\ell(y) \equiv \int_{\mathcal{X}_\ell \mathcal{Q}_\ell} \widehat{F}_\ell(y|x,q) d\widehat{F}_\ell(x,q),$$

and

$$\widehat{F}_\ell(x,q) = n_\ell^{-1} \sum_{i=1}^{n} 1(x_i \le x, q_i \le q, x_i \in \mathcal{X}_\ell, q_i \in \mathcal{Q}_\ell), (x', q)' \in \mathcal{X}_\ell \mathcal{Q}_\ell$$

with $n_\ell = \sum_{i=1}^{n} 1(x_i \in \mathcal{X}_\ell, q_i \in \mathcal{Q}_\ell)$ is the empirical distributional function of $(x', q)'$ on $\mathcal{X}_\ell \mathcal{Q}_\ell$. So the MD threshold effect covers the difference not only in the conditional distribution of y given $(x', q)'$ but also in

[2] We can estimate $F_\ell(y)$ by the empirical distribution of y_i's such that $(x_i', q_i)' \in \mathcal{X}_\ell \mathcal{Q}_\ell$. However, incorporating the information in the covariates can improve the efficiency of such estimation.

the marginal distribution of $(x', q)'$ between these two regimes. Note also that $\Delta_D(y)$ and $\widehat{\Delta}_D(y)$ are well defined only on $\mathcal{Y} \equiv \mathcal{Y}_1 \cap \mathcal{Y}_2$. Given $\widehat{\Delta}_D$, we can estimate the MQ threshold effect

$$\Delta_Q(\tau) = F_1^{-1}(\tau) - F_2^{-1}(\tau) \equiv Q_1(\tau) - Q_2(\tau), \tau \in \mathcal{T}$$

by

$$\widehat{\Delta}_Q(\tau) = \widehat{F}_1^{-1}(\tau) - \widehat{F}_2^{-1}(\tau) \equiv \widehat{Q}_1(\tau) - \widehat{Q}_2(\tau), \tau \in \mathcal{T},$$

where $F_\ell^{-1} : \mathcal{Y}_\ell \longmapsto \mathcal{T}$ is the left-inverse function of F_ℓ. Note here that $\widehat{F}_\ell(y)$ is necessarily weakly increasing.

3 Asymptotics for the IQTRE

In this section, we derive the asymptotic distributions of $\widehat{\gamma}$ in two frameworks of the quantile threshold effects. We also discuss some valid inference methods for γ.

3.1 Asymptotics with Fixed Threshold Effects

Before stating the asymptotic theory for the IQTRE, we first specify some regularity conditions.

Assumption D:

1. $w_i \in \mathbb{W} \equiv \mathbb{R} \times \mathbb{X} \times \mathbb{Q} \subset \mathbb{R}^d$ are i.i.d. sampled. $\gamma \in \Gamma = [\underline{\gamma}, \overline{\gamma}] \subset \mathbb{R}$ with Γ being compact. γ_0 is in the interior of Γ.
2. $\beta_\ell(\tau) \in B \subset \mathbb{R}^d$ with B being compact for all $\tau \in \mathcal{T}$. $\beta_\ell^0(\tau)$ is in the interior of B. $\beta_1^0(\tau_t) \neq \beta_2^0(\tau_t)$ for at least one t, where $\neq$ means at least one element of the vector is not equal.
3. The conditional density $f(y|x, q)$ exists, and is bounded and uniformly continuous in y, uniformly in $\mathbf{x} \in \mathbb{X}$ and $q \leq \gamma_0$ $(q > \gamma_0)$.
4. $f_q(\cdot)$ is continuous, and $0 < \underline{f}_q \leq f_q(\gamma) \leq \overline{f}_q < \infty$ for $\gamma \in \Gamma$. $P\left(q < \underline{\gamma}\right) > 0$ and $P(q > \overline{\gamma}) > 0$.
5. $E\left[f_{e_{1\tau}|x,q}(0|x, q)\mathbf{x}\mathbf{x}'\middle| q\right]$, $E\left[f_{e_{2\tau}|x,q}(0|x, q)\mathbf{x}\mathbf{x}'\middle| q\right]$ and $E[\mathbf{x}\mathbf{x}'|\, q]$ are bounded and continuous in q for $q \in \mathcal{N}_\gamma$, where $\mathcal{N}_\gamma$ is a neighborhood of γ_0.
6. The minimum eigenvalue of $E\left[f_{e_{1\tau}|x,q}(0|x, q)\mathbf{x}\mathbf{x}'\middle| q\right]$ $(E\left[f_{e_{2\tau}|x,q}(0|x, q)\mathbf{x}\mathbf{x}'\middle| q\right])$ is bounded away from zero uniformly over $\tau \in \mathcal{T}$ and $q \leq \gamma_0$ $(q > \gamma_0)$, and the minimum eigenvalue of and $E[\mathbf{x}\mathbf{x}'|\, q]$ is bounded away from zero uniformly over $q \in \Gamma$.
7. $E\left[\|\mathbf{x}\|^{2+\varepsilon}\right] < \infty$ for some $\varepsilon > 0$.
8. Both z_{1Ti} and z_{2Ti} have continuous distributions, where z_{1Ti} follows the conditional distribution of $\overline{z}_{1Ti}$ given $q_i = \gamma_0$ with

$$\overline{z}_{1Ti} = \sum_{t=1}^{T} \overline{z}_{1\tau_t i} = \sum_{t=1}^{T} \left[\rho_{\tau_t}\left(e_{1\tau_t i} + \mathbf{x}_i'\beta_1^0(\tau_t) - \mathbf{x}_i'\beta_2^0(\tau_t)\right) - \rho_{\tau_t}\left(e_{1\tau_t i}\right)\right]$$

and z_{2Ti} follows the limiting conditional distribution of $\overline{z}_{2Ti}$ given $q_i = \gamma_0$ with

$$\overline{z}_{2Ti} = \sum_{t=1}^{T} \overline{z}_{2\tau_t i} = \sum_{t=1}^{T} \left[\rho_{\tau_t}\left(e_{2\tau_t i} + \mathbf{x}_i'\beta_2^0(\tau_t) - \mathbf{x}_i'\beta_1^0(\tau_t)\right) - \rho_{\tau_t}\left(e_{2\tau_t i}\right)\right].$$

Denote $z_{\ell Ti}$ as $\sum_{t=1}^{T} z_{\ell\tau_t i}$.

Assumption D1 is a standard assumption on the sample space and the parameter space of γ. Assumption D2 implies that there is a quantile threshold effect among $\{\tau_1, \cdots, \tau_T\}$ although this need not be true for each τ_t. If $\beta_1^0(\tau_t) = \beta_2^0(\tau_t)$ for some t, then $z_{\ell\tau_t i}$ will not appear in $z_{\ell Ti}$. Nevertheless, as long as there is one τ_t such that $\beta_1^0(\tau_t) \neq \beta_2^0(\tau_t)$, $z_{\ell Ti}$ will not degenerate, and $E[z_{\ell Ti}] > 0$ given that $\mathbf{0}$ is the unique minimizer of $E[\rho_{\tau_t}(e_{\ell\tau_t i} + \mathbf{x}_i'\beta) | q_i = \gamma_0]$ which is implied by Assumption D6. Note that $z_{\ell Ti}$ is quite different from $z_{\ell i}$ in least squares regression, where

$$\overline{z}_{1i} = \left(\mathbf{x}_i'\left(\beta_1^0 - \beta_2^0\right) + e_{1i}\right)^2 - e_{1i}^2 = \left\{2\mathbf{x}_i'\left(\beta_1^0 - \beta_2^0\right)e_{1i} + \left(\beta_1^0 - \beta_2^0\right)_i'\mathbf{x}_i\mathbf{x}_i'\left(\beta_1^0 - \beta_2^0\right)\right\},$$
$$\overline{z}_{2i} = \left(\mathbf{x}_i'\left(\beta_2^0 - \beta_1^0\right) + e_{2i}\right)^2 - e_{2i}^2 = \left\{-2\mathbf{x}_i'\left(\beta_1^0 - \beta_2^0\right)e_{2i} + \left(\beta_1^0 - \beta_2^0\right)'\mathbf{x}_i\mathbf{x}_i'\left(\beta_1^0 - \beta_2^0\right)\right\},$$

see, e.g., Chan (1993). As long as $\mathbf{x}_i$ is bounded, $\overline{z}_{\ell Ti}$ is bounded. However, $\overline{z}_{\ell i}$ is unbounded as long as $e_{\ell i}$ is unbounded. Since $\overline{z}_{\ell Ti}$ and $\overline{z}_{\ell i}$ are the only difference in the asymptotic distributions of the IQTRE and the LSE, the IQTRE is robust to outliers of y in the estimation of γ. This is understandable, since γ_0 is identified by the quantile differences in its left and right neighborhoods, and the quantile estimation is robust to outliers of y. Assumptions D3 and D7 are borrowed from Angrist et al. (2006) (Assumptions 2 and 4 of Theorem 3) and Chernozhukov et al. (2012) (Assumptions (b) and (d) of Condition QR) to facilitate the derivation of the weak limits of the CQ, MD and MQ threshold processes. Assumption D4 implies that γ is not on the boundary of the q's support. Assumption D5 imposes some restrictions on the continuity of $f_{e_{\ell\tau}|x,q}(0|x,q)f_{x|q}(x|q)$ for $q \in \mathcal{N}_\gamma$. Assumption D6 is stronger than the usual assumptions in QR, e.g., Assumption 3 of Theorem 3 in Angrist et al. (2006) or Assumption (c) of Condition QR in Chernozhukov et al. (2012). Combining with Assumption D4, it implies the usual assumptions in the current context, e.g.,

$$J_1(\tau) = E\left[\mathbf{x}\mathbf{x}'f_{e_{1\tau}|x,q}(0|x,q)1(q \leq \gamma_0)\right] \text{ and } J_2(\tau) = E\left[\mathbf{x}\mathbf{x}'f_{e_{2\tau}|x,q}(0|x,q)1(q > \gamma_0)\right] \tag{4}$$

are positive definite uniformly over $\tau \in \mathcal{T}$. Also, $E[\mathbf{x}\mathbf{x}'|\, q = \gamma_0] > 0$, combined with Assumption D2, excludes the continuous threshold model of Chan and Tsay (1998) for all $t = 1, \cdots, T$.

Assumption D8 need further explanation. This assumption guarantees that $n(\widehat{\gamma} - \gamma_0)$ is uniquely defined even asymptotically. It is not obvious that $z_{\ell Ti}$ has a continuous distribution. In the supplementary materials, we show that when q is the only covariate, the distribution of $z_{\ell\tau i}$ is a mixture of discrete and continuous, violating Assumption D8. We also provide some sufficient conditions to guarantee the uniqueness of $n(\widehat{\gamma} - \gamma_0)$ in large samples when the distribution of $z_{\ell\tau i}$ has discrete components. However, these conditions are not satisfied in a typical setup of this paper. Nevertheless, as long as $\mathbf{x}_i'\left(\beta_1^0(\tau) - \beta_2^0(\tau)\right)$ has a continuous distribution, $\overline{z}_{\ell\tau i}$ will have a continuous distribution. Correspondingly, $\overline{z}_{\ell Ti}$ will have a continuous distribution. This is because, conditional on $e_{\ell\tau i}$, $\overline{z}_{\ell\tau i}$ has a continuous conditional distribution, while the distribution of $\overline{z}_{\ell\tau i}$ is just the average of these conditional distributions so is continuous. To guarantee $\mathbf{x}_i'\left(\beta_1^0(\tau) - \beta_2^0(\tau)\right)$ to have a continuous distribution, we require only one element of $\mathbf{x}$ to be continuously distributed. This of course allows $\mathbf{x}$ to include discrete covariates, e.g., the intercept or a dummy variable. Since $z_{\ell\tau i}$ is defined conditional on $q_i = \gamma_0$, the element associated with q_i in $\mathbf{x}_i'\left(\beta_1^0(\tau) - \beta_2^0(\tau)\right)$ becomes constant, so we must require at least one element of x to be continuously distributed conditional on $q = \gamma_0$.

We now state the asymptotic distribution of $n(\widehat{\gamma} - \gamma_0)$.

Theorem 1 *Under Assumption D,*

$$n(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \arg\min_v D_T(v) \equiv Z_\gamma,$$

where

$$D_T(v) = \begin{cases} \sum_{i=1}^{N_1(|v|)} z_{1Ti}, \text{ if } v \leq 0; \\ \sum_{i=1}^{N_2(v)} z_{2Ti}, \text{ if } v > 0; \end{cases}$$

is cadlag with $D_T(0) = 0$*, and* $N_\ell(\cdot)$ *is a Poisson process with intensity* $f_q(\gamma_0)$*,* $\ell = 1, 2$*. Furthermore,* $\{z_{1Ti}, z_{2Ti}\}_{i \geq 1}$*,* $N_1(\cdot)$ *and* $N_2(\cdot)$ *are independent of each other.*

The simplest corollary of Theorem 1 is that when $T = 1$ and $\beta_1^0(\tau) \neq \beta_2^0(\tau)$, $D_T(v)$ in the theorem is changed to $D_\tau(v)$, where $D_\tau(v)$ is defined similar as $D_T(v)$ except that $z_{\ell\tau i}$ replaces $z_{\ell Ti}$. In the supplementary materials, we discuss the conditions to guarantee the symmetry of the distribution of $\arg\min_v D_\tau(v)$. Also, given that $E[z_{\ell\tau i}] > 0$ when $\beta_1^0(\tau) \neq \beta_2^0(\tau)$, $E[z_{\ell Ti}] \geq E[z_{\ell\tau i}]$, which implies that $\widehat{\gamma}$ is more efficient than $\widehat{\gamma}_\tau$ which is based on a single quantile threshold regression.

In this framework, Yu (2008) shows that γ can be adaptively estimated by the semiparametric empirical Bayes estimator (SEBE) as long as an n-consistent estimator of γ is available. In his Section 5.2, the author also mentions that the initial estimator of γ is very important for the performance of the SEBE in finite samples.[3] Given that the IQTRE is more efficient than the usual LSE and the LADE, we expect that the SEBE started from the IQTRE performs better than that started from the LSE or LADE. The simulation studies in Section 6 will convince this result.

3.2 Asymptotics with Shrinking Threshold Effects

Another popular asymptotic framework in quantile threshold regression is to assume the quantile threshold effects δ_n shrinking to zero asymptotically, where

$$\delta_n = (\delta_{1n}', \cdots, \delta_{Tn}')' \text{ with } \delta_{tn} \equiv \beta_1^0(\tau_t) - \beta_2^0(\tau_t).$$

This framework is suitable to the case where the quantile threshold effects are relatively small for the given sample size.

Theorem 2 *Under Assumptions D1-D7 and* $\|\delta_n\| \to 0$*,* $\sqrt{n}\|\delta_n\| \to \infty$*,*

$$n f_q(\gamma_0) \left(\frac{\pi_{1Tn}}{\sigma_{Tn}}\right)^2 (\widehat{\gamma} - \gamma_0) \xrightarrow{d} \Lambda(\xi),$$

where

$$\pi_{\ell Tn} = \sum_{t=1}^T \delta_{tn}' E\left[f_{e_{\ell\tau_t}|x,q}(0|x,q)\mathbf{x}\mathbf{x}'|q = \gamma_0\right]\delta_{tn},$$
$$\sigma_{Tn}^2 = \sum_{t=1}^T \sum_{t'=1}^T (\tau_t \wedge \tau_{t'} - \tau_t\tau_{t'})\delta_{tn}' E[\mathbf{x}\mathbf{x}'|q = \gamma_0]\delta_{t'n},$$

$$\Lambda(\xi) = \arg\max_v \begin{cases} W_1(-v) - \frac{|v|}{2}, & \text{if } v \leq 0, \\ W_2(v) - \xi\frac{|v|}{2}, & \text{if } v > 0, \end{cases}$$

with

$$\xi = \lim_{n\to\infty} \frac{\pi_{2Tn}}{\pi_{1Tn}} \in (0, \infty),$$

[3] The simulation studies in Oka and Qu (2011) show that the performance of CIs for $\delta_{tn} \equiv \beta_1^0(\tau_t) - \beta_2^0(\tau_t)$ also critically depends on the precision of $\widehat{\gamma}$.

and $W_\ell(v)$, $\ell = 1, 2$, being two independent standard Wiener processes defined on $[0, \infty)$.

This asymptotic result is parallel to Corollary 1 of Oka and Qu (2011) and Theorem 1 of Caner (2002). However, we do not assume $\|\delta_{tn}\|$, $t = 1, \cdots, T$, to have the same convergence rate as in Oka and Qu (2011) or even each component of δ_{tn} for a fixed t to have the same convergence rate $n^{-\alpha}$, $0 < \alpha < 1/2$, as in Caner (2002), so some $\|\delta_{tn}\|$ may be smaller (or even zero) than others. Given that $E\left[f_{e_{1\tau}}(0|x,q)\mathbf{x}\mathbf{x}'|q = \gamma_0\right]$ and $E\left[\mathbf{x}\mathbf{x}'|q = \gamma_0\right]$ are positive definite for any $\tau \in \mathcal{T}$, the convergence rate is determined by $n \max_{t=1,\cdots,T} \|\delta_{tn}\|^2$ (or $n\delta_n'\delta_n$). In other words, the convergence rate is determined by the largest quantile threshold effect among all τ_t, $t = 1, \cdots, T$. If $\delta_{tn} = \delta_t n^{-\alpha}$ with $\delta_t \neq 0$ for all $t = 1, \cdots, T$, then the convergence rate is $n^{1-2\alpha}$, the same rate as in Theorem 1 of Caner (2002). In this case, we can calculate ξ and the normalizing coefficient $n^{2\alpha}\left(\frac{\pi_{1Tn}}{\sigma_{Tn}}\right)^2$ for the two examples in Section 2. In the first example,

$$\xi = \frac{\sum_{t=1}^{T}\delta_t' E\left[f_e(F_e^{-1}(\tau_t))/\sigma_{20}\mathbf{x}\mathbf{x}'|q = \gamma_0\right]\delta_t}{\sum_{t=1}^{T}\delta_t' E\left[f_e(F_e^{-1}(\tau_t))/\sigma_{10}\mathbf{x}\mathbf{x}'|q = \gamma_0\right]\delta_t} = \frac{\sigma_{10}}{\sigma_{20}},$$

$$n^{2\alpha}\left(\frac{\pi_{1Tn}}{\sigma_{Tn}}\right)^2 = \frac{\left(\sum_{t=1}^{T} f_e(F_e^{-1}(\tau_t))\delta_t' E\left[\mathbf{x}\mathbf{x}'|q = \gamma_0\right]\delta_t\right)^2}{\sigma_{10}^2 \sum_{t=1}^{T}\sum_{t'=1}^{T}(\tau_t \wedge \tau_{t'} - \tau_t\tau_{t'})\delta_t' E\left[\mathbf{x}\mathbf{x}'|q = \gamma_0\right]\delta_{t'}},$$

and in the second example,

$$\xi = \frac{\sum_{t=1}^{T} f_e(F_e^{-1}(\tau_t))\delta_t' E\left[\frac{\mathbf{x}\mathbf{x}'}{\mathbf{x}'\lambda_2}\middle| q = \gamma_0\right]\delta_t}{\sum_{t=1}^{T} f_e(F_e^{-1}(\tau_t))\delta_t' E\left[\frac{\mathbf{x}\mathbf{x}'}{\mathbf{x}'\lambda_1}|q = \gamma_0\right]\delta_t}, \; n^{2\alpha}\left(\frac{\pi_{1Tn}}{\sigma_{Tn}}\right)^2 = \frac{\left(\sum_{t=1}^{T} f_e(F_e^{-1}(\tau_t))\delta_t' E\left[\frac{\mathbf{x}\mathbf{x}'}{\mathbf{x}'\lambda_1}\middle| q = \gamma_0\right]\delta_t\right)^2}{\sum_{t=1}^{T}\sum_{t'=1}^{T}(\tau_t \wedge \tau_{t'} - \tau_t\tau_{t'})\delta_t' E\left[\mathbf{x}\mathbf{x}'|q = \gamma_0\right]\delta_{t'}}.$$

The distribution of $\Lambda(\xi)$ can be found in Appendix B of Bai (1997). When $\xi = 1$, this distribution is symmetric.

We now discuss the efficiency of $\widehat{\gamma}$ in this framework. First, different from the efficiency results in the framework with fixed threshold effects, it is hard to compare the efficiency of $\widehat{\gamma}$ and $\widehat{\gamma}_\tau$: $n^{2\alpha}\left(\frac{\pi_{1Tn}}{\sigma_{Tn}}\right)^2$ need not be an increasing functions of T. Second, it is hard to compare the efficiency of $\widehat{\gamma}_\tau$ and $\widehat{\gamma}_{LSE}$ even for $\tau = 0.5$ and $\delta_n = \delta n^{-\alpha}$ in the first example of Section 2. In this case,

$$n^{1-2\alpha}\left(\widehat{\gamma}_{0.5} - \gamma_0\right) \xrightarrow{d} \frac{\sigma_{10}^2}{4f_e(0)^2 f_q(\gamma_0)\delta' E\left[\mathbf{x}\mathbf{x}'|q = \gamma_0\right]\delta}\widetilde{\Lambda}\left(1, \frac{\sigma_{10}}{\sigma_{20}}\right),$$

while

$$n^{1-2\alpha}\left(\widehat{\gamma}_{LSE} - \gamma_0\right) \xrightarrow{d} \frac{\sigma_{10}^2}{f_q(\gamma_0)\delta' E\left[\mathbf{x}\mathbf{x}'|q = \gamma_0\right]\delta}\widetilde{\Lambda}\left(\frac{\sigma_{20}^2}{\sigma_{10}^2}, 1\right),$$

where

$$\widetilde{\Lambda}(\phi, \xi) = \arg\max_v \begin{cases} W_1(-v) - \frac{|v|}{2}, & \text{if } v \leq 0, \\ \sqrt{\phi}W_2(v) - \xi\frac{|v|}{2}, & \text{if } v > 0, \end{cases} \tag{5}$$

The distributions of $\widetilde{\Lambda}\left(1, \frac{\sigma_{10}}{\sigma_{20}}\right)$ and $\widetilde{\Lambda}\left(\frac{\sigma_{20}^2}{\sigma_{10}^2}, 1\right)$ are not the same and hard to compare, e.g., the density of $\widetilde{\Lambda}\left(1, \frac{\sigma_{10}}{\sigma_{20}}\right)$ is continuous while $\widetilde{\Lambda}\left(\frac{\sigma_{20}^2}{\sigma_{10}^2}, 1\right)$is not unless $\sigma_{10} = \sigma_{20}$ (see Appendix B of Bai (1997)). Only in the special case $\sigma_{10} = \sigma_{20}$, the relative efficiency of the LADE and the LSE is determined by the relative magnitude of $Var(e)$ and $(2f_e(0))^{-2}$, just as in the usual comparison between the LSE and the LADE in regular models (see page 415 of Bai (1995) and page 805 of Caner (2002)).

3.3 Inference Methods

In the framework with fixed threshold effects, Yu (2008) show that the nonparametric posterior interval (NPI) started from any n-consistent estimator of γ is a valid CI for γ and performs the best among all available CIs. Given that the IQTRE performs better than the LSE and the LADE, we expect that the NPI started from the IQTRE should perform better than that started from the LSE or the LADE in finite samples. The simulation studies in Section 6 convince this result. The algorithms for the SEBE and NPI are included in the supplementary materials.

In the framework with shrinking threshold effects, we can construct the Wald-type CI by inverting the asymptotic distribution of $\widehat{\gamma}$ in Theorem 2 as in Oka and Qu (2011). However, due to the identification failure when $\delta_n = 0$, this Wald-type CI performs poorly, which is confirmed in the simulation studies of Yu (2008). Hansen (2000) suggests to construct a CI for γ in the least squares estimation by inverting the likelihood ratio (LR) statistic; this method is also used in the LAD estimation of Caner (2002). The likelihood ratio statistic can be used to test whether $\gamma = \gamma_0$, and is constructed in our context as

$$LR_n(\gamma) = n\frac{\pi_{1Tn}}{\sigma_{Tn}^2}\left(Q_{Tn}(\gamma) - Q_{Tn}(\widehat{\gamma})\right)$$

Corollary 1 *Under Assumptions D1-D7 and* $\|\delta_n\| \to 0$, $\sqrt{n}\|\delta_n\| \to \infty$,

$$LR_n(\gamma_0) \xrightarrow{d} M,$$

where M *follows the distribution* $P(M \le z) = (1-e^{-z})(1-e^{-\xi z})$, *where* ξ *is defined in Theorem 2.*

To construct a CI for γ based on $LR_n(\gamma)$, we need to estimate π_{1Tn}, σ_{Tn}^2 and ξ.[4] Along the lines of Hansen (2000, Section 3.4), let $r_{\ell ti} = \left(\delta_{tn}'\mathbf{x}_i\right)^2 f_{e_{\ell\tau_t}}(0|x_i,q_i)$, and $r_{3ti} = \left(\delta_{tn}'\mathbf{x}_i\right)^2$; then

$$\pi_{\ell Tn} = \sum_{t=1}^{T} E\left[r_{\ell ti}|q_i = \gamma_0\right],\ \sigma_{Tn}^2 = \sum_{t=1}^{T}\sum_{t'=1}^{T}(\tau_t \wedge \tau_{t'} - \tau_t\tau_{t'})E\left[r_{3ti}|q = \gamma_0\right].$$

So we can estimate $\pi_{\ell Tn}$ and σ_{Tn}^2 by standard nonparametric techniques such as kernel smoothing or series approximation (see, e.g., Härdle and Linton (1994), Pagan and Ullah (1999), Ichimura and Todd (2007), and Li and Racine (2007) for an introduction). For example, suppose the kernel smoothing is used; then in finite samples, $r_{\ell ti}$ can be replaced by $\widehat{r}_{\ell ti} = \left(\widehat{\delta}_{tn}'\mathbf{x}_i\right)^2 K_{h_{\ell t}}(\widehat{e}_{\ell\tau_t i})$, r_{3ti} is replaced by $\widehat{r}_{3ti} = \left(\widehat{\delta}_{tn}'\mathbf{x}_i\right)^2$, and γ_0 is replaced by $\widehat{\gamma}$, where $\widehat{\delta}_{tn} = \widehat{\beta}_1(\tau_t) - \widehat{\beta}_2(\tau_t)$, and $K_h(\cdot) = K(\cdot/h)/h$ with $K(\cdot)$ being a kernel function and h being the bandwidth. Alternatively, $\widehat{r}_{\ell ti}$ can be $\left(\widehat{\delta}_{tn}'\mathbf{x}_i\right)^2 \frac{2h_{\ell t}}{\mathbf{x}_i'\widehat{\beta}_\ell(\tau_t+h_{\ell t}) - \mathbf{x}_i'\widehat{\beta}_\ell(\tau_t-h_{\ell t})}$, where $f_{e_{\ell\tau_t}}(0|x_i,q_i)$ is estimated by the difference quotient estimator. The kernel function is not crucial in kernel smoothing. As to the bandwidth selection, see Koenker (1994) for some practical suggestions. Given all these components, the $(1-\alpha)$ LR-CI for γ is

$$\left\{\gamma : \widehat{LR}_n(\gamma) \le \widehat{\text{crit}}\right\},$$

where $\widehat{LR}_n(\gamma)$ replaces π_{1Tn} and σ_{Tn}^2 in $LR_n(\gamma)$ by their estimates, and $\widehat{\text{crit}}$ is the $(1-\alpha)$ quantile of M with ξ being substituted by its estimate.

[4] Note that the LR-CI does not need to estimate $f_q(\gamma_0)$ as in the Wald-type CI.

4 Asymptotics for the CQ, MD and MQ Threshold Effects

To state the weak limits of the MD and MQ threshold processes, we first state the weak limit of $\widehat{\beta}(\cdot)$. The following theorem shows that the weak limit of $\widehat{\beta}(\cdot)$ is not affected by the estimation of γ, and is actually independent of the asymptotic distribution of $\widehat{\gamma}$ in both frameworks. We first define the meaning of the independence between a random variable (r.v.) and a stochastic process and between two stochastic processes.

Definition 1 *A r.v. X and a stochastic process $Y(\tau)$ indexed by $\tau \in \mathcal{T}$ are said to be independent if X is independent of all the finite dimensional marginals $(Y(\tau_1), \cdots, Y(\tau_k))$ of $Y(\tau)$. Two stochastic processes $X(\tau)$ and $Y(\tau)$ indexed by $\tau \in \mathcal{T}$ are said to be independent if all the finite dimensional marginals of the two processes $(X(\tau_1), \cdots, X(\tau_k))$, $(Y(\tau'_1), \cdots, Y(\tau'_l))$ are two vector r.v.'s independent from each other.*

Theorem 3 *Under Assumption D and $\sup_{\tau \in \mathcal{T}} \|\delta_0(\tau)\| = O(\|\delta_n\|)$, no matter δ_n is fixed or shrinks to zero,*

$$J_1(\cdot)\sqrt{n}\left(\widehat{\beta}_1(\cdot) - \beta_1^0(\cdot)\right) \rightsquigarrow Z_1(\cdot),$$
$$J_2(\cdot)\sqrt{n}\left(\widehat{\beta}_2(\cdot) - \beta_2^0(\cdot)\right) \rightsquigarrow Z_2(\cdot),$$

where $J_\ell(\cdot)$ is defined in (4), $Z_1(\cdot)$ is a zero-mean Gaussian process with the covariance function

$$\Sigma_1(\tau, \tau') \equiv E[Z_1(\tau)Z_1(\tau')] = (\min(\tau, \tau') - \tau\tau') E[\mathbf{x}\mathbf{x}' 1(q \leq \gamma_0)],$$

and $Z_2(\cdot)$ is similarly defined with the covariance function

$$\Sigma_2(\tau, \tau') \equiv E[Z_2(\tau)Z_2(\tau')] = (\min(\tau, \tau') - \tau\tau') E[\mathbf{x}\mathbf{x}' 1(q > \gamma_0)].$$

Furthermore, Z_γ in Theorem 1 (or $\Lambda(\xi)$ in Theorem 2), $Z_1(\cdot)$ and $Z_2(\cdot)$ are independent.

If δ_n is fixed, $\sup_{\tau \in \mathcal{T}} \|\delta_0(\tau)\| = O(\|\delta_n\|)$ automatically holds. If $\|\delta_n\| \to 0$, $\sup_{\tau \in \mathcal{T}} \|\delta_0(\tau)\| = O(\|\delta_n\|)$ means that the threshold effects at $\{\tau_t\}_{t=1}^T$ are the largest (in rate) possible threshold effects. An immediate corollary of the Theorem above is the weak limit of the CQ threshold process, $\widehat{\delta}(\cdot) \equiv \widehat{\beta}_1(\cdot) - \widehat{\beta}_2(\cdot)$.

Corollary 2 *Under Assumption D, no matter δ_n is fixed or shrinks to zero,*

$$\sqrt{n}\left[\widehat{\delta}(\cdot) - \delta_0(\cdot)\right] \rightsquigarrow J_1(\cdot)^{-1}Z_1(\cdot) - J_2(\cdot)^{-1}Z_2(\cdot),$$

where the process on the right-hand side is a zero-mean Gaussian process with the covariance function

$$\Sigma(\tau, \tau') = (\min(\tau, \tau') - \tau\tau')\left\{J_1(\tau)^{-1}E[\mathbf{x}\mathbf{x}' 1(q \leq \gamma_0)] J_1(\tau')^{-1} + J_2(\tau)^{-1}E[\mathbf{x}\mathbf{x}' 1(q > \gamma_0)] J_2(\tau')^{-1}\right\}.$$

Proof. This result follows by the continuous mapping theorem and the independence between $Z_1(\cdot)$ and $Z_2(\cdot)$. ■

We now state the weak limits of the MD and MQ threshold processes.

Theorem 4 *Suppose Assumption D holds, $\mathcal{X}_\ell$, $\mathcal{Q}_\ell$ and $\mathcal{Y}$ are compact, and $\mathcal{T}_\ell \equiv \{\tau : \mathbf{x}'\beta_\ell(\tau) \in \mathcal{Y}$ for some $x \in \mathcal{X}_\ell$ and $q \in \mathcal{Q}_\ell\} \subseteq \mathcal{T}$. Then*

$$\sqrt{n}\left(\widehat{\Delta}_D(y) - \Delta_D(y)\right) \rightsquigarrow \mathbb{G}_1(\kappa_{1,y}) - \mathbb{G}_2(\kappa_{2,y}),$$

where $\mathbb{G}_1(\kappa_{1,y})$ *and* $\mathbb{G}_2(\kappa_{2,y})$ *are two independent zero-mean Gaussian processes with the the covariance function*

$$E\left[\mathbb{G}_\ell(\kappa_{\ell,y})\mathbb{G}_\ell(\kappa_{\ell,y'})\right] = \int \kappa_{\ell,y}(\mathsf{y},\mathsf{x},\mathsf{q})\kappa_{\ell,y'}(\mathsf{y},\mathsf{x},\mathsf{q})dF(\mathsf{y}|\mathsf{x},\mathsf{q})dF_\ell(\mathsf{x},\mathsf{q}) - \int \kappa_{\ell,y}(\mathsf{y},\mathsf{x},\mathsf{q})dF(\mathsf{y}|\mathsf{x},\mathsf{q})dF_\ell(\mathsf{x},\mathsf{q})\int \kappa_{\ell,y'}(\mathsf{y},\mathsf{x},\mathsf{q})dF(\mathsf{y}|\mathsf{x},\mathsf{q})dF_\ell(\mathsf{x},\mathsf{q}),$$

and

$$\begin{aligned}
\kappa_{\ell,y}(\mathsf{y},\mathsf{x},\mathsf{q}) &= \int f(y|x,q)x'\psi_{\ell,F(y|x,q)}(\mathsf{y},\mathsf{x},\mathsf{q})dF_\ell(x,q) + \sqrt{s_\ell}F(y|\mathsf{x},\mathsf{q}),\\
\psi_{1,\tau}(\mathsf{y},\mathsf{x},\mathsf{q}) &= J_1(\tau)^{-1}\left\{\tau - 1(\mathsf{y}\le(1,\mathsf{x}',\mathsf{q})\beta_1(\tau))\right\}1(\mathsf{q}\le\gamma_0)(1,\mathsf{x}',\mathsf{q})',\\
\psi_{2,\tau}(\mathsf{y},\mathsf{x},\mathsf{q}) &= J_2(\tau)^{-1}\left\{\tau - 1(\mathsf{y}\le(1,\mathsf{x}',\mathsf{q})\beta_2(\tau))\right\}1(\mathsf{q}>\gamma_0)(1,\mathsf{x}',\mathsf{q})',
\end{aligned}$$

s_ℓ *is the probability limit of* n/n_ℓ*. If in addition* $F_\ell(y)$ *admits a positive continuous density* $f_\ell(y)$ *on an interval* $[a,b]$ *containing an* ϵ*-enlargement of the set* $\{Q_\ell(\tau):\tau\in\mathcal{T}\}$*, then*

$$\sqrt{n}\left(\widehat{\Delta}_Q(\tau)-\Delta_Q(\tau)\right) \rightsquigarrow \mathbb{G}_2\left(\kappa_{2,Q_2(\tau)}\right)/f_2(Q_2(\tau)) - \mathbb{G}_1\left(\kappa_{1,Q_1(\tau)}\right)/f_1(Q_1(\tau)).$$

Proof. The first part of this theorem is a direct corollary of Theorem 5.1 (3) in Chernozhukov et al. (2012) and Theorem 3 above. The second part of the theorem is a direct corollary of Theorem 4.1 (2) of Chernozhukov et al. (2012). ■

The randomness in $\kappa_{\ell,y}(\mathsf{y},\mathsf{x},\mathsf{q})$ includes two parts: the first part comes from the estimation of the conditional quantile, and the second part comes from the estimation of the marginal distribution of $(x',q)'$. $\mathbb{G}_1(\cdot)$ and $\mathbb{G}_2(\cdot)$ are independent because the randomnesses in $\widehat{F}_1(\cdot)$ and $\widehat{F}_2(\cdot)$ are independent.

4.1 Asymptotic Inference Methods

Inference on the QR process $\beta(\tau)$ is useful for testing basic hypotheses of the form

$$R(\tau)'\beta(\tau) = r(\tau) \text{ for all } \tau\in\mathcal{T}, \tag{6}$$

where $R(\tau)\in\mathbb{R}^{p\times 2d}$ and $r(\tau)\in\mathbb{R}^{p\times 1}$. We give a few examples here.

Example 1 *We may be interested in whether a variable or a subset of variables* $j\in\{l+1,\cdots,d\}$ *enters models for all conditional quantiles with zero coefficients, i.e., whether* $\beta_{\ell j}(\tau)=0$ *for all* $\tau\in\mathcal{T}$ *and* $j\in\{l+1,\cdots,d\}$*. This corresponds to*

$$R(\tau)' = \begin{pmatrix} 0_{(d-l)\times l} & I_{d-l} & 0_{(d-l)\times l} & 0_{(d-l)\times(d-l)} \\ 0_{(d-l)\times l} & 0_{(d-l)\times(d-l)} & 0_{(d-l)\times l} & I_{d-l}\end{pmatrix} \text{ and } r(\tau)=0_{2(d-l)},$$

where $0_{d_1\times d_2}$ *is a* $d_1\times d_2$ *matrix with all elements being zero, and* I_d *is a* $d\times d$ *identity matrix.*

Example 2 *Even if we reject the hypothesis that there is no quantile threshold effect (see the next section), we may still be interested in whether* $\beta_{1j}(\tau)=\beta_{2j}(\tau)$ *for all* $\tau\in\mathcal{T}$ *and* $j\in\{l+1,\cdots,d\}$*,* $l\ge 1$*; e.g., when* $l=1$*, we are interested in whether all the slope parameters are equal. Correspondingly,*

$$R(\tau)' = \begin{pmatrix} 0_{(d-l)\times l} & I_{d-l} & 0_{(d-l)\times l} & -I_{d-l}\end{pmatrix} \text{ and } r(\tau)=0_{d-l}.$$

Example 3 *We may want to check whether the model is from a location shift model or a location-scale shift model. In the former case,*

$$R(\tau)' = \begin{pmatrix} 0_{(d-1)\times 1} & I_{d-1} & 0_{(d-1)\times 1} & 0_{(d-1)\times(d-1)} \\ 0_{(d-1)\times 1} & 0_{(d-1)\times(d-1)} & 0_{(d-1)\times 1} & I_{d-1} \end{pmatrix} \text{ and } r(\tau) = \left(\underset{\sim}{\beta}_1', \underset{\sim}{\beta}_2'\right)'.$$

It asserts simply that the quantile regression slopes are constant, independent of τ. In the latter case,

$$R(\tau)' = \begin{pmatrix} \operatorname{diag}\left(\lambda_{1i}^{-1}\right) & 0_{d\times d} \\ 0_{d\times d} & \operatorname{diag}\left(\lambda_{2i}^{-1}\right) \end{pmatrix} \text{ and } r(\tau) = \begin{pmatrix} \mathbf{1}_d\xi_\tau \\ \mathbf{1}_d\xi_\tau \end{pmatrix} + \left(\beta_{11}\lambda_{11}^{-1}, \cdots, \beta_{1d}\lambda_{1d}^{-1}, \beta_{21}\lambda_{21}^{-1}, \cdots, \beta_{2d}\lambda_{2d}^{-1}\right)',$$

where $\mathbf{1}_d$ is a $d \times 1$ vector of ones.[5] *Usually, ξ_τ is unknown under the null and it is convenient to choose one coordinate, typically the intercept coefficient, to play the role of numeraire. Then*

$$R(\tau)' = \begin{pmatrix} \sigma_1 & -I_{d-1} & 0_{(d-1)\times 1} & 0_{(d-1)\times(d-1)} \\ 0_{(d-1)\times 1} & 0_{(d-1)\times(d-1)} & \sigma_2 & -I_{d-1} \end{pmatrix} \text{ and } r(\tau) = -\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$

where $\mu_{\ell i} = \beta_{\ell i} - \beta_{\ell 1}\lambda_{\ell i}/\lambda_{\ell 1}$ and $\sigma_{\ell i} = \lambda_{\ell i}/\lambda_{\ell 1}$, $i = 2, \cdots, d$.

Example 3 is different from Example 1 and 2 since nuisance parameters are involved in $R(\tau)$ and $r(\tau)$. Fortunately, as suggested by Koenker and Xiao (2002), the Khmaladze (1981)'s transformation can be applied to deal with such Durbin problems. We concentrate on the case where $R(\tau)$ and $r(\tau)$ are known. Especially, this case can be used to construct simultaneous (uniform) confidence intervals for linear functions of parameters $R(\tau)'\beta(\tau) - r(\tau)$ for all $\tau \in \mathcal{T}$, e.g., the CQ threshold effect $\beta_{1j}(\tau) - \beta_{2j}(\tau)$, $j = 1, \cdots, d$, for all $\tau \in \mathcal{T}$. The following corollary states the asymptotic distribution of the test statistic $\mathcal{K}_n \equiv \sup_{\tau\in\mathcal{T}} \left\| V(\tau)^{-1/2}\sqrt{n}\left(R(\tau)'\widehat{\beta}(\tau) - r(\tau)\right)\right\|$ for testing (6), where $\|\cdot\|$ can be $\|\cdot\|_1$, $\|\cdot\|_2$ or $\|\cdot\|_\infty$, $V(\tau) \equiv R(\tau)' \mathsf{J}(\tau)^{-1}\mathbf{\Sigma}(\tau,\tau)\mathsf{J}(\tau)^{-1}R(\tau)$ with $\mathsf{J}(\tau) \equiv \operatorname{diag}\{J_1(\tau), J_2(\tau)\}$ and $\mathbf{\Sigma}(\tau,\tau) \equiv \operatorname{diag}\{E\left[\mathbf{x}\mathbf{x}'1\left(q \le \gamma_0\right)\right], E\left[\mathbf{x}\mathbf{x}'1\left(q > \gamma_0\right)\right])\}$.

Corollary 3 *Under Assumption D, no matter δ_n is fixed or shrinks to zero, $\mathcal{K}_n$ converges in distribution to $\sup_{\tau\in\mathcal{T}} \|B_p(\tau)\|$, where B_p is the standard p-dimensional Brownian bridge. The result is not affected by replacing $\mathsf{J}(\tau)$ and $\mathbf{\Sigma}(\tau,\tau)$ with estimates that are consistent uniformly in $\tau \in \mathcal{T}$.*

Proof. This result follows by the continuous mapping theorem in $\ell^\infty(\mathcal{T})$. ■

Thus, $\mathcal{K}_n$ has a well-behaved limit distribution. In practice, by stochastic equicontinuity of the QR process, we can replace any continuum of quantile indices $\mathcal{T}$ by a finite grid $\mathcal{T}_{\mathcal{K}_n}$, where the distance between adjacent grid points goes to zero as $n \to \infty$. The critical values of $\sup_{\tau\in\mathcal{T}} \|B_p(\tau)\|_1$ and $\sup_{\tau\in\mathcal{T}} \|B_p(\tau)\|_2$ can be found in the electronic appendix of Koenker and Xiao (2002) and Table 1 of Andrews (1993), respectively.[6] Given the α quantile of $\sup_{\tau\in\mathcal{T}} \|B_p(\tau)\|$, say, $\kappa(\alpha)$, the asymptotic simultaneous $(1-\alpha)$ confidence band is $I_n(\tau) = \left[R(\tau)'\widehat{\beta}(\tau) - r(\tau) \pm \kappa(1-\alpha)\cdot\sqrt{\widehat{V}(\tau)}\Big/\sqrt{n}\right]$ for a uniformly consistent estimator $\widehat{V}(\tau)$ of $V(\tau)$ over $\tau \in \mathcal{T}$.

The inference procedure above requires estimators of $\mathsf{J}(\tau)$ and $\mathbf{\Sigma}(\tau,\tau)$ that are uniformly consistent in $\tau \in \mathcal{T}$. For $\mathbf{\Sigma}(\tau,\tau)$, we need only estimate $E\left[\mathbf{x}\mathbf{x}'1\left(q \le \gamma_0\right)\right]$ and $E\left[\mathbf{x}\mathbf{x}'1\left(q > \gamma_0\right)\right]$ by their sample analogs.

[5] Obviously, there is some difficulty if there are $\lambda_{\ell i}$ equal to zero. In such cases, we can take $\lambda_{\ell i} = 1$, and set the corresponding elements $r_i(\tau) = \beta_{\ell i}$.

[6] Strictly speaking, the critical values in Koenker and Xiao (2002) are designed for $\sup_{\tau\in\mathcal{T}} \|W_p(\tau)\|$, where $W_p(\tau)$ is the standard p-dimensional Brownian motion, and the critical values in Andrews (1993) are for $\sup_{\tau\in\mathcal{T}} \|Q_p(\tau)\|$, where $Q_p(\tau) = B_p(\tau)/\tau(1-\tau)$ is the standard p-dimensional Bessel process. Nevertheless, the simulation method used in these papers can be applied without any difficulty in our context to obtain the critical values. Also, the resampling methods in the next subsection are very convenient in practice.

As to $\mathsf{J}(\tau)$, a popular estimator is the kernel estimator

$$\widehat{J}_1(\tau) = \frac{1}{n}\sum_{i=1}^{n} K_{h_1}\left(\widehat{e}_{1\tau i}\right)\mathbf{x}_i\mathbf{x}_i' 1\left(q_i \leq \widehat{\gamma}\right), \widehat{J}_2(\tau) = \frac{1}{n}\sum_{i=1}^{n} K_{h_2}\left(\widehat{e}_{2\tau i}\right)\mathbf{x}_i\mathbf{x}_i' 1\left(q_i > \widehat{\gamma}\right), \tag{7}$$

where $K_h(\cdot) = K(\cdot/h)/h$ with $K(\cdot)$ being a kernel function and h being the bandwidth, $\widehat{e}_{1\tau i} = y_i - \mathbf{x}_i'\widehat{\beta}_1(\tau)$ for i such that $q_i \leq \widehat{\gamma}$ and $\widehat{e}_{2\tau i} = y_i - \mathbf{x}_i'\widehat{\beta}_2(\tau)$ for i such that $q_i > \widehat{\gamma}$. The following corollary states the uniform consistency of this estimator.

Corollary 4 *Suppose $K(\cdot)$ is a nonnegative, symmetric function of bounded variation (but not necessarily continuous), and $\int K(u)du = 1$, $\int |K(u)|\,du < \infty$, $uK(u) \to 0$ as $|u| \to \infty$. $h_\ell \to 0$ and $nh_\ell^2 \to \infty$. Then under Assumption D and the additional assumption that $E\left[\|\mathbf{x}\|^4\right] < \infty$ and $\int |f(y|x,q)|\,dy < \infty$ for a.s. $(x',q)'$, $\widehat{J}_\ell(\tau)$ is consistent uniformly in $\tau \in \mathcal{T}$.*

Powell (1986) and Buchinsky and Hahn (1998) use the uniform kernel, and Buchinsky (1998) also considers the normal kernel; see further discussions in Powell (1989). When $K(\cdot)$ has a bounded support, the assumptions $\int |K(u)|\,du < \infty$, $uK(u) \to 0$ as $|u| \to \infty$, and $\int |f(y|x,q)|\,dy < \infty$ for a.s. $(x',q)'$ are not required. As to the bandwidth, Koenker (1994) suggests $h_\ell = C_\ell \cdot n_\ell^{-1/3}$, where $n_1 = \sum_{i=1}^n 1(q_i \leq \widehat{\gamma})$, $n_2 = \sum_{i=1}^n 1(q_i > \widehat{\gamma})$, and C_ℓ can be obtained from Hall and Sheather (1988). Another popular estimator of $f_{e_{\ell\tau}|x,q}(0|x,q)$ in $\mathsf{J}(\tau)$ is the difference quotient estimator, but its uniform consistency is doubtful.

4.2 Resampling Inference Methods

The asymptotic methods in the last subsection may be useful for the uniform inference of the CQ threshold effects. However, for the MD and MQ threshold effects, the asymptotic methods are not applicable since the limit processes in Theorem 4 are non-pivotal and their covariance functions depend on complicated unknown, though estimable, nuisance parameters. In other words, the Durbin problem appears again in this context. A popular alternative of the asymptotic methods is the resampling methods, especially, the *exchangeable bootstrap*. This procedure incorporates many popular forms of resampling as special cases, namely the empirical bootstrap, weighted bootstrap, m out of n bootstrap, and subsampling, see Section 3.6.2 of van der Vaart and Wellner (1996) for concrete descriptions. Each bootstrap scheme is useful to a specific application. For example, in small samples, we might want to use the weighted bootstrap to gain good accuracy and robustness to "small cells", whereas in large samples, where computational tractability can be an important consideration, we might prefer subsampling.

Previously, the bootstrap validity are only proved for pointwise cases (e.g., Hahn (1995), and Feng et al. (2011)), and the process result was available only for subsampling (see, Chernozhukov and Fernández-Val (2005), and Chernozhukov and Hansen (2006)). Chernozhukov et al. (2012) prove the validity of the general exchangeable bootstrap for estimating the limit law of the entire QR coefficient process (see their Corollary 5.1) and the MD and MQ threshold processes (see their Theorem 5.1(2)). We will not repeat their results in this paper, but only provide the bootstrap procedures in our context.

Let $(\omega_1, \cdots, \omega_n)$ be a vector of nonnegative random variables that satisfy Condition EB in Chernozhukov et al. (2012) or the conditions (3.6.8) of van der Vaart and Wellner (1996). For example, $(\omega_1, \cdots, \omega_n)$ is a multinomial vector with dimension n and probabilities $(1/n, \cdots, 1/n)$ in the empirical bootstrap. The exchangeable bootstrap uses the components of $(\omega_1, \cdots, \omega_n)$ as random sampling weights in the construction of the bootstrap version of the estimators. Thus the bootstrap version of the MD threshold effects is

$$\widehat{\Delta}_D^*(y) = \widehat{F}_1^*(y) - \widehat{F}_2^*(y),$$

where

$$\widehat{F}_\ell^*(y) \equiv \int_{\mathcal{X}_\ell \mathcal{Q}_\ell} \widehat{F}_\ell^*(y|x,q) d\widehat{F}_\ell^*(x,q).$$

The component

$$\widehat{F}_\ell^*(x,q) = (n_\ell^*)^{-1} \sum\nolimits_{i=1}^n \omega_i 1(x_i \leq x, q_i \leq q, x_i \in \mathcal{X}_\ell, q_i \in \mathcal{Q}_\ell), (x',q)' \in \mathcal{X}_\ell \mathcal{Q}_\ell$$

with $n_\ell^* = \sum_{i=1}^n \omega_i 1(x_i \in \mathcal{X}_\ell, q_i \in \mathcal{Q}_\ell)$ is a bootstrap version of $\widehat{F}_\ell(x,q)$. The component

$$\widehat{F}_\ell^*(y|x,q) = \varepsilon + \int_\varepsilon^{1-\varepsilon} 1\left(\mathbf{x}'\widehat{\beta}_\ell^*(\tau) \leq y\right) d\tau, (y,x',q)' \in \mathcal{Y}_\ell \mathcal{X}_\ell \mathcal{Q}_\ell$$

with

$$\widehat{\beta}_1^*(\tau) = \arg\min_{\beta_1} \sum_{i=1}^n \omega_i \rho_\tau(y_i - \mathbf{x}_i'\beta_1) 1(q_i \leq \widehat{\gamma}),$$
$$\widehat{\beta}_2^*(\tau) = \arg\min_{\beta_2} \sum_{i=1}^n \omega_i \rho_\tau(y_i - \mathbf{x}_i'\beta_2) 1(q_i > \widehat{\gamma}),$$

is a bootstrap version of $\widehat{F}_\ell(y|x,q)$.[7] Correspondingly, the MQ threshold effect

$$\widehat{\Delta}_Q^*(\tau) = \widehat{F}_1^{*-1}(\tau) - \widehat{F}_2^{*-1}(\tau).$$

Given $\widehat{\Delta}_D^*(y)$, we can conduct uniform inferences for $\Delta_D(y)$. An asymptotic simultaneous $(1-\alpha)$ confidence band for $\Delta_D(y)$ over $y \in \mathcal{Y}$ is defined by the end-point functions

$$\widehat{\Delta}_D^{\pm}(y) = \widehat{\Delta}_D(y) \pm \widehat{t}_{1-\alpha} \widehat{\Sigma}(y)^{1/2} / \sqrt{n},$$

such that

$$\lim_{n\to\infty} P\left(\Delta_D(y) \in \left[\widehat{\Delta}_D^-(y), \widehat{\Delta}_D^+(y)\right] \text{ for all } y \in \mathcal{Y}\right) = 1-\alpha. \tag{8}$$

Here, $\widehat{\Sigma}(y)$ is a uniformly consistent estimator of $\Sigma(y)$, the asymptotic variance function of $\sqrt{n}\left(\widehat{\Delta}_D(y) - \Delta_D(y)\right)$. In order to achieve the coverage property (8), we set the critical value $\widehat{t}_{1-\alpha}$ as a consistent estimator of the $(1-\alpha)$-quantile of the maximal t-statistic:

$$t = \sup_{y\in\mathcal{Y}} \sqrt{n} \widehat{\Sigma}(y)^{-1/2} \left|\widehat{\Delta}_D(y) - \Delta_D(y)\right|.$$

It remains to obtain $\widehat{\Sigma}(y)$ and $\widehat{t}_{1-\alpha}$. For this purpose, we first get $\widehat{Z}_{D,b}^*(y)$, $b = 1, \cdots, B$, as i.i.d. realization of $\widehat{Z}_D^*(y) = \sqrt{n}\left(\widehat{\Delta}_D^*(y) - \widehat{\Delta}_D(y)\right)$ for $y \in \mathcal{Y}$. Then compute a bootstrap estimate of $\Sigma(y)^{1/2}$ such as the bootstrap interquartile range[8] rescaled with the normal distribution: $\widehat{\Sigma}(y)^{1/2} = (q_{0.75}(y) - q_{0.25}(y))/1.349$ for $y \in \mathcal{Y}$, where $q_p(y)$ is the p-th quantile of $\left\{\widehat{Z}_{D,b}^*(y), b = 1, \cdots, B\right\}$. Finally, $\widehat{t}_{1-\alpha}$ is set as the $(1-\alpha)$ sample quantile of $\left\{\widehat{t}_b, b = 1, \cdots, B\right\}$, where $\widehat{t}_b = \sup_{y\in\mathcal{Y}} \widehat{\Sigma}(y)^{-1/2} \left|\widehat{Z}_{D,b}^*(y)\right|$.

[7] Note here that $\widehat{\gamma}$ is not replaced by its bootstrap counterpart $\widehat{\gamma}^*$ to simplify the bootstrap procedure. Actually, from Yu (2013a), the invalidity of the bootstrap for γ does not affect the bootstrap validity for regular parameters.

[8] Here, the interquartile range rather than the standard deviation is used to avoid technical complexities, see Remark 3.2 of Chernozhukov et al. (2012).

The uniform band for $\Delta_Q(\tau)$ can be obtained similarly by replacing $\widehat{\Delta}_D^*(y)$ and $\widehat{\Delta}_D(y)$ by $\widehat{\Delta}_Q^*(\tau)$ and $\widehat{\Delta}_Q(\tau)$. We can also estimate the critical value for $\mathcal{K}_n$ in the last section by the $(1-\alpha)$ sample quantile of $\left\{\sup_{\tau\in\mathcal{T}}\left\|\widehat{V}(\tau)^{-1/2}\sqrt{n}R(\tau)'\left(\widehat{\beta}_b^*(\tau)-\widehat{\beta}(\tau)\right)\right\|, b=1,\cdots,B\right\}$, where $\widehat{V}(\tau)$ is a uniformly consistent estimator of $V(\tau)$ over $\tau\in\mathcal{T}$, and $\left\{\widehat{\beta}_b^*(\tau)\equiv\left(\widehat{\beta}_{1b}^*(\tau)',\widehat{\beta}_{2b}^*(\tau)'\right)', b=1,\cdots,B\right\}$ are i.i.d. realizations of $\widehat{\beta}^*(\tau)\equiv\left(\widehat{\beta}_1^*(\tau)',\widehat{\beta}_2^*(\tau)'\right)'$. For the construction of the uniform confidence band for a single element of $R(\tau)'\beta(\tau)-r(\tau)$, $\widehat{V}(\tau)^{-1/2}$ can be substituted by the corresponding rescaled bootstrap interquartile range.

5 Specification Testing

To estimate γ or the MD and MQ threshold processes in the previous sections, we must first guarantee that there are CQ threshold effects. For such specification testing, it is more convenient to reparametrize the model as

$$y=\mathbf{x}'\beta_0(\tau)+\mathbf{x}'\delta(\tau)1(q\leq\gamma_0)+e_\tau,\ Q_\tau[e_\tau|\mathbf{x}]=0,$$

where the true threshold point γ_0 is unknown. The null hypothesis is that there are not CQ threshold effects, or

$$H_0:\delta(\tau)=0 \text{ for all } \tau\in\mathcal{T},$$

and correspondingly, the alternative is

$$H_1:\delta(\tau)\neq 0 \text{ for some } \tau\in\mathcal{T},$$

and the local alternative is

$$H_1^c:\delta(\tau)=n^{-1/2}c(\tau) \text{ for some } \tau\in\mathcal{T}.$$

To facilitate the development of our asymptotic results, we impose the following additional assumptions.

Assumption T:

1. The minimum eigenvalues of $J(\tau)\equiv E[f_{e_\tau|x,q}(0|x_i,q_i)\mathbf{x}_i\mathbf{x}_i']$ and $J(\gamma,\tau)\equiv E[f_{e_\tau|x,q}(0|x_i,q_i)\mathbf{x}_i(\gamma)\mathbf{x}_i(\gamma)']$ are uniformly bounded away from zero uniformly over $(\tau,\gamma)\in\mathcal{T}\times\Gamma$, where $\mathbf{x}_i(\gamma)=\mathbf{x}_i1(q_i\leq\gamma)$.
2. $c(\tau)$ is uniformly bounded over $\tau\in\mathcal{T}$.
3. $f(e_\tau|x,q)$ is bounded and uniformly continuous in e_τ uniformly over $(\tau,x',q)'\in\mathcal{T}\times\mathbb{X}\times\mathbb{Q}$.

From Assumption D3, $c(\tau)$ should be uniformly continuous, but the proof does not require this assumption.

5.1 Test Statistics

A straightforward test is the Wald-type test which is based on the estimate of $\delta(\tau)$. The test statistics are functionals of

$$\widehat{W}_n(\gamma,\tau)=\left(\widehat{J}_1(\gamma,\tau)^{-1}\widehat{\Sigma}_1(\gamma,\tau)\widehat{J}_1(\gamma,\tau)^{-1}+\widehat{J}_2(\gamma,\tau)^{-1}\widehat{\Sigma}_2(\gamma,\tau)\widehat{J}_2(\gamma,\tau)^{-1}\right)^{-1/2}\sqrt{n}\widehat{\delta}(\gamma,\tau),$$

where $\widehat{\delta}(\gamma,\tau) \equiv \widehat{\beta}_1(\gamma,\tau) - \widehat{\beta}_2(\gamma,\tau)$, $\widehat{\beta}_1(\gamma,\tau)$ is the τth QR estimator using the data with $q_i \leq \gamma$, $\widehat{J}_1(\gamma,\tau) = n^{-1}\sum_{i=1}^n K_{h_1}(\widehat{e}_{1\tau i})\mathbf{x}_i\mathbf{x}_i' 1(q_i \leq \gamma)$ with $\widehat{e}_{1\tau i} = y_i - \mathbf{x}_i'\widehat{\beta}_1(\gamma,\tau)$, $\widehat{\Sigma}_1(\gamma,\tau) = \tau(1-\tau)n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i' 1(q_i \leq \gamma)$, and $\widehat{\beta}_2(\gamma,\tau)$, $\widehat{J}_2(\gamma,\tau)$ and $\widehat{\Sigma}_2(\gamma,\tau)$ are similarly defined but using the data with $q_i > \gamma$. This test is hard to apply in practice due to two reasons. First, many quantile regressions (indexed by (γ,τ)) should be conducted, which is quite time-consuming. Second, the critical values are hard (although not impossible) to obtain even by the simulation method in the next subsection.

In this paper, we suggest the score-type test which constructs the test statistics under the null rather than under the alternative as in the Wald-type test. This type of test is based on the subgradient of $\rho_\tau(\cdot)$, just like the CUSUM test which is based on the gradient of the objective function of least squares (see, inter alia, Ploberger and Krämer (1992) and Bai (1996) in the structural change testing). Yu (2009) uses similar ideas in the specification testing of threshold regression with endogeneity. The test statistics are functionals of

$$\widehat{T}_n(\gamma,\tau) = \left[\tau(1-\tau)\cdot n^{-1}\sum_{i=1}^n \left(\mathbf{x}_i(\gamma) - \widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right)\left(\mathbf{x}_i(\gamma) - \widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right)'\right]^{-1/2}$$
$$\cdot n^{-1/2}\sum_{i=1}^n \left[\mathbf{x}_i(\gamma) - \widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right]\varphi_\tau(\widehat{e}_{\tau i}),$$

where

$$\widehat{J}(\gamma,\tau) = \frac{1}{n}\sum_{i=1}^n K_h(\widehat{e}_{\tau i})\mathbf{x}_i\mathbf{x}_i' 1(q_i \leq \gamma), \widehat{J}(\tau) = \frac{1}{n}\sum_{i=1}^n K_h(\widehat{e}_{\tau i})\mathbf{x}_i\mathbf{x}_i'$$

are similarly defined as in (7), $\varphi_\tau(u) = 1(u<0) - \tau$, $\widehat{e}_{\tau i} = y_i - \mathbf{x}_i'\widehat{\beta}(\tau)$, and $\widehat{\beta}(\tau)$ is the QR estimator of y_i on $\mathbf{x}_i$ (so the null hypothesis is imposed). Different from $\widehat{W}_n(\gamma,\tau)$, we need only run one quantile regression for each τ to construct $\widehat{T}_n$. Note here that although $\widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}n^{-1/2}\sum_{i=1}^n \mathbf{x}_i\varphi_\tau(\widehat{e}_{\tau i}) = o_p(1)$, $\mathbf{x}_i(\gamma)$ is recentered by $\widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i$. This is because the effect of $\widehat{\beta}(\tau)$ will not disappear asymptotically so the asymptotic distribution of $n^{-1/2}\sum_{i=1}^n \mathbf{x}_i(\gamma)\varphi_\tau(\widehat{e}_{\tau i})$ is different from $n^{-1/2}\sum_{i=1}^n \mathbf{x}_i(\gamma)\varphi_\tau(e_{\tau i}) = n^{-1/2}\sum_{i=1}^n \mathbf{x}_i(\gamma)\varphi_\tau(y_i - \mathbf{x}_i'\beta_0(\tau))$ under H_0. In other words, the Durbin problem reappears in this context. Recentering is to offset the effect of $\widehat{\beta}(\tau)$.

Given $\widehat{T}_n$, we usually consider two test statistics. The first is the Kolmogorov-Smirnov sup-type statistic

$$\widehat{K}_n = \sup_{\tau\in\mathcal{T}}\sup_{\gamma\in\Gamma}\left\|\widehat{T}_n(\gamma,\tau)\right\|,$$

and the second is the Cramér–von Mises average-type statistic

$$\widehat{C}_n = \int_{\mathcal{T}}\int_{\Gamma}\left\|\widehat{T}_n(\gamma,\tau)\right\|\omega_1(\gamma)\omega_2(\tau)d\gamma d\tau,$$

where $\omega_1(\gamma)$ and $\omega_2(\tau)$ in $\widehat{C}_n$ are known positive weight functions with $\int_\Gamma \omega_1(\gamma)d\gamma = 1$ and $\int_{\mathcal{T}}\omega_2(\tau)d\tau = 1$. For example, $\omega_2(\tau) = 1/|\mathcal{T}|$ with $|\mathcal{T}|$ being the length of $\mathcal{T}$; if we have some information on the quantile indices where threshold effects are most likely to happen, we can impose larger weights on the neighborhoods of such indices.

The choice of the norm $\|\cdot\|$ is also an issue. Euclidean norm $\|\cdot\|_2$ is obviously natural, but has the possibly undesirable effect of accentuating extreme behavior in a few coordinates. Instead, we will employ the ℓ_1 norm in the simulations and the empirical application below. Also, $\|\cdot\|_\infty$ is used for $\widehat{K}_n$ in the structural change test of Qu (2008). Define $g_n = g(\widehat{T}_n)$, where g is the functional defined in $\widehat{K}_n$ or $\widehat{C}_n$. The following theorem

states the weak limits of g_n under H_1^c.

Theorem 5 *Suppose the same assumptions on $K(\cdot)$, h, and $f(y,x,q)$ as in Corollary 4 are satisfied; then under Assumptions T, D1 and D4,*

$$g_n \xrightarrow{d} g^c = g(T^c),$$

where

$$T^c(\gamma,\tau) = H(\gamma,\tau)^{-1/2}\left\{S(\gamma,\tau) - \left[J(\gamma\wedge\gamma_0,\tau) - J(\gamma,\tau)J(\tau)^{-1}J(\gamma_0,\tau)\right]c(\tau)\right\},$$

with $S(\gamma,\tau)$ being a zero-mean Gaussian process with the covariance kernel

$$H((\gamma_1,\tau_1),(\gamma_2,\tau_2)) = (\tau_1\wedge\tau_2 - \tau_1\tau_2)\,E\left[\left(\mathbf{x}_i(\gamma_1) - J(\gamma_1,\tau_1)J(\tau_1)^{-1}\mathbf{x}_i\right)\left(\mathbf{x}_i(\gamma_2) - J(\gamma_2,\tau_2)J(\tau_2)^{-1}\mathbf{x}_i\right)'\right],$$

and $H(\gamma,\tau) = H((\gamma,\tau),(\gamma,\tau))$.

To understand $S(\gamma,\tau)$, consider a simple case where $\mathbf{x} = (1,x')'$, q follows the uniform distribution on $[0,1]$ and is independent of $(x',e_\tau)'$. In this case,

$$H((\gamma_1,\tau_1),(\gamma_2,\tau_2)) = (\tau_1\wedge\tau_2 - \tau_1\tau_2)(\gamma_1\wedge\gamma_2 - \gamma_1\gamma_2)\,E[\mathbf{x}\mathbf{x}'];$$

in other words, $E[\mathbf{x}\mathbf{x}']^{-1/2}S(\gamma,\tau)$ is the standard p-dimensional Brownian Pillow (or tucked Brownian Sheet). Now, the local power is generated by $\left[J(\gamma\wedge\gamma_0,\tau) - J(\gamma,\tau)J(\tau)^{-1}J(\gamma_0,\tau)\right]c(\tau) = (\gamma\wedge\gamma_0 - \gamma\gamma_0)\,E[\mathbf{x}\mathbf{x}']\,c(\tau)$. These results are similar to those in the structural change testing; see, e.g., Proposition 2 and Corollary 1 of Qu (2008). Of course, the construction of $\widehat{T}_n$ can be greatly simplified in this simple case, e.g.,

$$\widehat{T}_n(\gamma,\tau) = \left[n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i'\right]^{-1/2}\cdot n^{-1/2}\sum_{i=1}^n \mathbf{x}_i\varphi_\gamma(q_i-\gamma)\,\varphi_\tau(\widehat{e}_{\tau i}) \tag{9}$$

will converge to the standard p-dimensional Brownian Pillow under H_0.

5.2 Simulating the Critical Values

The asymptotic distribution of g_n is not pivotal. Following Hansen (1996), we obtain the critical values by simulating $\widehat{T}_n(\gamma,\tau)$. More specifically, let $\{\xi_i^*\}_{i=1}^n$ be i.i.d. $N(0,1)$ random variables, and set

$$\begin{aligned}\widehat{T}_n^*(\gamma,\tau) = &\left[n^{-1}\sum_{i=1}^n \varphi_\tau(\widehat{e}_{\tau i})^2\left(\mathbf{x}_i(\gamma) - \widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right)\left(\mathbf{x}_i(\gamma) - \widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right)'\right]^{-1/2} \\ &n^{-1/2}\sum_{i=1}^n\left[\mathbf{x}_i(\gamma) - \widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right]\varphi_\tau(\widehat{e}_{\tau i})\,\xi_i^*.\end{aligned} \tag{10}$$

Here, $\tau(1-\tau)$ in $\widehat{T}_n(\gamma,\tau)$ is replaced by $\varphi_\tau(\widehat{e}_{\tau i})^2$. This is because under H_1, $\widehat{\beta}(\tau)$ is only an approximate of the true conditional quantile function (see, e.g., Angrist et al. (2006)), and the asymptotic variance of $n^{-1/2}\sum_{i=1}^n\left[\mathbf{x}_i(\gamma) - \widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right]\varphi_\tau(\widehat{e}_{\tau i})$ is not $H(\gamma,\tau)$ any more. Since the true data generating process (DGP) is unknown, we must use this robust asymptotic variance estimator to mimic the behavior under H_0.[9] Given this observation, to make sure the conditional distribution of $\widehat{T}_n^*(\gamma,\tau)$ given the original data is

[9] Of course, under H_1^c, $\tau(1-\tau)$ can be used in $\widehat{T}_n^*$.

close to the distribution of $\widehat{T}_n(\gamma,\tau)$ under H_0, $\widehat{T}_n(\gamma,\tau)$ can be replaced by

$$\widetilde{T}_n(\gamma,\tau) = \left[n^{-1}\sum_{i=1}^{n} \varphi_\tau\left(\widehat{e}_{\tau i}\right)^2 \left(\mathbf{x}_i(\gamma) - \widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right)\left(\mathbf{x}_i(\gamma) - \widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right)' \right]^{-1/2}$$
$$n^{-1/2}\sum_{i=1}^{n}\left[\mathbf{x}_i(\gamma) - \widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right]\varphi_\tau\left(\widehat{e}_{\tau i}\right),$$

and the corresponding K_n and C_n are denoted as $\widetilde{K}_n$ and $\widetilde{C}_n$, respectively. The tests based on $\widetilde{T}_n$ may have more precise sizes but lose some powers.

Our test is to reject H_0 if g_n is greater than the $(1-\alpha)$th conditional quantile of $g(\widehat{T}_n^*)$. Equivalently, the p-value transformation can be employed. Define $p_n^* = 1 - F_n^*(g_n)$, and $p_n = 1 - F_0(g_n)$, where F_n^* is the conditional distribution of $g(\widehat{T}_n^*)$ given the original data, and F_0 is the asymptotic distribution of $g(\widehat{T}_n)$ under the null. Our test is to reject H_0 if $p_n^* \leq \alpha$. The following theorem states the validity of the above procedure.

Theorem 6 *Under the Assumptions of Theorem 5, $p_n^* = p_n + o_p(1)$ under both H_0 and H_1^c. Hence $p_n^* \xrightarrow{d} p^c = 1 - F_0(g^c)$ under H_1^c, and $p_n^* \xrightarrow{d} U$, the uniform distribution on $[0,1]$, under H_0.*

By the stochastic equicontinuity of the $\widehat{T}_n(\gamma,\tau)$ process, we can replace $\mathcal{T}$ and Γ by finite grids with the distance between adjacent grid points going to zero as $n \to \infty$.[10] Also, the conditional distribution can be approximated by standard simulation techniques. More specifically, the following procedure is used.

Step 1: generate $\left\{\xi_{ij}^*\right\}_{i=1}^n$ be i.i.d. $N(0,1)$ random variables.
Step 2: set $\widehat{T}_n^{j*}(\gamma_l,\tau_t)$ as in (10), where $\{\gamma_l\}_{l=1}^L$ and $\{\tau_t\}_{t=1}^T$ are grid approximation of Γ and $\mathcal{T}$. Note here that the same $\left\{\xi_{ij}^*\right\}_{i=1}^n$ are used for all (γ_l,τ_t), $l = 1,\cdots,L$, $t = 1,\cdots,T$.
Step 3: set $g_n^{j*} = g(\widehat{T}_n^{j*})$.
Step 4: repeat Step 1-3 J times to generate $\left\{g_n^{j*}\right\}_{j=1}^J$.
Step 5: if $p_n^{J*} = J^{-1}\sum_{j=1}^J 1(g_n^{j*} \geq g_n) \leq \alpha$, we reject H_0; otherwise, accept H_0.

6 Monte Carlo Experiments

In this section, we conduct some Monte Carlo experiments to check the performance of the estimators and tests in the previous sections. Given that the SEB procedure of Yu (2008) and simulating the critical values of the score-type tests in Section 5 are very time-consuming, we will consider the following simple DGPs to save simulation time.

$$y = \begin{cases} (1\ x)\,\beta_1 + \sigma_1 e, & q \leq \gamma; \\ (1\ x)\,\beta_2 + \sigma_2 e, & q > \gamma; \end{cases}$$

where $x \sim N(0,1)$, $q \sim U[0,1]$, $e \sim N(0,1)$ or the double exponential distribution with scale $1/\sqrt{2}$ (which has variance 1 and is denoted as $DExp(1/\sqrt{2})$), and x, q and e are independent of each other. The double exponential distribution of e is also used in the simulation study of Bai (1995), corresponding to the heavy-tailed error case. $\gamma_0 = 0.5$, $\beta_1^{0\prime} = (0\ 0)$, $\beta_2^{0\prime} = n^{-1/2}c \cdot (1/\sqrt{2}, 1/\sqrt{2})$ for some positive numbers of c, $n = 200$, and the number of repetitions is set as 500. We consider two setups for σ_ℓ^0. In the first setup, $\sigma_{10} = \sigma_{20} = 1$, and in the second setup, $\sigma_{10} = 1$ and $\sigma_{20} = 2$. The first setup only considers the threshold effect in conditional mean (or median), while the second setup also covers the threshold effect in variance

[10] A natural choice of the grids for Γ is the q_i's in Γ, and for $\mathcal{T}$ is the breakpoints in $\mathcal{T}$ whose number is at most $O_p(n \ln n)$ from Portnoy (1991).

(and other quantiles). We label the case with $\sigma_{10} = \sigma_{20} = 1$ and $e \sim N(0,1)$ as scenario 1, $\sigma_{10} = \sigma_{20} = 1$ and $e \sim DExp(1/\sqrt{2})$ as scenario 2, $\sigma_{10} = 1, \sigma_{20} = 2$ and $e \sim N(0,1)$ as scenario 3, $\sigma_{10} = 1, \sigma_{20} = 2$ and $e \sim DExp(1/\sqrt{2})$ as scenario 4, respectively. $\Gamma = [q_{(0.2n)}, q_{(0.8n)}]$, where $q_{(i)}$ is the ith order statistic of $\{q_i\}_{i=1}^n$. $\mathcal{T} = [0.1, 0.9]$.

6.1 Specification Testing

The specification testing is extremely time-consuming when a fine approximation of $\mathcal{T}$ is used. To save simulation time, we only check the performance of $\widehat{K}_n$ with a rough approximation of $\mathcal{T}$ being used. Specifically, we use 11 approximation points $\{\tau_t\}_{t=1}^T$ which are evenly distributed on $\mathcal{T}$. Such a rough approximation is not suitable for $\widehat{C}_n$ or $\widetilde{C}_n$. Also, given the special structure of the joint distribution of $(e, x, q)'$, we use the simple form (9) of $\widehat{T}_n$ to further save simulation time. For the tests based on the LSE, we only consider the score-type tests for comparison; see Yu (2009) for descriptions on these tests. In simulating the critical values, we let $J = 400$. The size and power is evaluated at the 5% nominal level. Totally, we consider only three test statistics: $\widehat{K}_n$ with $\mathcal{T} = \{\tau_t\}_{t=1}^T$ or based on the LADE only, and the sup-form of the score-type test based on the LSE.

We report the simulation results of specification testing in Figure 1. From Figure 1, a few results of interest are summarized as follows. First, in scenario 1, the score-test based on the LSE performs best, and in scenario 2, $\widehat{K}_n$ based on the LADE works best. This is understandable given that the threshold effect appears only in conditional mean in scenario 1 and only in conditional median in scenario 2. Nevertheless, the performance of $\widehat{K}_n$ with $\mathcal{T} = \{\tau_t\}_{t=1}^T$ is close to the best case in scenario 1 and is identical to the best case in scenario 2. Second, in scenario 3 and 4, $\widehat{K}_n$ with $\mathcal{T} = \{\tau_t\}_{t=1}^T$ performs much better than the other two tests. Especially, the tests based on the LSE or the LADE do not have any power when $c = 0$; however, the tests based on $\widehat{K}_n$ have significant powers even when $c = 0$. Third, in scenario 3, the test based on the LSE performs better than the test based on the LADE and in scenario 4, the converse conclusion can be drawn. These results obviously parallel those in scenario 1 and 2. In summary, $\widehat{K}_n$ based on multiple quantiles performs stably and among the best in all kinds of scenarios, no matter whether the error has a heavy tail, or whether there is a threshold effect in variance.

6.2 Estimation

In the estimation, our first goal is to compare the efficiency of the IQTRE with the LSE, LADE and MLE and also the SEBE started from the IQTRE with that started from the LSE and LADE. Note that in all setups, $E[e_\ell | x, q] = 0$, so the LSE can be applied. Also, since $\text{Med}(e_\ell | x, q) = 0$, the LADE can only identify γ from the threshold effect in β_ℓ (rather than in σ_ℓ), and is comparable to the LSE. Given that e_ℓ has the maximum density at its median, the LADE should be the most efficient among all QRE's, so we only report the results for the LADE. In the IQTRE, the same $\{\tau_t\}_{t=1}^T$ as in the specification testing are used. As to the MLE, the algorithm can be found in Section 3.2 of Yu (2012). Our second goal is to compare the coverage and length of various CIs, including the LR-CIs in Section 3.3 based on the LSE, LADE and IQTRE, and the NPI started from the LSE, LADE and IQTRE. In constructing $LR_n(\gamma)$, we use the kernel smoother to estimate $f_e(F_e^{-1}(\tau_t))$ in π_{1Tn}. The bandwidth h is set to be $n^{-1/3} z_\alpha^{2/3} \left[1.5\phi^2\left(\Phi^{-1}(\tau)\right) / \left(\left(2\Phi^{-1}(\tau)\right)^2 + 1\right)\right]^{1/3}$ as suggested in Hall and Sheather (1988), where ϕ and Φ stand for the standard normal density and distribution function and z_α satisfies $\Phi(z_\alpha) = 1 - \alpha/2$. We let $c = 20$ and the resulting $\beta_2^0 = (1, 1)'$. From the specification testing in the last subsection, this value of c should be out of the contiguous neighborhood of the null.

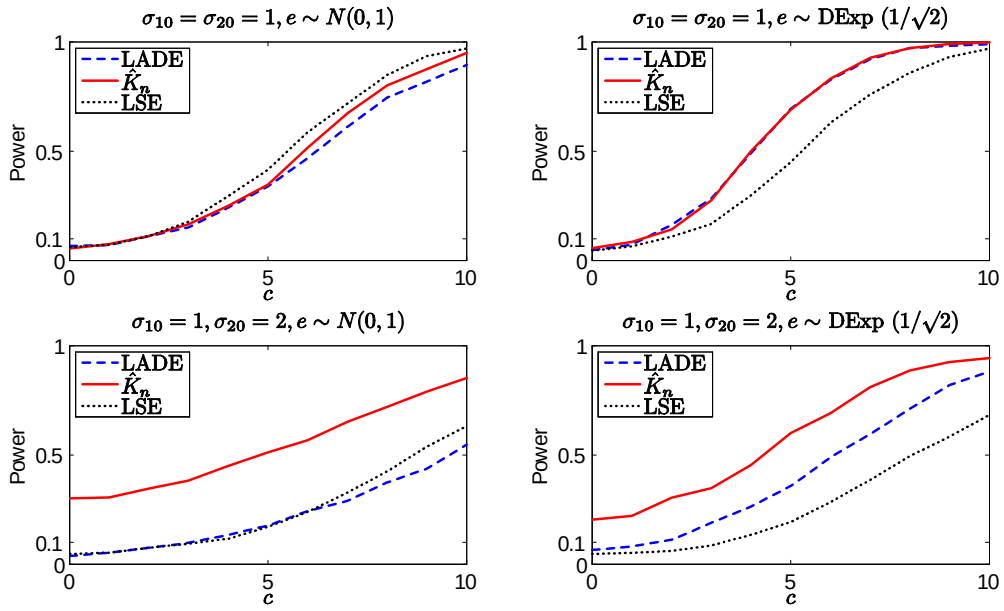

Figure 1: Power Comparison with Different Specifications of $\sigma_{\ell 0}$ and Error Distributions

The performance of various estimators are summarized in Table 1. From Table 1, the following conclusions can be drawn. First, among the LSE, LADE and IQTRE, the LSE performs best in scenario 1 and the LADE performs best in scenario 2. They perform even better than the MLE since they are equivalent to the MLE with the restriction $\sigma_{10} = \sigma_{20}$ imposed. The performance of the IQTRE is the close to that of the MLE (and the best semiparametric estimator) in these two scenarios. Second, the IQTRE performs best among the LSE, LADE and IQTRE in scenario 3 and 4, so the threshold effects at other quantiles indeed provide information for γ. On the other hand, by comparing the IQTRE and the MLE, we can see that the threshold effects at only finite quantiles cannot cover the whole CQ threshold effects. Also, as expected from the specification testing, the LSE performs better than the LADE in scenario 3 and the LADE performs better than the LSE in scenario 4. This is understandable since the LADE is more robust to the heavy-tailed error than the LSE; see Section 3.1. Third, the SEBE started from the best-performed estimator performs the best. This verifies our expectation in Section 3.3: the starting value of the SEBE indeed affects the efficiency of the SEBE in finite samples. In summary, it is safe to claim that the IQTRE performs stably well in all scenarios.

The performance of various CIs are summarized in Table 2. Four main conclusions from Table 2 are as follows. First, the LR-CIs suffer from the overcoverage problem and the NPIs suffer from the undercoverage problem. On the other hand, the NPI is much shorter than the corresponding LR-CI. Second, among the three LR-CIs, the CI based on the IQTRE performs best if taking both the coverage and length into consideration, which matches the efficiency results in Table 1. Third, although all NPIs have the undercoverage problem, the NPI based on the IQTRE suffers the least. Also, the NPI based on the IQTRE is shorter than the other two NPIs, especially in scenario 3 and 4. Fourth, the NPI in the scenario with a heavy-tailed error generally has a worse coverage than that in the scenario with a light-tailed error. In summary, for both the LR-CI and the NPI, the CIs based on the IQTRE perform the best.

$\sigma_{\ell 0}$	$\sigma_{10}=\sigma_{20}=1$				$\sigma_{10}=1$ and $\sigma_{20}=2$			
$f_e(\cdot)$	$N(0,1)$		DExp$(1/\sqrt{2})$		$N(0,1)$		DExp$(1/\sqrt{2})$	
Risk$(\times 10^{-2})$	RMSE	MAD	RMSE	MAD	RMSE	MAD	RMSE	MAD
LSE	2.00	1.21	1.77	1.06	4.93	2.77	4.73	2.58
LADE	2.19	1.34	1.43	0.90	5.54	3.10	2.91	1.76
IQTRE	2.06	1.24	1.54	0.96	3.70	1.98	2.74	1.63
SEBE_LSE	1.71	1.16	1.34	0.89	3.62	1.87	3.45	1.84
	1.77	1.11	1.34	0.86	3.84	1.91	3.62	1.84
SEBE_LADE	1.78	1.18	1.26	0.84	4.26	2.09	2.30	1.38
	1.83	1.15	1.30	0.81	4.39	2.11	2.45	1.40
SEBE_IQTRE	1.72	1.16	1.29	0.86	2.97	1.54	2.24	1.36
	1.77	1.12	1.33	0.83	3.04	1.55	2.35	1.36
MLE	2.03	1.23	1.50	0.94	1.99	1.19	1.92	1.18

Table 1: Estimator Performance for γ (Based on 500 Repetitions)

Note: MAD: mean absolute deviation, SEBE_$\cdot$: SEBE started from $\cdot$ (the upper value is the posterior mean, and the lowe value is the posterior median)

$\sigma_{\ell 0}$	$\sigma_{10}=\sigma_{20}=1$				$\sigma_{10}=1$ and $\sigma_{20}=2$			
$f_e(\cdot)$	$N(0,1)$		DExp$(1/\sqrt{2})$		$N(0,1)$		DExp$(1/\sqrt{2})$	
Coverage and Length $(\times 10^{-2})$	Cov	Leng	Cov	Leng	Cov	Leng	Cov	Leng
LR_LSE	0.986	5.95	0.970	5.78	0.972	17.02	0.974	17.29
LR_LADE	0.996	9.86	0.992	5.85	0.976	21.01	0.978	11.57
LR_IQTRE	0.994	6.86	0.976	4.98	0.984	10.38	0.982	8.59
NPI_LSE	0.936	4.93	0.900	3.93	0.884	5.45	0.860	5.32
NPI_LADE	0.934	4.94	0.918	3.82	0.872	5.65	0.884	5.00
NPI_IQTRE	0.938	4.96	0.918	3.82	0.896	4.92	0.884	4.80

Table 2: Comparison of Inference Methods: Coverage and Average Length of Nominal 95% Confidence Intervals for γ (Based on 500 Repetitions)

Note: LR_$\cdot$: equal-tailed LR-CI based on $\cdot$, NPI_$\cdot$: NPI started from $\cdot$

7 Application

In this section, we apply the estimation and testing procedures in Section 3, 4 and 5 to the growth data used in Durlauf and Johnson (1995) and reanalyzed in Hansen (2000) and Yu (2008). A similar dataset is used in Koenker and Machado (1999), but no threshold effects are considered there. The growth theory with multiple equilibria motivates the following threshold regression model:

$$\begin{aligned}&\ln\left(\tfrac{Y}{L}\right)_{i,1985}-\ln\left(\tfrac{Y}{L}\right)_{i,1960}\\ &=\begin{cases}\beta_{10}+\beta_{11}\ln\left(\tfrac{Y}{L}\right)_{i,1960}+\beta_{12}\ln\left(\tfrac{I}{Y}\right)_i+\beta_{13}\ln\left(n_i+g+\delta\right)+\beta_{14}\ln S_i+\sigma_1 e_i, & \text{if } \left(\tfrac{Y}{L}\right)_{i,1960}\le\gamma;\\ \beta_{20}+\beta_{21}\ln\left(\tfrac{Y}{L}\right)_{i,1960}+\beta_{22}\ln\left(\tfrac{I}{Y}\right)_i+\beta_{23}\ln\left(n_i+g+\delta\right)+\beta_{24}\ln S_i+\sigma_2 e_i, & \text{if } \left(\tfrac{Y}{L}\right)_{i,1960}>\gamma.\end{cases}\end{aligned}$$

For each country i, $\left(\frac{Y}{L}\right)_{i,t}$ is the real GDP per member of the population aged 15-64 in year t, $\left(\frac{I}{Y}\right)_i$ is the investment to GDP ratio, n_i is the growth rate of the working-age population, and S_i is the fraction of working-age population enrolled in secondary schools. The variables not indexed by t are annual averages

over the period 1960-1985. Following Durlauf and Johnson (1995), we set $g+\delta = 0.05$. The data are assumed to be i.i.d. sampled. This assumption is approximately true, since there are not many interactions, such as trade, international capital flows, etc., between any two countries during this period. The objective is to check whether the growth depends on the starting point.

The LSE, LADE, IQTRE and the corresponding LR-CIs, SEBEs and NPIs for γ are summarized in Table 3, where $\Gamma = [q_{(0.1n)}, q_{(0.9n)}]$, $\mathcal{T} = [0.2, 0.8]$, and $T = 11$.[11] It is quite surprising that the three estimators coincide for this data set. This indicates that the threshold point is indeed a lower (18.85%) percentile of q, below which are mostly poor African countries. Nevertheless, the IQTRE is the most efficient since the length of the corresponding LR-CI is the shortest. Anyway, the three LR-CIs are not informative given that they are all too wide. The length of the NPI improves a lot relative to the LR-CI. As expected, the NPI based on the IQTRE is narrower than that based on the LSE. From the simulation study in Section 6.2, the NPI may suffer from the undercoverage problem and the LR-CI may suffer from the overcoverage problem, so a better CI should be in-between.

		LSE	LADE	IQTRE
Original Estimators		871		
LR-CI		[594,1842)	[755,1842)	[755,1623)
Length of LR-CI		1248	1087	868
Ratio of Countries Covered by LR-CI		40/96	37/96	33/96
SEBE	Posterior Mean	828.8	843.8	
	Posterior Median	831.7	863.7	
NPI		[756.4, 877.9]	[778.9, 878.3]	
Length of NPIs		121.5	99.4	
Ratio of Countries Covered by NPIs		6/96	4/96	

Table 3: Comparison of Estimators and Inference Methods for γ

Figure 2 shows the quantile processes in the two regimes generated by the threshold point 871.[12] Since the sample size is relatively small, the uniform confidence bands are not informative, so are not drawn on. The last graph of Figure 2 shows the MQ threshold effects. For comparison, the corresponding least squares estimators are also shown in Figure 2. The quantile processes in Figure 2 reveal more information about the growth patterns in the two regimes than the LSE. For example, schooling does not have any effect on growth for poor countries below the median, while it has significant effects for all rich countries. For another example, the effect of the starting point of growth is quite uniform among all quantiles for rich countries, while for poor countries, its effects at all quantiles are below the mean effect, with the minimum reached around the median. Another observation is that the error term e is not generated by $(\mathbf{x}'\lambda)\,\epsilon$ with ϵ being i.i.d. sampled; otherwise, each component of $\widehat{\beta}_\ell$ should be a location-scale transformation of any other component. But this is obviously not the case from Figure 2, which means that there are complicated heteroskedasticities in the error term.[13] At last, the MQ threshold effects are negative over all $\tau \in \mathcal{T}$, just as expected. The MQ threshold effect is larger for a larger quantile index τ, which indicates that the marginal distribution of $(x', q)'$ is very important in the MQ threshold effect evaluation. From the estimation of $\beta_\ell(\cdot)$, the threshold effects in the conditional distribution concentrates on the lower especially medium τ. However, the marginal distributions of $(x', q)'$ are very different, e.g., the mean of $\mathbf{x}$ in the right regime is much larger than that in the left regime.

[11] For $T = 3$ to 22, the IQTRE is the same as the LADE.

[12] The quantile processes based on the SEBEs are qualitatively similar.

[13] Durlauf and Johnson (1995) also observe the heteroskedasticity of the error term, so our results convince and refine their results.

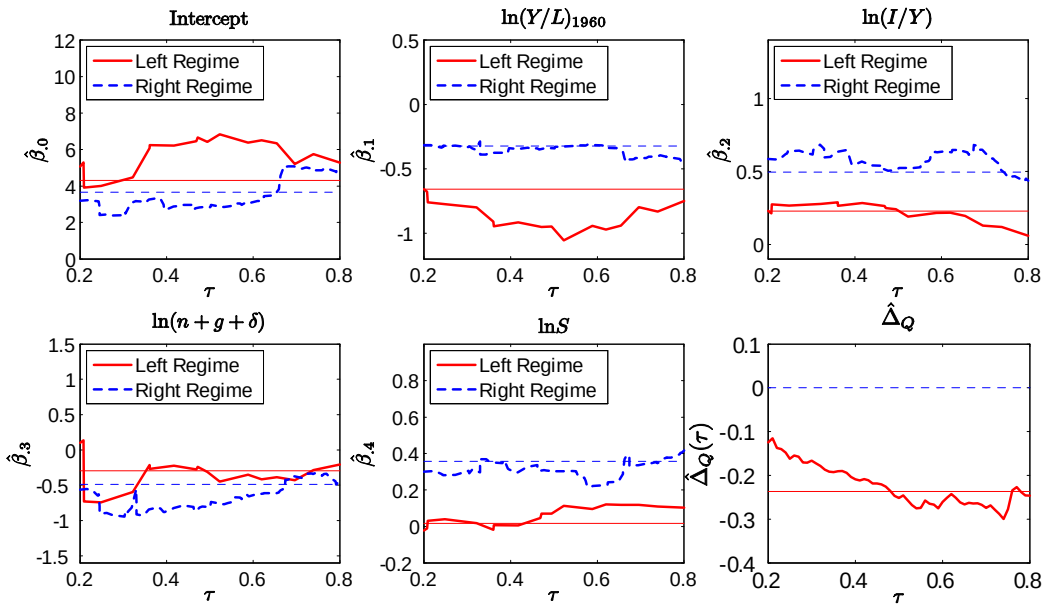

Figure 2: $\widehat{\beta}_\ell(\cdot)$ and $\widehat{\Delta}_Q(\cdot)$

Next, we conduct the specification testing on CQ threshold effects. the corresponding p_n^{J*} for $\widehat{K}_n$ is 0.09, so the null hypothesis is rejected by $\widehat{K}_n$ at 10% significance level, where $J = 500$. The corresponding p-values for the sup-score test based on the LSE is 0.262, so the null hypothesis cannot be rejected. To explore further which quantile contributes mostly to the quantile threshold effects, we calculate $\widehat{K}_n$ for $\mathcal{T} = \{\tau\}$ and the corresponding p_n^* values, and then graph these p-values against τ in Figure 3. From Figure 3, most threshold effects concentrate on the lower especially medium indices of $\mathcal{T}$, just as predicted by Figure 2. Also, Figure 3 indicates that the specification test based solely on the LADE cannot reject the null hypothesis. From these testing results, we can conclude that the test based on a range of quantile threshold effects indeed has a larger power than that based solely on the LSE or the LADE.

8 Conclusion

We have considered the estimation and specification testing in quantile threshold regression. First, we propose the IQTRE for the threshold point, and derive its asymptotic distribution in two asymptotic frameworks. This estimator is more efficient than the existing estimators based on a single characteristic of the conditional distribution of the response variable, such as the LSE and LADE, and is comparable to the MLE, so can serve as a better starting point in the adaptive estimation of the threshold point. Second, we estimate two new threshold processes: the marginal distributional threshold process and the marginal quantile threshold process, and provide both the asymptotic and resampling inference methods for these processes. Third, we put forward a new score-type test in the specification testing of quantile threshold regression. This type of test is more powerful than the tests based solely on the LSE or the LADE. Comparing with the usual Wald-type test, it is computationally less intensive, and its critical values are easier to obtain by the simulation method.

Possible extensions of the analyses in the paper can be along the following directions. First, the insights in this paper are ready to extend to time series, repeated cross-sections, and panels. Second, we can extend the

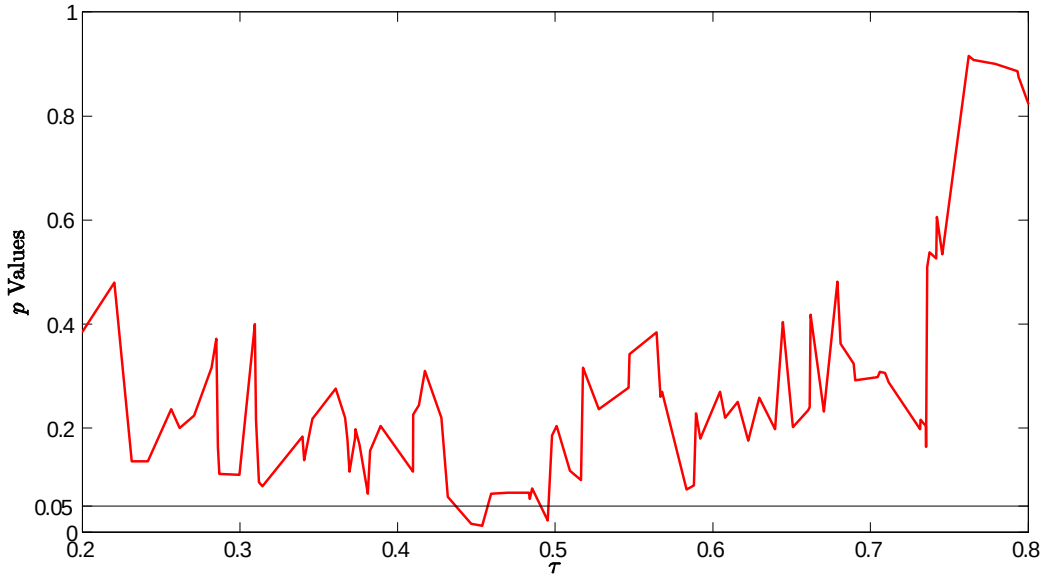

Figure 3: p-Values of $\widehat{K}_n$ for each τ

one-regime analysis to the multiple-regime case. Especially, the procedure to determine the number of breaks in Section 6 of Oka and Qu (2011) can be extended to determine the number of threshold points. Third, the analysis in this paper is based on quantile regression, and an alternative way is based on distribution regression; see Section 3.2 of Chernozhukov et al. (2012) for an introduction. Fourth, we assume the model is correctly specified in this paper. Actually, the application in Galvao et al. (2011) indicates that there may exist misspecification in the setup (2) since the QRE of the threshold point depends on the quantile index. Extension to incorporate misspecification can be done along the line of Yu (2013b).

References

Andrews, D., 1993, Tests for Parameter Instability and Structural Change with Unknown Change Point, *Econometrica*, 61, 821-856. A Corrigendum in *Econometrica*, 71, 395-397.

Angrist, J. et al., 2006, Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure, *Econometrica*, 74, 539-563.

Bai, J., 1995, Least Absolute Deviation Estimation of a Shift, *Econometric Theory*, 11, 403-436.

Bai, J., 1996, Testing for Parameter Constancy in Linear Regressions: An Empirical Distribution Function Approach, *Econometrica*, 64, 597-622.

Bai, J., 1997, Estimation of a Change Point in Multiple Regression Models, *Review of Economics and Statistics*, 79, 551-563.

Bai, J., 1998, Estimation of Multiple-Regime Regressions with Least Absolutes Deviation, *Journal of Statistical Planning and Inference*, 74, 103-134.

Bhattacharya, P.K. and P.J. Brockwell, 1976, The Minimum of an Additive Process with Applications to Signal Estimation and Storage Theory, *Z. Wahrschein. Verw. Gebiete*, 37, 51-75.

Buchinsky, M., 1998, Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Reserach, *Journal of Human Resources*, 33, 88-126.

Buchinsky, M., and J. Hahn, 1998, An Alternative Estimator for the Censored Quantile Regression Model, *Econometrica*, 653-671.

Cai, Y., 2010, Forecasting for Quantile Self-Exciting Threshold Autoregressive Time Series Models, *Biometrika*, 97, 199-208.

Cai, Y. and J. Stander, 2008, Quantile Self-Exciting Threshold Autoregressive Time Series Models, 2008, *Journal of Time Series Analysis*, 29, 187-202.

Caner, M., 2002, A Note on Least Absolute Deviation Estimation of a Threshold Model, *Econometric Theory*, 18, 800-814.

Chan, K.S., 1993, Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model, *The Annals of Statistics*, 21, 520-533.

Chan, K.S., and R.S. Tsay, 1998, Limiting Properties of the Least Squares Estimator of a Continuous Threshold Autoregressive Model, *Biometrika*, 85, 413-426.

Chen, J., 2008, Estimating and Testing Quantile Regression with Structural Changes, unpublished manuscript, Department of Economics, New York Univeristy.

Chernozhukov, V., and I. Fernández-Val, 2005, Subsampling Inference on Quantile Regression Processes, *Sankhyā*, 67, 253-276.

Chernozhukov, V. et al., 2012, Inference on Counterfactual Distributions, *Econometrica*, forthcoming.

Chernozhukov, V., and C. Hansen, 2006, Instrumental Quantile Regression Inference for Structural and Treatment Effect Models, *Journal of Econometrics*, 132, 491-525.

Durlauf, S.N. and P.A. Johnson, 1995, Multiple Regimes and Cross-Country Growth Behavior, *Journal of Applied Econometrics*, 10, 365-384.

Feng, X. et al., 2011, Wild Bootstrap for Quantile Regression, *Biometrika*, 98, 995-999.

Galvao, A.F. et al., 2011, Threshold Quantile Autoregressive Models, *Journal of Time Series Analysis*, 32, 253-267.

Gutenbrunner, C. and J. Jurečková, 1992, Regression Rank Scores and Regression Quantiles, *The Annals of Statistics*, 20, 305-330.

Hahn, J., 1995, Bootstrapping Quantile Regression Estimators, *Econometric Theory*, 11, 105-121.

Hall, P. and S. Sheather, 1988, On the Distribution of a Studentized Quantile, *Journal of the Royal Statistical Society, Series B*, 50, 381-391.

Hansen, B.E., 1996, Inference when a Nuisance Parameter is not Identified under the Null Hypothesis, *Econometrica*, 64, 413-430.

Hansen, B.E., 2000, Sample Splitting and Threshold Estimation, *Econometrica*, 68, 575-603.

Hansen, B.E., 2011, Threshold Autoregression in Economics, *Statistics and Its Interface*, 4, 123-127.

Härdle, W. and O. Linton, 1994, Applied Nonparametric Methods, *Handbook of Econometrics*, Vol. 4, R.F. Eagle and D.L. McFadden, eds., Elsevier Science B.V., Ch. 38, 2295-2339.

Ichimura, H. and P. Todd, 2007, Implementing Nonparametric and Semiparametric Estimators, *Handbook of Econometrics*, Vol. 6B, J. Heckman and E. Leamer, eds., Elsevier Science B.V., Ch. 74, 5369-5468.

Kato, K., 2009, Asymptotics for Argmin Processes: Convexity Arguments, *Journal of Multivariate Analysis*, 1816-1829.

Khmaladze, E.V., 1981, Martingale Approach in the Theory of Goodness-of-fit Test, *Theory of Probability and its Applications*, 26, 240-257.

Kim, J. and D. Pollard, 1990, Cube Root Asymptotics, *The Annals of Statistics*, 18, 191-219.

Knight, K., 1998, Limiting Distributions for L_1 Regression Estimators Under General Conditions, *The Annals of Statistics,* 26, 755-770.

Koenker, R., 1994, Confidence Intervals for Regression Quantiles, in *Asymptotic Statistics: Proceeding of the 5th Prague Symposium on Asymptotic Statistics*, ed. by P. Mandl and M. Hušková, Heidelburg: Physica-Verlag.

Koenker, R., and G. Bassett, 1978, Regression Quantiles, *Econometrica*, 46, 33-50.

Koenker, R., and G. Bassett, 1982, Robust Tests for Heteroscedasticity Based on Regression Quantiles, *Econornetrica*, 50, 43-61.

Koenker, R. and J. Machado, 1999, Goodness of Fit and Related Inference Processes for Quantile Regression, *Journal of American Statistics Association*, 94, 1296-1310.

Koenker, R. and Z. Xiao, 2002, Inference on the Quantile Regression Process, *Econometrica*, 70, 1583-1612.

Li, Q. and J.S. Racine, 2007, *Nonparametric Econometrics: Theory and Practice*, Princeton, N.J. : Princeton University Press.

Newey, W.K. and D.L. McFadden, 1994, Large Sample Estimation and Hypothesis Testing, *Handbook of Econometrics*, Vol. 4, R.F. Eagle and D.L. McFadden, eds., Elsevier Science B.V., Ch. 36, 2113-2245.

Oka, T., and Z. Qu, 2011, Estimating Structural Changes in Regression Quantiles, *Journal of Econometrics*, 162, 248-267.

Pagan, A. and A. Ullah, 1999, *Nonparametric Econometrics*, New York: Cambridge University Press.

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

Ploberger, W. and W. Krämer, 1992, The CUSUM Test with OLS Residuals, *Econometrica*, 60, 271-285.

Portnoy, S., 1991, Asymptotic Behavior of the Number of Regression Quantile Breakpoints, *SIAM Journal on Scientific and Statistical Computing*, 12, 867-883.

Powell, J., 1986, Censored Regression Quantiles, *Journal of Econometrics,* 32, 143-55.

Powell, J., 1989, Estimation of Monotonic Regression Models under Quantile Restricitons, *Nonparametric and Semiparametric Methods in Econom*trics *and Statistics*, ed. by: W.A. Barnett, J.L. Powell, and G.E. Tauchen, Cambridge, MA: Cambridge University Press, 357-384.

Qu, Z., 2008, Testing for Structural Change in Regression Quanitles, *Journal of Econometrics*, 146, 170-184.

Su, L, and Z. Xiao, 2008, Testing for Parameter Stability in Quantile Regression Models, *Statistics and Probability Letters*, 78, 2768-2775.

Tong, H., 1978, On a Threshold Model, In C. H. Chen, eds., *Pattern Recognition and Signal Processing*, Amsterdam: Sijthoff & Noordhoff, 575-586.

Tong, H., 1983, *Threshold Models in Nonlinear Time Series Analysis*. Lecture Notes in Statistics, New York: Springer-Verlag.

Tong, H., 1990, *Non-linear Time Series: A Dynamical System Approach*, Oxford: Clarendon Press.

Tong, H., 2011, Threshold Models in Time Series Analysis - 30 Years On, *Statistics and Its Interface*, 4, 107-118.

van der Vaart, A.W. and J.A. Wellner, 1996, *Weak Convergence and Empirical Processes: With Applications to Statistics*, New York : Springer.

Yu, P., 2008, Adaptive Estimation of the Threshold Point in Threshold Regression, unpublished manuscript, Department of Economics, University of Auckland.

Yu, P., 2009, Threshold Regression with Endogeneity, unpublished manuscript, Department of Economics, University of Auckland.

Yu, P., 2012, Likelihood Estimation and Inference in Threshold Regression, *Journal of Econometrics*, 167, 274-294.

Yu, P., 2013a, The Bootstrap in Threshold Regression, forthcoming, *Econometric Theory*.

Yu, P., 2013b, Threshold Regression Under Misspecification, unpublished manuscript, Department of Economics, University of Auckland.

Appendix A: Proofs

First, some notations are collected for reference in all lemmas and proofs. The letter C is used as a generic positive constant, which need not be the same in each occurrence. $a_n = n\delta_n'\delta_n$.

$$Q_n(\theta) = P_n(m(\cdot|\theta)), Q(\theta) = P(m(\cdot|\theta)), \mathbb{G}_n(m(w|\theta)) = \sqrt{n}(Q_n(\theta) - Q(\theta)),$$

where $m(w|\theta) = \sum_{t=1}^{T} \rho_{\tau_t}(y - \mathbf{x}'\beta_1(\tau_t) 1(q \le \gamma) - \mathbf{x}'\beta_2(\tau_t) 1(q > \gamma))$, $\theta = (\gamma, \beta_T')'$ with $\beta_T = (\beta_{1T}', \beta_{2T}')'$and $\beta_{\ell T} = \left(\beta_\ell(\tau_1)', \cdots, \beta_\ell(\tau_T)'\right)'$, and for a function f,

$$P_n[f(w)] = n^{-1}\sum_{i=1}^{n} f(w_i), \mathbb{G}_n(f(w)) = n^{-1/2}\sum_{i=1}^{n}(f(w_i) - E[f(w_i)]),$$
$$\mathbb{G}_n\left(\widehat{f}(w)\right) = n^{-1/2}\sum_{i=1}^{n}(f(w_i) - E[f(w_i)])\Big|_{f=\widehat{f}}, \text{ where } \widehat{f} \text{ is an estimated function,}$$

$$\psi_\tau(u) = \tau - 1(u \le 0), \varphi_\tau(u) = -\psi_\tau(u).$$

Proof of Theorem 1. The consistency of $\widehat{\gamma}$ is proved in Lemma 1, and the convergence rate is shown in Lemma 3. From Lemma 5, $n(\widehat{\gamma} - \gamma_0)$ has the same asymptotic distribution as $\arg\min_v D_{Tn}(v)$ with

$$D_{Tn}(v) = \sum_{i=1}^{n} \overline{z}_{1Ti} 1\left(\gamma_0 + \frac{v}{n} < q_i \le \gamma_0\right) + \sum_{i=1}^{n} \overline{z}_{2Ti} 1\left(\gamma_0 < q_i \le \gamma_0 + \frac{v}{n}\right).$$

Now, a modified version of the argmax continuous mapping theorem (Theorem 3.2.2 in van der Vaart and Wellner (1996)) is used to derive the asymptotic distribution.

(i) $D_{Tn}(v) \rightsquigarrow D_T(v)$ on any compact set of v. This is proved in Lemma 7.

(ii) $n(\widehat{\gamma} - \gamma_0) = O_p(1)$. This is proved in Lemma 3.

(iii) $\arg\min_v D_T(v) = O_p(1)$. This is shown in Appendix D of Yu (2012).

(iv) $\arg\min_v D_T(v)$ is unique. This is guaranteed by Assumption D8.

■

Proof of Theorem 2. The consistency of $\widehat{\gamma}$ is proved in Lemma 2, and the convergence rate is shown in Lemma 4. From Lemma 6, $a_n(\widehat{\gamma} - \gamma_0)$ has the same asymptotic distribution as $\arg\min_v C_{Tn}(v)$, where

$$C_{Tn}(v) = \begin{cases} \sum_{t=1}^{T} \delta_{tn}' \left[\sum_{i=1}^{n} \mathbf{x}_i \psi_{\tau_t}(e_{1\tau_t i}) 1\left(\gamma_0 + \frac{v}{a_n} < q_i \le \gamma_0\right)\right] + \frac{f_q(\gamma_0)}{2} \frac{\pi_{1Tn}}{\delta_n'\delta_n} |v|, & \text{if } v \le 0, \\ -\sum_{t=1}^{T} \delta_{tn}' \left[\sum_{i=1}^{n} \mathbf{x}_i \psi_{\tau_t}(e_{2\tau_t i}) 1\left(\gamma_0 < q_i \le \gamma_0 + \frac{v}{a_n}\right)\right] + \frac{f_q(\gamma_0)}{2} \frac{\pi_{2Tn}}{\delta_n'\delta_n} v, & \text{if } v > 0. \end{cases}$$

We now apply Theorem 2.7 of Kim and Pollard (1990) to find the asymptotic distribution of $a_n(\widehat{\gamma} - \gamma_0)$.

(i) $C_{Tn}(v) \rightsquigarrow C_T(v) \in \mathbf{C}_{\min}(\mathbb{R})$, where $\mathbf{C}_{\min}(\mathbb{R})$ is defined as the subset of continuous functions $x(\cdot) \in \mathbf{B}_{\text{loc}}(\mathbb{R})$ for which (i) $x(t) \to \infty$ as $|t| \to \infty$ and (ii) $x(t)$ achieves its minimum at a unique point in $\mathbb{R}$, and $\mathbf{B}_{\text{loc}}(\mathbb{R})$ is the space of all locally bounded real functions on $\mathbb{R}$, endowed with the uniform metric on compacta. The weak convergence is proved in Lemma 8. We now check $C_T(v) \in \mathbf{C}_{\min}(\mathbb{R})$. It is not hard to check $C_T(v)$ is continuous, has a unique minimum (see Lemma 2.6 of Kim and Pollard (1990)), and $\lim_{|v|\to\infty} C_T(v) = \infty$ almost surely (which is true since $\lim_{|v|\to\infty} W_\ell(v)/|v| = 0$ almost surely).

(ii) $a_n(\widehat{\gamma}-\gamma_0)=O_p(1)$. This is proved in Lemma 4.

So

$$a_n(\widehat{\gamma}-\gamma_0)\xrightarrow{d}\arg\min_v C_T(v).$$

Making the change-of-variables $v=\frac{\sigma_T^2}{\pi_{1T}^2 f_q(\gamma_0)}r$, noting the distributional equality $W_\ell(a^2r)=aW_\ell(r)$, we can rewrite the asymptotic distribution as

$$\begin{aligned}
\arg\min_v C_T(v) &= \arg\max_v\{-C_T(v)\}\\
&= \frac{\sigma_T^2}{\pi_{1T}^2 f_q(\gamma_0)}\arg\max_v\left\{-C_T\left(\frac{\sigma_T^2}{\pi_{1T}^2 f_q(\gamma_0)}r\right)\right\}\\
&= \frac{\sigma_T^2}{\pi_{1T}^2 f_q(\gamma_0)}\arg\max_v\begin{cases}-\sqrt{f_q(\gamma_0)}\sigma_T W_1\left(-\frac{\sigma_T^2}{\pi_{1T}^2 f_q(\gamma_0)}r\right)-\frac{1}{2}\frac{\sigma_T^2}{\pi_{1T}}|r|, & \text{if } r\le 0,\\ -\sqrt{f_q(\gamma_0)}\sigma_T W_2\left(\frac{\sigma_T^2}{\pi_{1T}^2 f_q(\gamma_0)}r\right)-\frac{\sigma_T^2}{2}\frac{\pi_{2T}}{\pi_{1T}^2}|r|, & \text{if } r>0,\end{cases}\\
&= \frac{\sigma_T^2}{\pi_{1T}^2 f_q(\gamma_0)}\arg\max_v\begin{cases}\frac{\sigma_T^2}{\pi_{1T}}W_1(-r)-\frac{1}{2}\frac{\sigma_T^2}{\pi_{1T}}|r|, & \text{if } r\le 0,\\ \frac{\sigma_T^2}{\pi_{1T}}W_2(r)-\frac{\sigma_T^2}{2}\frac{\pi_{2T}}{\pi_{1T}^2}|r|, & \text{if } r>0,\end{cases}\\
&= \frac{\sigma_T^2}{\pi_{1T}^2 f_q(\gamma_0)}\arg\max_v\begin{cases}W_1(-r)-\frac{1}{2}|r|, & \text{if } r\le 0,\\ W_2(r)-\frac{1}{2}\frac{\pi_{2T}}{\pi_{1T}}|r|, & \text{if } r>0,\end{cases}\\
&= \frac{\sigma_T^2}{\pi_{1T}^2 f_q(\gamma_0)}\Lambda(\xi).
\end{aligned}$$

By Slutsky's theorem,

$$nf_q(\gamma_0)\left(\frac{\pi_{1Tn}}{\sigma_{Tn}}\right)^2(\widehat{\gamma}-\gamma_0)\xrightarrow{d}\Lambda(\xi).$$

■

Proof of Corollary 1. From the proof of Theorem 2 and the continuous mapping theorem,

$$n\left(Q_{Tn}(\gamma_0)-Q_{Tn}(\widehat{\gamma})\right)\xrightarrow{d}\sup_v\{-C_T(v)\}.$$

Note that

$$\begin{aligned}
\sup_v\{-C_T(v)\} &= \sup_r\begin{cases}-\sqrt{f_q(\gamma_0)}\sigma_T W_1\left(-\frac{\sigma_T^2}{\pi_{1T}^2 f_q(\gamma_0)}r\right)-\frac{1}{2}\frac{\sigma_T^2}{\pi_{1T}}|r|, & \text{if } r\le 0,\\ -\sqrt{f_q(\gamma_0)}\sigma_T W_2\left(\frac{\sigma_T^2}{\pi_{1T}^2 f_q(\gamma_0)}r\right)-\frac{\sigma_T^2}{2}\frac{\pi_{2T}}{\pi_{1T}^2}|r|, & \text{if } r>0,\end{cases}\\
&= \sup_r\begin{cases}\frac{\sigma_T^2}{\pi_{1T}}W_1(-r)-\frac{1}{2}\frac{\sigma_T^2}{\pi_{1T}}|r|, & \text{if } r\le 0,\\ \frac{\sigma_T^2}{\pi_{1T}}W_2(r)-\frac{\sigma_T^2}{2}\frac{\pi_{T2}}{\pi_{1T}^2}|r|, & \text{if } r>0,\end{cases}\\
&= \frac{\sigma_T^2}{\pi_{1T}}\sup_r\begin{cases}W_1(-r)-\frac{1}{2}|r|, & \text{if } r\le 0,\\ W_2(r)-\frac{1}{2}\frac{\pi_{2T}}{\pi_{1T}}|r|, & \text{if } r>0,\end{cases}
\end{aligned}$$

so $\sup_v\{-C_T(v)\}=\frac{\sigma_T^2}{\pi_{1T}}\max\{M_1,M_2\}\equiv\frac{\sigma_T^2}{\pi_{1T}}M$, where $M_1=\sup_{r\le 0}\left\{W_1(-r)-\frac{1}{2}|r|\right\}$, $M_2=\sup_{r\ge 0}\left\{W_2(r)-\frac{1}{2}\xi|r|\right\}$, and M_1 and M_2 are independent. From Bhattacharya and Brockwell (1976), M_1 follows the standard exponential function, and M_2 follows an exponential distribution with mean $1/\xi$. It follows that

$$P(M\le x)=P(M_1\le x,M_2\le x)=P(M_1\le x)P(M_2\le x)=(1-e^{-x})(1-e^{-\xi x}).$$

By Slutsky's theorem, the result of the theorem follows. ■

Remark 1 *In the least squares estimation, $C_T(v)$ is changed to*

$$C(v) = \begin{cases} \sqrt{f_q(\gamma_0)V_1}W_1(-v) + \frac{f_q(\gamma_0)}{2}D|v|, & \text{if } v \leq 0, \\ \sqrt{f_q(\gamma_0)V_2}W_2(v) + \frac{f_q(\gamma_0)}{2}D|v|, & \text{if } v > 0, \end{cases}$$

where $D = \lim_{n\to\infty} \delta_n' E[\mathbf{x}\mathbf{x}'|q=\gamma_0]\delta_n/\delta_n'\delta_n$, *and* $V_\ell = \lim_{n\to\infty} \delta_n' E\left[\mathbf{x}\mathbf{x}'e_\ell^2|q=\gamma_0\right]\delta_n/\delta_n'\delta_n$. *So the asymptotic distribution of the LSE is*

$$a_n(\widehat{\gamma}_{LSE} - \gamma_0) \xrightarrow{d} \arg\min_v C(v),$$

and

$$nf_q(\gamma_0)\frac{D_n^2}{V_{1n}}(\widehat{\gamma}_{LSE} - \gamma_0) \xrightarrow{d} \widetilde{\Lambda}\left(\frac{V_2}{V_1}, 1\right).$$

where $\widetilde{\Lambda}(\phi,\xi)$ *is defined in (5),* $D_n = \delta_n' E[\mathbf{x}\mathbf{x}'|q=\gamma_0]\delta_n$, *and* $V_{1n} = \delta_n' E\left[\mathbf{x}\mathbf{x}'e_1^2|q=\gamma_0\right]\delta_n$. *Correspondingly,*

$$S_n(\gamma_0) - S_n(\widehat{\gamma}_{LSE}) \xrightarrow{d} \sup_v\{-C(v)\} = \frac{V_1}{D}\sup_r \begin{cases} W_1(-r) - \frac{1}{2}|r|, & \text{if } r \leq 0, \\ \sqrt{\frac{V_2}{V_1}}W_2(r) - \frac{1}{2}|r|, & \text{if } r > 0, \end{cases}$$

so

$$\frac{D_n}{V_{1n}}(S_n(\gamma_0) - S_n(\widehat{\gamma}_{LSE})) \xrightarrow{d} M_{LSE},$$

where $S_n(\gamma) = \frac{1}{2}\sum_{i=1}^n (y_i - \mathbf{x}_i'\widehat{\beta}_1(\gamma)1(q_i \leq \gamma) - \mathbf{x}_i'\widehat{\beta}_2(\gamma)1(q_i > \gamma))^2$ *is the profiled objective function of the LSE for a given* γ, *and*

$$P(M_{LSE} \leq x) = (1 - e^{-x})(1 - e^{-V_1 x/V_2}).$$

When $e_\ell = \sigma_\ell e$ *with* e *being independent of* $\mathbf{x}$, $V_1/V_2 = \sigma_{10}^2/\sigma_{20}^2$.

Proof of Theorem 3. As in the proof of Theorem 3 of Angrist et al. (2006), we divide the proof of the weak limit of $\widehat{\beta}_\ell(\cdot)$ into two steps. For simplicity, take $\widehat{\beta}_1(\cdot)$ as an example.

Step 1: Uniform consistency of $\widehat{\beta}_1(\cdot)$.

For each τ in $\mathcal{T}$, $\widehat{\beta}_1(\cdot)$ minimizes $Q_n(\tau,\beta_1,\widehat{\gamma}) \equiv P_n\left[\left(\rho_\tau(y - \mathbf{x}'\beta_1) - \rho_\tau\left(y - \mathbf{x}'\beta_1^0(\tau)\right)\right)1(q \leq \widehat{\gamma})\right]$. Define $Q(\tau,\beta_1,\gamma_0) \equiv E\left[\left(\rho_\tau(y - \mathbf{x}'\beta_1) - \rho_\tau\left(y - \mathbf{x}'\beta_1^0(\tau)\right)\right)1(q \leq \gamma_0)\right]$. It is easy to show that $E[\|\mathbf{x}\|] < \infty$ implies that $E\left[\left|\left(\rho_\tau(y - \mathbf{x}'\beta_1) - \rho_\tau\left(y - \mathbf{x}'\beta_1^0(\tau)\right)\right)\right|1(q \leq \gamma_0)\right] < \infty$. Therefore, $Q(\tau,\beta_1,\gamma_0)$ is finite and by the stated assumptions (especially, Assumption D6), it is uniquely minimized at $\beta_1^0(\tau)$ for each τ in $\mathcal{T}$.

We first show the uniform convergence, namely for any compact set B, $Q_n(\tau,\beta_1,\gamma) = Q(\tau,\beta_1,\gamma) + o_p(1)$ uniformly in $(\tau,\beta_1,\gamma) \in \mathcal{T}\times B\times\Gamma$. This statement holds pointwise by the weak law of large numbers (WLLN). The empirical process $(\tau,\beta_1,\gamma) \mapsto Q_n(\tau,\beta_1,\gamma)$ is stochastic equicontinuous because

$$\begin{aligned}
&|Q_n(\tau',\beta_1',\gamma') - Q_n(\tau,\beta_1,\gamma)| \\
&\leq |Q_n(\tau',\beta_1',\gamma') - Q_n(\tau,\beta_1,\gamma')| + |Q_n(\tau,\beta_1,\gamma') - Q_n(\tau,\beta_1,\gamma)| \\
&\leq 2P_n[\|\mathbf{x}\|]\sup_{\beta_1\in B}\|\beta_1\||\tau' - \tau| + 2P_n[\|\mathbf{x}\|]\|\beta_1' - \beta_1\| + 2P_n\left[\|\mathbf{x}\|\frac{1(\gamma'\wedge\gamma < q \leq \gamma'\vee\gamma)}{\sqrt{|\gamma' - \gamma|}}\right]\sup_{\beta_1\in B}\|\beta_1\|\sqrt{|\gamma' - \gamma|} \\
&= O_p(1)|\tau' - \tau| + O_p(1)\|\beta_1' - \beta_1\| + O_p(1)\sqrt{|\gamma' - \gamma|}.
\end{aligned}$$

Hence, the convergence also holds uniformly.

Next, we show uniform consistency. Consider a collection of closed balls $B_M\left(\beta_1^0(\tau)\right)$ of radius M and center $\beta_1^0(\tau)$, and let $\beta_{1M}(\tau) = \beta_1^0(\tau) + \delta_M(\tau)\cdot v(\tau)$, where $v(\tau)$ is a direction vector with unity norm

$\|v(\tau)\| = 1$ and $\delta_M(\tau)$ is a positive scalar such that $\delta_M(\tau) \geq M$. Then uniformly in $\tau \in \mathcal{T}$,

$$(M/\delta_M(\tau)) \cdot \left(Q_n\left(\tau, \beta_{1M}(\tau), \widehat{\gamma}\right) - Q_n\left(\tau, \beta_1^0(\tau), \widehat{\gamma}\right)\right) \overset{(1)}{\geq} Q_n\left(\tau, \beta_{1M}^*(\tau), \widehat{\gamma}\right) - Q_n\left(\tau, \beta_1^0(\tau), \widehat{\gamma}\right)$$
$$\overset{(2)}{\geq} Q\left(\tau, \beta_{1M}^*(\tau), \gamma_0\right) - Q\left(\tau, \beta_1^0(\tau), \gamma_0\right) + o_p(1) \overset{(3)}{\geq} \epsilon_M + o_p(1)$$

for some $\epsilon_M > 0$, where (1) follows by convexity in β_1 for $\beta_{1M}^*(\tau)$ the point of the boundary of $B_M\left(\beta_1^0(\tau)\right)$ on the line connecting $\beta_{1M}(\tau)$ and $\beta_1^0(\tau)$; (2) follows by the uniform convergence established above and $\widehat{\gamma} - \gamma_0 = o_p(1)$; and (3) follows because $\beta_1^0(\tau)$ is the unique minimizer of $Q\left(\tau, \beta_1^0(\tau), \gamma_0\right)$ uniformly in $\tau \in \mathcal{T}$, by convexity and Assumption D6. Hence for any $M > 0$, the minimizer $\widehat{\beta}_1(\cdot)$ must be in the radius-M ball centered at $\beta_1^0(\tau)$ uniformly for all $\tau \in \mathcal{T}$, with probability approaching 1.

Step 2: Asymptotic Gaussianity of $\sqrt{n}\left(\widehat{\beta}_1(\cdot) - \beta_1^0(\cdot)\right)$.

First, by the computational properties of $\widehat{\beta}_1(\cdot)$, for all $\tau \in \mathcal{T}$ (cf. Theorem 3.3 in Koenker and Bassett (1978)) we have that

$$\left\| P_n\left(\psi_\tau\left(y - \mathbf{x}'\widehat{\beta}_1(\tau)\right)\mathbf{x}1(q \leq \widehat{\gamma})\right)\right\| \leq d \cdot \sup_{i \leq n} \|\mathbf{x}_i\| / n.$$

Note that $E\left[\|\mathbf{x}_i\|^{2+\varepsilon}\right] < \infty$, $\varepsilon > 0$, implies $\sup_{i \leq n} \|\mathbf{x}_i\| = o_p(n^{1/2})$ because $P\left(\sup_{i \leq n} \|\mathbf{x}_i\| > \sqrt{n}\right) \leq nP\left(\|\mathbf{x}_i\| > \sqrt{n}\right) \leq n\, E\left[\|\mathbf{x}_i\|^{2+\varepsilon}\right] \Big/ n^{(2+\varepsilon)/2} = o(1)$. Hence uniformly in $\tau \in \mathcal{T}$,

$$\sqrt{n} P_n\left(\psi_\tau\left(y - \mathbf{x}'\widehat{\beta}_1(\tau)\right)\mathbf{x}1(q \leq \widehat{\gamma})\right) = o_p(1). \tag{11}$$

Second, $(\tau, \beta_1, \gamma) \mapsto \mathbb{G}_n\left(\psi_\tau\left(y - \mathbf{x}'\beta_1\right)\right)\mathbf{x}1(q \leq \gamma))$ is stochastic equicontinuous over $\mathcal{T} \times B \times \Gamma$, where B is any compact set, with respect to the $L_2(P)$ pseudo-metric

$$\rho\left(\left(\tau', \beta_1', \gamma'\right), (\tau, \beta_1, \gamma)\right)^2 \equiv \max_{j \in 1, \cdots, d} E\left[\left(\psi_{\tau'}\left(y - \mathbf{x}'\beta_1'\right)\mathbf{x}_j 1(q \leq \gamma') - \psi_\tau\left(y - \mathbf{x}'\beta_1\right)\mathbf{x}_j 1(q \leq \gamma)\right)^2\right]$$

for $j = 1, \cdots, d$ indexing the components of $\mathbf{x}$. Note that the functional class $\{\psi_\tau\left(y - \mathbf{x}'\beta_1\right)\mathbf{x}1(q \leq \gamma), \tau \in \mathcal{T}, \beta_1 \in B, \gamma \in \Gamma\}$ is formed as $(\mathcal{T} - \mathcal{F})\,\mathbf{x}\mathcal{Q}$, where $\mathcal{F} = \{1\left(y \leq \mathbf{x}'\beta_1\right), \beta_1 \in B\}$ and $\mathcal{Q} = \{1(q \leq \gamma), \gamma \in \Gamma\}$ is a VC subgraph class and hence a bounded Donsker class. Hence $(\mathcal{T} - \mathcal{F})$ is also bounded Donsker and $(\mathcal{T} - \mathcal{F})\,\mathbf{x}\mathcal{Q}$ is, therefore, Donsker with a square-integrable envelope $2 \cdot \max_{j \in 1, \cdots, d} |\mathbf{x}_j|$ by Theorem 2.10.6 in van der Vaart and Wellner (1996). Stochastic equicontinuity then is part of being Donsker.

Third, by stochastic equicontinuity of $(\tau, \beta_1, \gamma) \mapsto \mathbb{G}_n\left(\psi_\tau\left(y - \mathbf{x}'\beta_1\right)\right)\mathbf{x}1(q \leq \gamma))$ we have that

$$\mathbb{G}_n\left(\psi_\tau\left(y - \mathbf{x}'\widehat{\beta}_1(\tau))\right)\mathbf{x}1(q \leq \widehat{\gamma})\right) = \mathbb{G}_n\left(\psi_\tau\left(y - \mathbf{x}'\beta_1^0(\tau))\right)\mathbf{x}1(q \leq \gamma_0)\right) + o_p(1) \text{ in } \ell^\infty(\mathcal{T}), \tag{12}$$

which follows from $\sup_{\tau \in \mathcal{T}} \left\|\widehat{\beta}_1(\tau) - \beta_1^0(\tau)\right\| = o_p(1)$, $\widehat{\gamma} - \gamma_0 = o_p(1)$, and resulting convergence with respect to the pseudo-metric $\sup_{\tau \in \mathcal{T}} \rho\left(\left(\tau', \widehat{\beta}_1(\tau), \widehat{\gamma}\right), \left(\tau, \beta_1^0(\tau), \gamma_0\right)\right)^2 = o_p(1)$. The last result is immediate from $\sup_{\tau \in \mathcal{T}} \rho\left(\left(\tau', \beta_1'(\tau), \gamma'\right), (\tau, \beta_1(\tau), \gamma)\right)^2 \leq C \cdot \left(\sup_{\tau \in \mathcal{T}} \|\beta_1'(\tau) - \beta_1(\tau)\|^{\varepsilon/(2+\varepsilon)} + |\gamma' - \gamma|\right)$ by the Hölder's inequality, where C can take $\left(\overline{f} \cdot \left(E\left[\|\mathbf{x}\|^2\right]\right)^{1/2}\right)^{\varepsilon/(2(2+\varepsilon))} \cdot \left(E\left[\|\mathbf{x}\|^{2+\varepsilon}\right]\right)^{2/(2+\varepsilon)} + 2\overline{f}_q \sup_{\gamma \in \mathcal{N}_\gamma} E\left[\|\mathbf{x}\|^2 \middle| q = \gamma\right]$, $\overline{f}$ is the a.s. upper bound on $f(y|x,q)$, and $\mathcal{N}_\gamma$ is a neighborhood of γ_0.

Furthermore, the following expansion is valid uniformly in $\tau \in \mathcal{T}$:

$$E\left[\psi_\tau\left(y-\mathbf{x}'\beta_1\right)\right)\mathbf{x}1(q\leq\gamma)]\big|_{\beta_1=\widehat{\beta}_1(\tau),\gamma=\widehat{\gamma}} = \left[-J_1(\tau)+o_p(1)\right]\left(\widehat{\beta}_1(\tau)-\beta_1^0(\tau)\right)+o_p(n^{-1/2}). \tag{13}$$

Indeed, by Taylor expansion,

$$\begin{aligned} E\left[\psi_\tau\left(y-\mathbf{x}'\beta_1\right)\right)\mathbf{x}1(q\leq\gamma)]\big|_{\beta_1=\widehat{\beta}_1(\tau),\gamma=\widehat{\gamma}} = & -E\left[f_{y|x,q}\left(\mathbf{x}'b_1(\tau)|x,q\right)\mathbf{x}\mathbf{x}'1(q\leq\gamma_0)\right]\big|_{b_1(\tau)=\beta_1^*(\tau)}\left(\widehat{\beta}_1(\tau)-\beta_1^0(\tau)\right) \\ & + E\left[\psi_\tau\left(y-\mathbf{x}'\beta_1^0(\tau)\right)\right)\mathbf{x}|q=\gamma\right]f_q(\gamma)\big|_{\gamma=\gamma^*}\left(\widehat{\gamma}-\gamma_0\right), \end{aligned}$$

where $\beta_1^*(\tau)$ is on the line connecting $\widehat{\beta}_1(\tau)$ and $\beta_1^0(\tau)$ for each τ and can be different for each row of the Jacobian matrix, and γ^* is between $\widehat{\gamma}$ and γ_0. $E\left[f_{y|x,q}\left(\mathbf{x}'b_1(\tau)|x,q\right)\mathbf{x}\mathbf{x}'1(q\leq\gamma_0)\right]\big|_{b_1(\tau)=\beta_1^*(\tau)} = J_1(\tau)+o_p(1)$ by the uniform consistency of $\widehat{\beta}_1(\tau)$, and the assumed uniform continuity and boundedness of the mapping $y\mapsto f\left(y|x,q\right)$, uniformly for $q\leq\gamma_0$ and $x\in\mathbb{X}$. $E\left[\psi_\tau\left(y-\mathbf{x}'\beta_1^0(\tau)\right)\right)\mathbf{x}|q=\gamma\right]f_q(\gamma)\big|_{\gamma=\gamma^*}\left(\widehat{\gamma}-\gamma_0\right)=0$ if $\gamma^*\leq\gamma_0$ and is $O_p(\left\|\beta_1^0(\tau)-\beta_2^0(\tau)\right\|\left(\widehat{\gamma}-\gamma_0\right))$ if $\gamma^*>\gamma_0$. In whatever case, $E\left[\psi_\tau\left(y-\mathbf{x}'\beta_1^0(\tau)\right)\right)\mathbf{x}|q=\gamma\right]f_q(\gamma)\big|_{\gamma=\gamma^*}\left(\widehat{\gamma}-\gamma_0\right) = O_p(\left\|\delta_n\right\|a_n^{-1}) = o_p(n^{-1/2})$.

Fourth, we have that

$$o_p(1) = \left[\left(-J_1(\tau)+o_p(1)\right)\sqrt{n}\left(\widehat{\beta}_1(\tau)-\beta_1^0(\tau)\right)\right]+\mathbb{G}_n\left(\psi_\cdot\left(y-\mathbf{x}'\beta_1^0(\cdot)\right)\right)\mathbf{x}1(q\leq\gamma_0)\right), \tag{14}$$

because the left-hand side of (11) is equal to the left-hand side of $n^{1/2}$(13) plus the left-hand side of (12). Therefore, using the mineig$[J_1(\tau)]\geq\lambda_1>0$ uniformly in $\tau\in\mathcal{T}$,

$$\sup_{\tau\in\mathcal{T}}\left\|\mathbb{G}_n\left(\psi_\tau\left(y-\mathbf{x}'\beta_1^0(\tau)\right)\right)\mathbf{x}1(q\leq\gamma_0)\right)+o_p(1)\right\| \geq \left(\sqrt{\lambda_1}+o_p(1)\right)\cdot\sup_{\tau\in\mathcal{T}}\sqrt{n}\left\|n\left(\widehat{\beta}_1(\tau)-\beta_1^0(\tau)\right)\right\|, \tag{15}$$

where for a matrix A, mineig$[A]$ denotes the minimum eigenvalue of A.

Fifth, the mapping $\tau\mapsto\beta_1^0(\tau)$ is continuous by the implicit function theorem and stated assumptions. In fact, because $\beta_1^0(\tau)$ solves $E\left[\left(\tau-1\left(y\leq\mathbf{x}'\beta_1\right)\right)1(q\leq\gamma_0)\right]=0$, $d\beta_1^0(\tau)/d\tau = J_1(\tau)^{-1}E[\mathbf{x}]$. Hence $\tau\mapsto\mathbb{G}_n\left(\psi_\tau\left(y-\mathbf{x}'\beta_1^0(\tau)\right)\right)\mathbf{x}1(q\leq\gamma_0)\right)$ is stochastic equicontinuous over $\mathcal{T}$ for the pseudo-metric given by $\rho\left(\tau',\tau\right)=\rho\left(\left(\tau',\beta_1^0(\tau'),\gamma_0\right),\left(\tau,\beta_1^0(\tau),\gamma_0\right)\right)$. Stochastic equicontinuity of $\tau\mapsto\mathbb{G}_n\left(\psi_\tau\left(y-\mathbf{x}'\beta_1^0(\tau)\right)\right)\mathbf{x}1(q\leq\gamma_0)\right)$ and a multivariate central limit theorem imply that

$$\mathbb{G}_n\left(\psi_\cdot\left(y-\mathbf{x}'\beta_1^0(\cdot)\right)\right)\mathbf{x}1(q\leq\gamma_0)\right)\rightsquigarrow Z_1(\cdot)\text{ in }\ell^\infty\left(\mathcal{T}\right), \tag{16}$$

where $Z_1(\cdot)$ is a Gaussian process with covariance function $\Sigma_1\left(\cdot,\cdot\right)$ specified in the statement of Theorem 3. Therefore, the left-hand side of (15) is $O_p(n^{-1/2})$, implying $\sup_{\tau\in\mathcal{T}}\sqrt{n}\left\|n\left(\widehat{\beta}_1(\tau)-\beta_1^0(\tau)\right)\right\|=O_p(1)$.

Finally, the latter fact and (14)-(16) imply that in $\ell^\infty\left(\mathcal{T}\right)$,

$$J_1(\cdot)\sqrt{n}\left(\widehat{\beta}_1(\cdot)-\beta_1^0(\cdot)\right)=\mathbb{G}_n\left(\psi_\cdot\left(y-\mathbf{x}'\beta_1^0(\cdot)\right)\right)\mathbf{x}1(q\leq\gamma_0)\right)+o_p(1)\rightsquigarrow Z_1(\cdot).$$

The proof for the weak limit of $\widehat{\beta}_\ell\left(\cdot\right)$ does not rely on whether δ_n is fixed or shrinking, so can be applied to both cases.

At last, we prove the asymptotic independence among $\widehat{\gamma}$, $\widehat{\beta}_1(\cdot)$ and $\widehat{\beta}_2(\cdot)$. From Lemma 5, $n(\widehat{\gamma}-\gamma_0)$ has the same asymptotic distribution as $\arg\min_v D_{Tn}(v)$. From the above proof, $J_1(\cdot)\sqrt{n}\left(\widehat{\beta}_1(\cdot)-\beta_1^0(\cdot)\right)$ has the same weak limit as $\mathbb{G}_n\left(\psi_\cdot\left(y-\mathbf{x}'\beta_1^0(\cdot)\right)\right)\mathbf{x}1(q\leq\gamma_0)\right)$, and $J_2(\cdot)\sqrt{n}\left(\widehat{\beta}_2(\cdot)-\beta_2^0(\cdot)\right)$ has the same weak limit as $\mathbb{G}_n\left(\psi_\cdot\left(y-\mathbf{x}'\beta_2^0(\cdot)\right)\right)\mathbf{x}1(q>\gamma_0)\right)$. We only prove the result for a pair of fixed v_1 and v_2, and fixed τ_1 and

τ_2, or the Cramér-Wold device can be used. Define

$$S_{1i} = \overline{z}_{1Ti} 1\left(\gamma_0 + \frac{v_1}{n} < q_i \leq \gamma_0\right),\ S_{2i} = \overline{z}_{2Ti} 1\left(\gamma_0 < q_i \leq \gamma_0 + \frac{v_2}{n}\right),$$
$$S_{3i} = \frac{1}{\sqrt{n}}\psi_{\tau_1}\left(y - \mathbf{x}'\beta_1^0(\tau_1))\right)\mathbf{x}1(q \leq \gamma_0) \equiv \frac{1}{\sqrt{n}}s_{3i}, S_{4i} = \frac{1}{\sqrt{n}}\psi_{\tau_2}\left(y - \mathbf{x}'\beta_2^0(\tau_2))\right)\mathbf{x}1(q > \gamma_0) \equiv \frac{1}{\sqrt{n}}s_{4i},$$

where $v_1 \leq 0$ and $v_2 > 0$. Since

$$\exp\left\{\sqrt{-1}t_1 S_{1i}\right\} = 1 + 1\left(\gamma_0 + \frac{v_1}{n} < q_i \leq \gamma_0\right)\left[\exp\left\{\sqrt{-1}t_1\overline{z}_{1Ti}\right\} - 1\right],$$
$$\exp\left\{\sqrt{-1}t_2 S_{2i}\right\} = 1 + 1\left(\gamma_0 < q_i \leq \gamma_0 + \frac{v_2}{n}\right)\left[\exp\left\{\sqrt{-1}t_2\overline{z}_{2Ti}\right\} - 1\right],$$

it follows

$$\begin{aligned}
&E\left[\exp\left\{\sqrt{-1}\left[t_1 S_{1i} + t_2 S_{2i} + t_3' S_{3i} + t_4' S_{4i}\right]\right\}\right] \\
&= E\left[\exp\left\{\sqrt{-1}t_{34}' s_{34i}/\sqrt{n}\right\}\right] - \frac{v_1}{n} f_q(\gamma_0) E\left[\exp\left\{\sqrt{-1}t_{34}' s_{34i}/\sqrt{n}\right\}\left[\exp\left\{\sqrt{-1}t_1\overline{z}_{1Ti}\right\} - 1\right] \middle| q_i = \gamma_0\right] \\
&+ \frac{v_2}{n} f_q(\gamma_0) E\left[\exp\left\{\sqrt{-1}t_{34}' s_{34i}/\sqrt{n}\right\}\left[\exp\left\{\sqrt{-1}t_2\overline{z}_{2Ti}\right\} - 1\right] \middle| q_i = \gamma_0\right] + o\left(\frac{1}{n}\right) \\
&= 1 + \frac{1}{n}\left[-\frac{1}{2}t_{34}'\Sigma(\tau_1,\tau_2)t_{34}' - f_q(\gamma_0) v_1 \left(E\left[\left[\exp\left\{\sqrt{-1}t_1\overline{z}_{1Ti}\right\}\right] \middle| q_i = \gamma_0\right] - 1\right)\right. \\
&\left. + f_q(\gamma_0) v_2 \left(E\left[\left[\exp\left\{\sqrt{-1}t_2\overline{z}_{2Ti}\right\}\right] \middle| q_i = \gamma_0\right] - 1\right)\right] + o\left(\frac{1}{n}\right),
\end{aligned}$$

where $t_{34} = (t_3', t_4')'$, $s_{34i} = (s_{3i}', s_{4i}')'$, $o(1)$ in the first equality is a quantity going to zero uniformly over $i = 1, \cdots, n$ from Assumption D4, the last equality is from the Taylor expansion of $\exp\left\{\sqrt{-1}t_{34}' s_{34i}/\sqrt{n}\right\}$, and

$$\Sigma(\tau_1,\tau_2) = E\left[s_{34i}s_{34i}'\right] = \operatorname{diag}\left\{\Sigma_1(\tau_1,\tau_1), \Sigma_2(\tau_2,\tau_2)\right\}.$$

So

$$\begin{aligned}
&E\left[\exp\left\{\sqrt{-1}\left[t_1\sum_{i=1}^n S_{1i} + t_2\sum_{i=1}^n S_{2i} + t_3'\sum_{i=1}^n S_{3i} + t_4'\sum_{i=1}^n S_{4i}\right]\right\}\right] \\
&= \prod_{i=1}^n E\left[\exp\left\{\sqrt{-1}\left[t_1 S_{1i} + t_2 S_{2i} + t_3' S_{3i} + t_4' S_{4i}\right]\right\}\right] \\
&\to \exp\left\{-\frac{1}{2}t_3'\Sigma_1(\tau_1,\tau_1)t_3' - \frac{1}{2}t_4'\Sigma_2(\tau_2,\tau_2)t_4'\right. \\
&- f_q(\gamma_0) v_1 \left(E\left[\left[\exp\left\{\sqrt{-1}t_1\overline{z}_{1Ti}\right\}\right] \middle| q_i = \gamma_0\right] - 1\right) \\
&\left. + f_q(\gamma_0) v_2 \left(E\left[\left[\exp\left\{\sqrt{-1}t_2\overline{z}_{2Ti}\right\}\right] \middle| q_i = \gamma_0\right] - 1\right)\right\}.
\end{aligned}$$

As a result, $\widehat{\gamma}$, $\widehat{\beta}_1(\tau_1)$ and $\widehat{\beta}_2(\tau_2)$ are asymptotically independent.

In the case with shrinking threshold effects, $n(\widehat{\gamma} - \gamma_0)$ has the same asymptotic distribution as $\arg\min_v C_{Tn}(v)$ from Lemma 6. Redefine

$$S_{1i} = \sum_{t=1}^T \delta_{tn}'\mathbf{x}_i\psi_{\tau_t}\left(e_{1\tau_t i}\right)1\left(\gamma_{v_1} < q_i \leq \gamma_0\right), S_{2i} = -\sum_{t=1}^T \delta_{tn}'\mathbf{x}_i\psi_{\tau_t}\left(e_{2\tau_t i}\right)1\left(\gamma_0 < q_i \leq \gamma_{v_2}\right),$$

where $\gamma_v = \gamma_0 + v/a_n$. Then

$$
\begin{aligned}
&E\left[\exp\left\{\sqrt{-1}\left[t_1 S_{1i} + t_2 S_{2i} + t_3' S_{3i} + t_4' S_{4i}\right]\right\}\right] \\
&= 1 - \frac{1}{2} t_1^2 \sum_{t=1}^{T} \sum_{t'=1}^{T} (\tau_t \wedge \tau_{t'} - \tau_t \tau_{t'}) \delta_{tn}' E\left[\mathbf{x}_i \mathbf{x}_i' 1\left(\gamma_{v_1} < q_i \leq \gamma_0\right)\right] \delta_{t'n} \\
&- \frac{1}{2} t_2^2 \sum_{t=1}^{T} \sum_{t'=1}^{T} (\tau_t \wedge \tau_{t'} - \tau_t \tau_{t'}) \delta_{tn}' E\left[\mathbf{x}_i \mathbf{x}_i' 1(\gamma_0 < q_i \leq \gamma_{v_2})\right] \delta_{t'n} - \frac{1}{n} \frac{1}{2} t_{34}' \Sigma(\tau_1, \tau_2) t_{34}' + o\left(\frac{1}{n}\right) \\
&= 1 - \frac{f_q(\gamma_0)}{2n} |v_1| t_1^2 \sum_{t=1}^{T} \sum_{t'=1}^{T} (\tau_t \wedge \tau_{t'} - \tau_t \tau_{t'}) \frac{\delta_{tn}'}{\|\delta_n\|} E\left[\mathbf{x}_i \mathbf{x}_i' | q_i = \gamma_0\right] \frac{\delta_{t'n}}{\|\delta_n\|} \\
&- \frac{f_q(\gamma_0)}{2n} v_2 t_2^2 \sum_{t=1}^{T} \sum_{t'=1}^{T} (\tau_t \wedge \tau_{t'} - \tau_t \tau_{t'}) \frac{\delta_{tn}'}{\|\delta_n\|} E\left[\mathbf{x}_i \mathbf{x}_i' | q_i = \gamma_0\right] \frac{\delta_{t'n}}{\|\delta_n\|} - \frac{1}{2n} t_{34}' \Sigma(\tau_1, \tau_2) t_{34}' + o\left(\frac{1}{n}\right),
\end{aligned}
$$

so

$$
\begin{aligned}
&E\left[\exp\left\{\sqrt{-1}\left[t_1 \sum_{i=1}^{n} S_{1i} + t_2 \sum_{i=1}^{n} S_{2i} + t_3' \sum_{i=1}^{n} S_{3i} + t_4' \sum_{i=1}^{n} S_{4i}\right]\right\}\right] \\
&= \prod_{i=1}^{n} E\left[\exp\left\{\sqrt{-1}\left[t_1 S_{1i} + t_2 S_{2i} + t_3' S_{3i} + t_4' S_{4i}\right]\right\}\right] \\
&= \exp\left\{-\frac{1}{2} t_3' \Sigma_1(\tau_1, \tau_1) t_3' - \frac{1}{2} t_4' \Sigma_2(\tau_2, \tau_2) t_4' \right. \\
&- \frac{f_q(\gamma_0)}{2} |v_1| t_1^2 \sum_{t=1}^{T} \sum_{t'=1}^{T} (\tau_t \wedge \tau_{t'} - \tau_t \tau_{t'}) \frac{\delta_{tn}'}{\|\delta_n\|} E\left[\mathbf{x}_i \mathbf{x}_i' | q_i = \gamma_0\right] \frac{\delta_{t'n}}{\|\delta_n\|} \\
&\left. - \frac{f_q(\gamma_0)}{2} v_2 t_2^2 \sum_{t=1}^{T} \sum_{t'=1}^{T} (\tau_t \wedge \tau_{t'} - \tau_t \tau_{t'}) \frac{\delta_{tn}'}{\|\delta_n\|} E\left[\mathbf{x}_i \mathbf{x}_i' | q_i = \gamma_0\right] \frac{\delta_{t'n}}{\|\delta_n\|} \right\} + o(1).
\end{aligned}
$$

As a result, $\widehat{\gamma}$, $\widehat{\beta}_1(\tau_1)$ and $\widehat{\beta}_2(\tau_2)$ are asymptotically independent. ■

Proof of Corollary 4. Take $\widehat{J}_1(\tau)$ as an example. Recall that $\widehat{J}_1(\tau) = \frac{1}{n} \sum_{i=1}^{n} K_{h_1}\left(y_i - \mathbf{x}_i' \widehat{\beta}_1(\tau)\right) \mathbf{x}_i \mathbf{x}_i' 1\left(q_i \leq \widehat{\gamma}\right)$. We will show that

$$
\widehat{J}_1(\tau) - J_1(\tau) = o_p(1) \text{ uniformly in } \tau \in \mathcal{T}.
$$

Note that $h_1 \widehat{J}_1(\tau) = P_n\left[f_i(\widehat{\beta}_1(\tau), \widehat{\gamma}, h_1)\right]$, where $f_i(\beta, \gamma, h) = K\left(\frac{y_i - \mathbf{x}_i'\beta}{h}\right) \mathbf{x}_i \mathbf{x}_i' 1\left(q_i \leq \gamma\right)$. For any compact set B, Γ and positive constant H, the functional class $\{f_i(\beta, \gamma, h), \beta \in B, \gamma \in \Gamma, \gamma \in (0, H]\}$ is a Donsker class with a square-integrable envelope by Theorem 2.10.6 in van der Vaart and Wellner (1996), because this is a product of a square-integrable random matrix $\mathbf{x}_i \mathbf{x}_i'$ (recall $E\left[\|\mathbf{x}\|^4\right] < \infty$ by assumption) and two VC classes $\left\{K\left(\frac{y_i - \mathbf{x}_i'\beta}{h}\right), \beta \in B, h \in (0, H]\right\}$ (see Example 2.10 of Pakes and Pollard (1989)) and $\{1\left(q_i \leq \gamma\right), \gamma \in \Gamma\}$. Therefore, $(\beta, \gamma, h) \mapsto \mathbb{G}_n\left[f_i(\beta, \gamma, h)\right]$ converges to a Gaussian process in $\ell^\infty\left(B \times \Gamma \times (0, H]\right)$, which implies that $\sup_{\beta \in B, \gamma \in \Gamma, \gamma \in (0, H]} \|P_n\left[f_i(\beta, \gamma, h)\right] - E\left[f_i(\beta, \gamma, h)\right]\| = O_p(n^{-1/2})$. Letting B be the parameter space of

$\beta_1(\tau)$, this implies $\sup_{\tau\in\mathcal{T}}\left\|P_n\left[f_i(\widehat{\beta}_1(\tau),\widehat{\gamma},h_1)\right]-E\left[f_i(\beta,\gamma,h_1)\right]|_{\beta=\widehat{\beta}_1(\tau),\gamma=\widehat{\gamma}}\right\|=O_p(n^{-1/2})$. Hence,

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}}\left\|\widehat{J}_1(\tau)-J_1(\tau)\right\|=\sup_{\tau\in\mathcal{T}}\left\|\frac{1}{h_1}P_n\left[f_i(\widehat{\beta}_1(\tau),\widehat{\gamma},h_1)\right]-J_1(\tau)\right\|\\
&\leq\sup_{\tau\in\mathcal{T}}\left\|\frac{1}{h_1}\left(P_n\left[f_i(\widehat{\beta}_1(\tau),\widehat{\gamma},h_1)\right]-E\left[f_i(\beta,\gamma,h_1)\right]|_{\beta=\widehat{\beta}_1(\tau),\gamma=\widehat{\gamma}}\right)\right\|\\
&+\sup_{\tau\in\mathcal{T}}\left\|\frac{1}{h_1}\left(E\left[f_i(\beta,\gamma,h_1)\right]|_{\beta=\widehat{\beta}_1(\tau),\gamma=\widehat{\gamma}}-E\left[f_i(\beta,\gamma_0,h_1)\right]|_{\beta=\widehat{\beta}_1(\tau)}\right)\right\|\\
&+\sup_{\tau\in\mathcal{T}}\left\|\frac{1}{h_1}E\left[f_i(\beta,\gamma_0,h_1)\right]\Big|_{\beta=\widehat{\beta}_1(\tau)}-J_1(\tau)\right\|=O_p\left(n^{-1/2}h_1^{-1}\right)+o_p(1),
\end{aligned}$$

where the $o_p(1)$ in the last equality is from two facts:

$$\begin{aligned}
h_1^{-1}E\left[f_i(\beta,\gamma,h_1)\right]\big|_{\beta=\widehat{\beta}_1(\tau),\gamma=\widehat{\gamma}}&=E\left[\mathbf{x}_i\mathbf{x}_i'1\left(q_i\leq\gamma\right)\int K(u)f_{y|x,q}(\mathbf{x}_i'\beta+uh|x_i,q_i)du\right]\Big|_{\beta=\widehat{\beta}_1(\tau),\gamma=\widehat{\gamma}}\\
&=E\left[\mathbf{x}_i\mathbf{x}_i'1\left(q_i\leq\gamma_0\right)\int K(u)f_{y|x,q}(\mathbf{x}_i'\beta+uh|x_i,q_i)du\right]\Big|_{\beta=\widehat{\beta}_1(\tau)}+o_p(1),
\end{aligned}$$

and

$$h_1^{-1}E\left[f_i(\beta,\gamma_0,h_1)\right]\big|_{\beta=\widehat{\beta}_1(\tau)}=E\left[\mathbf{x}_i\mathbf{x}_i'1\left(q_i\leq\gamma_0\right)\int K(u)f_{y|x,q}(\mathbf{x}_i'\beta+uh|x_i,q_i)du\right]\Big|_{\beta=\widehat{\beta}_1(\tau)}=J_1(\tau)+o_p(1)$$

by the assumptions on $K(\cdot)$ and $f(y|x,q)$; see (A.55) in Pagan and Ullah (1999). By $nh_1^2\to\infty$, the result follows. ■

Proof of Theorem 5. This proof is based on $\widetilde{T}_n(\gamma,\tau)$; the proof for $\widehat{T}_n(\gamma,\tau)$ is easier.

First, $\widehat{\beta}(\tau)$ is uniformly consistent to $\beta_0(\tau)$ for $\tau\in\mathcal{T}$. The proof is similar to Appendix A.1.1 of Angrist et al. (2006). Given Assumption T1, we need only show that $Q_n(\tau,\beta)\equiv P_n\left[\rho_\tau\left(y-\mathbf{x}'\beta\right)-\rho_\tau\left(y-\mathbf{x}'\beta_0(\tau)\right)\right]$ converges to $Q_\infty(\tau,\beta)\equiv E\left[\rho_\tau\left(Y-\mathbf{x}'\beta\right)-\rho_\tau\left(e_\tau\right)\right]$ which is uniquely minimized at $\beta_0(\tau)$, where $Y=\mathbf{x}'\beta_0(\tau)+e_\tau$. For this purpose, we need only to show that

$$Q_n(\tau,\beta)-P_n\left[\rho_\tau\left(Y-\mathbf{x}'\beta\right)-\rho_\tau\left(e_\tau\right)\right]=o_p(1)$$

uniformly over $(\tau,\beta)\in\mathcal{T}\times B$. Note that

$$\left|Q_n(\tau,\beta)-P_n\left[\rho_\tau\left(Y-\mathbf{x}'\beta\right)-\rho_\tau\left(e_\tau\right)\right]\right|\leq 3n^{-1/2}\sup_{i\leq n}\left|\mathbf{x}_i'c(\tau)\right|1(q_i\leq\gamma_0)=o_p(1),$$

given that $E[\|\mathbf{x}\|^4]<\infty$, and $c(\tau)$ is uniformly bounded on $\tau\in\mathcal{T}$.

Second, $n^{-1}\sum_{i=1}^n\varphi_\tau\left(y_i-\mathbf{x}_i'\widehat{\beta}(\tau)\right)^2\left(\mathbf{x}_i(\gamma)-\widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right)\left(\mathbf{x}_i(\gamma)-\widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right)'$ converges to $H(\gamma,\tau)$ uniformly in $(\tau,\gamma)\in\mathcal{T}\times\Gamma$. From the proof of Corollary 4, $\widehat{J}(\gamma,\tau)$ and $\widehat{J}(\tau)$ are uniformly consistent to $J(\gamma,\tau)$ and $J(\tau)$, respectively, so we need only to show that $n^{-1}\sum_{i=1}^n g_i\left(\gamma,\tau,\widehat{\beta}(\tau)\right)$ converges to $H(\gamma,\tau)$ uniformly, where

$$g_i(\gamma,\tau,\beta)=\varphi_\tau\left(y_i-\mathbf{x}_i'\beta\right)^2\left(\mathbf{x}_i(\gamma)-J(\gamma,\tau)J(\tau)^{-1}\mathbf{x}_i\right)\left(\mathbf{x}_i(\gamma)-J(\gamma,\tau)J(\tau)^{-1}\mathbf{x}_i'\right).$$

It is easy to verify that $\{g_i(\gamma,\tau,\beta),(\tau,\beta,\gamma)\in\mathcal{T}\times B\times\Gamma\}$ is Donsker, and hence a Glivenko-Cantelli class, e.g., using Theorem 2.10.6 in van der Vaart and Wellner (1996). This implies that $P_n[g_i(\gamma,\tau,\beta)]-E[g_i(\gamma,\tau,\beta)]=$

$o_p(1)$ uniformly in $(\tau,\beta,\gamma)\in\mathcal{T}\times B\times\Gamma$. This latter and continuity of $E[g_i(\gamma,\tau,\beta)]$ in (τ,β,γ), combined with the uniformly consistency of $\widehat{\beta}(\tau)$ and the definition of y_i, imply the result.

Third, $n^{-1/2}\sum_{i=1}^n\left[\mathbf{x}_i(\gamma)-\widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right]\varphi_\tau\left(y_i-\mathbf{x}_i'\widehat{\beta}(\tau)\right)\rightsquigarrow S(\gamma,\tau)+\left[J(\gamma,\tau)J(\tau)^{-1}J(\gamma_0,\tau)-J(\gamma\wedge\gamma_0,\tau)\right]c(\tau)$ in $\ell^\infty(\mathcal{T}\times\Gamma)$.

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\mathbf{x}_i(\gamma)\varphi_\tau\left(y_i-\mathbf{x}_i'\widehat{\beta}(\tau)\right)=\mathbb{G}_n\left(\mathbf{x}_i(\gamma)\varphi_\tau\left(y_i-\mathbf{x}_i'\widehat{\beta}(\tau)\right)\right)+\sqrt{n}\,E\left[\mathbf{x}_i(\gamma)\varphi_\tau\left(y_i-\mathbf{x}_i'\beta\right)\right]\big|_{\beta=\widehat{\beta}(\tau)}.$$

By Assumption T2 and stochastic equicontinuity of $(\tau,\beta,\gamma)\mapsto\mathbb{G}_n\left(\mathbf{x}(\gamma)\varphi_\tau\left(y-\mathbf{x}'\beta\right)\right)$, the first term on the right hand side $\mathbb{G}_n\left(\mathbf{x}_i(\gamma)\varphi_\tau\left(y_i-\mathbf{x}_i'\widehat{\beta}(\tau)\right)\right)=\mathbb{G}_n\left(\mathbf{x}_i(\gamma)\varphi_\tau\left(e_{\tau i}\right)\right)+o_p(1)$ in $\ell^\infty(\mathcal{T}\times\Gamma)$. By Taylor expansion, the second term

$$\begin{aligned}
&\sqrt{n}\,E\left[\mathbf{x}_i(\gamma)\varphi_\tau\left(y_i-\mathbf{x}_i'\beta\right)\right]\big|_{\beta=\widehat{\beta}(\tau)}\\
&=\sqrt{n}\,E\left[\mathbf{x}_i(\gamma)\varphi_\tau\left(\mathbf{x}_i'\beta_0(\tau)+n^{-1/2}\mathbf{x}_i(\gamma_0)'c(\tau)+e_{\tau i}-\mathbf{x}_i'\beta\right)\right]\Big|_{\beta=\widehat{\beta}(\tau)}\\
&=J(\gamma,\tau)\sqrt{n}\left(\widehat{\beta}(\tau)-\beta_0(\tau)\right)-J(\gamma\wedge\gamma_0,\tau)c(\tau)+o_p(1).
\end{aligned}$$

From the proof of Corollary 4, $\widehat{J}(\gamma,\tau)$ and $\widehat{J}(\tau)$ are uniformly consistent to $J(\gamma,\tau)$ and $J(\tau)$, respectively, so

$$\begin{aligned}
&\widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\mathbf{x}_i\varphi_\tau\left(y_i-\mathbf{x}_i'\widehat{\beta}(\tau)\right)\\
&=\left(J(\gamma,\tau)J(\tau)^{-1}+o_p(1)\right)\left\{\mathbb{G}_n\left(\mathbf{x}_i\varphi_\tau\left(e_{\tau i}\right)\right)-J(\gamma_0,\tau)c(\tau)+J(\tau)\sqrt{n}\left(\widehat{\beta}(\tau)-\beta_0(\tau)\right)+o_p(1)\right\}\\
&=J(\gamma,\tau)J(\tau)^{-1}\mathbb{G}_n\left(\mathbf{x}_i\varphi_\tau\left(e_{\tau i}\right)\right)+J(\gamma,\tau)\sqrt{n}\left(\widehat{\beta}(\tau)-\beta_0(\tau)\right)-J(\gamma,\tau)J(\tau)^{-1}J(\gamma_0,\tau)c(\tau)+o_p(1)
\end{aligned}$$

in $\ell^\infty(\mathcal{T}\times\Gamma)$. As a result,

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^n\left[\mathbf{x}_i(\gamma)-\widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right]\varphi_\tau\left(y_i-\mathbf{x}_i'\widehat{\beta}(\tau)\right)\\
&=\mathbb{G}_n\left(\mathbf{x}_i(\gamma)\varphi_\tau\left(e_{\tau i}\right)\right)-J(\gamma,\tau)J(\tau)^{-1}\mathbb{G}_n\left(\mathbf{x}_i\varphi_\tau\left(e_{\tau i}\right)\right)-\left[J(\gamma\wedge\gamma_0,\tau)-J(\gamma,\tau)J(\tau)^{-1}J(\gamma_0,\tau)\right]c(\tau),
\end{aligned}$$

where $\mathbb{G}_n\left(\mathbf{x}_i(\gamma)\varphi_\tau\left(e_{\tau i}\right)\right)-J(\gamma,\tau)J(\tau)^{-1}\mathbb{G}_n\left(\mathbf{x}_i\varphi_\tau\left(e_{\tau i}\right)\right)$ converges weakly to $S(\gamma,\tau)$ in $\ell^\infty(\mathcal{T}\times\Gamma)$. ■

Proof of Theorem 6. First, conditional on the original sample path, $n^{-1/2}\sum_{i=1}^n\left[\mathbf{x}_i(\gamma)-\widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right]\varphi_\tau(y_i-\mathbf{x}_i'\widehat{\beta}(\tau))\xi_i^*$ is a zero-mean Gaussian process with covariance function

$$\begin{aligned}
&H_n((\gamma_1,\tau_1),(\gamma_2,\tau_2))\\
&=n^{-1}\sum_{i=1}^n\varphi_{\tau_1}\left(y_i-\mathbf{x}_i'\widehat{\beta}(\tau_1)\right)\varphi_{\tau_2}\left(y_i-\mathbf{x}_i'\widehat{\beta}(\tau_2)\right)\left(\mathbf{x}_i(\gamma_1)-\widehat{J}(\gamma_1,\tau_1)\widehat{J}(\tau_1)^{-1}\mathbf{x}_i\right)\left(\mathbf{x}_i(\gamma_2)-\widehat{J}(\gamma_2,\tau_2)\widehat{J}(\tau_2)^{-1}\mathbf{x}_i\right)'.
\end{aligned}$$

Extending the second step in the proof of Theorem 5, we have $H_n((\gamma_1,\tau_1),(\gamma_2,\tau_2))\xrightarrow{p}H((\gamma_1,\tau_1),(\gamma_2,\tau_2))$ uniformly over $(\tau_1,\tau_2,\gamma_1,\gamma_2)\in\mathcal{T}\times\mathcal{T}\times\Gamma\times\Gamma$. Second, also by the second step in the proof of Theorem 5, $n^{-1}\sum_{i=1}^n\varphi_\tau\left(y_i-\mathbf{x}_i'\widehat{\beta}(\tau)\right)^2\left(\mathbf{x}_i(\gamma)-\widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right)\left(\mathbf{x}_i(\gamma)-\widehat{J}(\gamma,\tau)\widehat{J}(\tau)^{-1}\mathbf{x}_i\right)'\xrightarrow{p}H(\gamma,\tau)$ uniformly over $(\tau,\gamma)\in\mathcal{T}\times\Gamma$. In summary, $\widehat{T}_n^*(\gamma,\tau)\overset{*}{\rightsquigarrow}H(\gamma,\tau)^{-1/2}S(\gamma,\tau)=T^0(\gamma,\tau)$ in $\ell^\infty(\mathcal{T}\times\Gamma)$, where $\overset{*}{\rightsquigarrow}$ signifies the weak convergence in probability. ■

Appendix B: Lemmas

Lemma 1 *Under Assumption D,* $\widehat{\theta} \xrightarrow{p} \theta_0$.

Proof. Theorem 2.1 of Newey and McFadden (1994) is used in this proof. The objective function is

$$Q_n(\gamma, \beta_T) = \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T} \rho_{\tau_t} \left(y_i - \mathbf{x}_i' \beta_1 (\tau_t) 1(q_i \leq \gamma) - \mathbf{x}_i' \beta_2 (\tau_t) 1(q_i > \gamma) \right).$$

It is convenient to consider the recentered version of $Q_n(\gamma, \beta_T)$:

$$S_n(\gamma, \beta_T) = Q_n(\gamma, \beta_T) - Q_n(\gamma_0, \beta_T^0).$$

We need only show that $\sup_{\theta \in \Theta} |S_n(\gamma, \beta_T) - S(\gamma, \beta_T)| \xrightarrow{p} 0$, where

$$S(\gamma, \beta_T) = \sum\nolimits_{t=1}^{T} E \left[\rho_{\tau_t} \left(y_i - \mathbf{x}_i' \beta_1 (\tau_t) 1(q_i \leq \gamma) + \mathbf{x}_i' \beta_2 (\tau_t) 1(q_i > \gamma) \right) - \rho_{\tau_t} \left(e_{1\tau_t} 1(q_i \leq \gamma_0) + e_{2\tau_t} 1(q_i > \gamma_0) \right) \right]$$

is continuous in θ and is uniquely minimized at θ_0.

Step 1: $\sup_{\theta \in \Theta} |S_n(\gamma, \beta_T) - S(\gamma, \beta_T)| \xrightarrow{p} 0$. We apply Lemma 2.8 of Pakes and Pollard (1989) to prove this result. So we need to check the class of functions $\left\{ \sum_{t=1}^{T} \rho_{\tau_t} \left(y - \mathbf{x}' \beta_1 (\tau_t) 1(q \leq \gamma) - \mathbf{x}' \beta_2 (\tau_t) 1(q > \gamma) \right), \theta \in \Theta \right\}$ is Euclidean with an envelope that has a finite first moment.

$$\begin{aligned}
&\sum\nolimits_{t=1}^{T} \left[\rho_{\tau_t} \left(y - \mathbf{x}' \beta_1 (\tau_t) 1(q \leq \gamma) + \mathbf{x}' \beta_2 (\tau_t) 1(q > \gamma) \right) - \rho_{\tau_t} \left(e_{1\tau_t} 1(q \leq \gamma_0) + e_{2\tau_t} 1(q > \gamma_0) \right) \right] \\
&= 1(q \leq \gamma_0 \wedge \gamma) \sum\nolimits_{t=1}^{T} \left[\rho_{\tau_t} \left(e_{1\tau_t} + \mathbf{x}' \beta_1^0 (\tau_t) - \mathbf{x}' \beta_1 (\tau_t) \right) - \rho_{\tau_t} \left(e_{1\tau_t} \right) \right] \\
&+ 1 (q > \gamma \vee \gamma_0) \sum\nolimits_{t=1}^{T} \left[\rho_{\tau_t} \left(e_{2\tau_t} + \mathbf{x}' \beta_2^0 (\tau_t) - \mathbf{x}' \beta_2 (\tau_t) \right) - \rho_{\tau_t} \left(e_{2\tau_t} \right) \right] \\
&+ 1 (\gamma \wedge \gamma_0 < q \leq \gamma_0) \sum\nolimits_{t=1}^{T} \left[\rho_{\tau_t} \left(e_{1\tau_t} + \mathbf{x}' \beta_1^0 (\tau_t) - \mathbf{x}' \beta_2 (\tau_t) \right) - \rho_{\tau_t} \left(e_{1\tau_t} \right) \right] \\
&+ 1 (\gamma_0 < q \leq \gamma \vee \gamma_0) \sum\nolimits_{t=1}^{T} \left[\rho_{\tau_t} \left(e_{2\tau_t} + \mathbf{x}' \beta_2^0 (\tau_t) - \mathbf{x}' \beta_1 (\tau_t) \right) - \rho_{\tau_t} \left(e_{2\tau_t} \right) \right] \\
&\equiv \sum\nolimits_{t=1}^{T} A_{\tau_t} \left(\mathsf{w}_{1\tau_t} | \theta \right) + \sum\nolimits_{t=1}^{T} B_{\tau_t} \left(\mathsf{w}_{2\tau_t} | \theta \right) + \sum\nolimits_{t=1}^{T} C_{\tau_t} \left(\mathsf{w}_{1\tau_t} | \theta \right) + \sum\nolimits_{t=1}^{T} D_{\tau_t} \left(\mathsf{w}_{2\tau_t} | \theta \right),
\end{aligned}$$

where $\mathsf{w}_{\ell\tau} = (e_{\ell\tau}, \mathbf{x}')'$. $\{1 (q \leq \gamma \wedge \gamma_0), \gamma \in \Gamma\}$ is Euclidean with envelope 1 by Lemma 2.4 of Pakes and Pollard (1989). $\sum_{t=1}^{T} [\rho_{\tau_t} (e_{1\tau_t} + \mathbf{x}' \beta_1^0 (\tau_t) - \mathbf{x}_i' \beta_1 (\tau_t)) - \rho_{\tau_t} (e_{1\tau_t})]$ is Lipschitz by the following arguments:

$$\sum\nolimits_{t=1}^{T} \rho_{\tau_t} \left(e_{1\tau_t} + \mathbf{x}' \beta_1^0 (\tau_t) - \mathbf{x}' \beta_1 (\tau_t) \right) - \sum\nolimits_{t=1}^{T} \rho_{\tau_t} \left(e_{1\tau_t} + \mathbf{x}' \beta_1^0 (\tau_t) - \mathbf{x}' \widetilde{\beta}_1 (\tau_t) \right) \leq 2 \|\mathbf{x}\| \sum\nolimits_{t=1}^{T} \left\| \beta_1 (\tau_t) - \widetilde{\beta}_1 (\tau_t) \right\|,$$

where $E[\|\mathbf{x}\|] < \infty$ by Assumption D7. By Lemma 2.13 of Pakes and Pollard (1989), $\{\sum_{t=1}^{T} [\rho_{\tau_t} (e_{1\tau_t} + \mathbf{x}' \beta_1^0 (\tau_t) - \mathbf{x}_i' \beta_1 (\tau_t)) - \rho_{\tau_t} (e_{1\tau_t})], \beta_1 (\tau_t) \in B, t = 1, \cdots T\}$ is Euclidean with the envelope $C \|\mathbf{x}\|$. So $\{\sum_{t=1}^{T} A_{\tau_t} (\mathsf{w}_{1\tau_t} | \theta), \gamma \in \Gamma, \beta_{1\tau_t} \in B, t = 1, \cdots T\}$ is Euclidean with envelope $C \|\mathbf{x}\|$ by Lemma 2.14 (ii) of Pakes and Pollard (1989). We can show other terms are Euclidean by similar arguments.

Step 2: $S(\gamma, \beta_T)$ is continuous in θ and is uniquely maximized at θ_0. This continuity of $S(\gamma, \beta_T)$ is obvious given that $f(q)$ is bounded on Γ. To show $S(\gamma, \beta_T)$ is uniquely minimized at θ_0, we consider four cases. (i) $\gamma = \gamma_0$, $\beta_T \neq \beta_T^0$. From standard arguments in quantile regression, Assumption D6 guarantees

that $S(\gamma_0, \beta_T)$ is uniquely minimized at β_T^0. (ii) $\gamma \neq \gamma_0$ (say $\gamma < \gamma_0$), $\beta_T = \beta_T^0$.

$$S(\gamma, \beta_T^0) - S(\gamma_0, \beta_T^0) = \int_{\gamma}^{\gamma_0} \left[\sum_{t=1}^{T} \int E \left[\rho_{\tau_t} \left(e_{1\tau_t} + \mathbf{x}' \beta_1^0 (\tau_t) - \mathbf{x}' \beta_2^0 (\tau_t) \right) - \rho_{\tau_t} (e_{1\tau_t}) \, | x, q \right] dF(x|q) \right] f(q) dq.$$

Assumptions D2 and D6 guarantee that $\sum_{t=1}^{T} \int E \left[\rho_{\tau_t} (e_{1\tau_t} + \mathbf{x}' \beta_1^0 (\tau_t) - \mathbf{x}' \beta_2^0 (\tau_t)) - \rho_{\tau_t} (e_{1\tau_t}) \, | x, q \right] dF(x|q)$ is strictly greater than zero. Given that $f(q)$ is greater than zero on Γ, $S(\gamma, \beta_T^0) - S(\gamma_0, \beta_T^0) > 0$ if $\gamma \neq \gamma_0$. (iii) $\gamma > \gamma_0$, $\beta_T \neq \beta_T^0$, $\beta_{1T} = \beta_{2T}^0$ or $\gamma < \gamma_0$, $\beta_T \neq \beta_T^0$, $\beta_{2T} = \beta_{1T}^0$. Take the former case as an example,

$$S(\gamma, \beta_T) - S(\gamma_0, \beta_T^0) \geq \int_{-\infty}^{\gamma} \left[\sum_{t=1}^{T} \int E \left[\rho_{\tau_t} \left(e_{1\tau_t} + \mathbf{x}' \beta_1^0 (\tau_t) - \mathbf{x}' \beta_2^0 (\tau_t) \right) - \rho_{\tau_t} (e_{1\tau_t}) \, | x, q \right] dF(x|q) \right] f(q) dq.$$

Given Assumptions D2, D4 and D6, $S(\gamma, \beta_T) - S(\gamma_0, \beta_T^0) > 0$ if $\gamma > \gamma_0$. (iv) $\gamma > \gamma_0$, $\beta_{1T} \neq \beta_{2T}^0$ or $\gamma < \gamma_0$, $\beta_{2T} \neq \beta_{1T}^0$. Similar arguments as in Case (ii) lead to $S(\gamma, \beta_T) - S(\gamma_0, \beta_T^0) > 0$. ■

Remark 2 *This proof cannot be extend to the objective function* $\frac{1}{|\mathcal{T}|} \int_{\mathcal{T}} Q_{\tau n} (\theta) \, d\tau$ *with* $\theta = (\gamma', \beta' (\cdot))'$. *This is because* $\widehat{\beta} (\cdot)$ *can be any discontinuous function on* $\mathcal{T}$ *such that the parameter space for* $\beta (\cdot)$ *is not compact. If we impose some smoothness assumptions on* $\beta (\cdot)$, *it is quite possible to prove the consistency of* $\widehat{\gamma}$ *under such an objective function. However, this is not how* $\widehat{\gamma}$ *is defined.*

Lemma 2 *Under Assumptions D1-D7 and* $\|\delta_n\| \to 0$, $\sqrt{n} \|\delta_n\| \to \infty$, $\widehat{\beta}_{\ell T} - \beta_{\ell T}^0 = o_p(\|\delta_n\|)$, *and* $\widehat{\gamma} - \gamma_0 = o_p(1)$.

Proof. We use the notations in the last lemma to prove this result. Consider the case of $\gamma \geq \gamma_0$ without loss of generality because of symmetry. By Step 1 of the last lemma, and $\|\delta_n\| \to 0$,

$$\sup_{\theta \in \Theta} |S_n(\gamma, \beta_T) - S(\gamma, \beta_T)| \xrightarrow{p} 0,$$

where $S(\gamma, \beta_T)$ is redefined as

$$\begin{aligned}
&E \left[\sum_{t=1}^{T} \left[\rho_{\tau_t} \left(e_{1\tau_t} + \mathbf{x}' \beta_1^0 (\tau_t) - \mathbf{x}' \beta_1 (\tau_t) \right) - \rho_{\tau_t} (e_{1\tau_t}) \right] 1(q \leq \gamma_0) \right] \\
&+ E \left[\sum_{t=1}^{T} \left[\rho_{\tau_t} \left(e_{2\tau_t} + \mathbf{x}' \beta_2^0 (\tau_t) - \mathbf{x}' \beta_2 (\tau_t) \right) - \rho_{\tau_t} (e_{2\tau_t}) \right] 1 (q > \gamma) \right] \\
&+ E \left[\sum_{t=1}^{T} \left[\rho_{\tau_t} \left(e_{2\tau_t} + \mathbf{x}' \beta_1^0 (\tau_t) - \mathbf{x}' \beta_1 (\tau_t) \right) - \rho_{\tau_t} (e_{2\tau_t}) \right] 1 (\gamma_0 < q \leq \gamma) \right].
\end{aligned}$$

From Assumption D6, $S(\gamma, \beta_T)$ is uniquely minimized at β_T^0 for any $\gamma \in \Gamma$, so by Theorem 2.1 of Newey and McFadden (1994), $\widehat{\beta}$ is consistent for any $\gamma \in \Gamma$. However, $S(\gamma, \beta_T)$ is not uniquely minimized at θ_0. For example, $S(\gamma, \beta_T^0) = 0$ for any $\gamma \in \Gamma$. To prove the consistency of $\widehat{\gamma}$, the normalization in $Q_n(\gamma, \beta_T)$ should be a_n^{-1} rather than n^{-1}. We denote the objective function still as $Q_n(\gamma, \beta_T)$. Also, without loss of generality, the parameter space can be restricted as $\left\{ \left\| \beta_T - \beta_T^0 \right\| \leq \epsilon, \, \gamma \in \Gamma \right\}$ for a small positive number ϵ.

Write $Q_n(\gamma, \beta_T)$ as

$$
\begin{aligned}
& Q_n(\gamma, \beta_T) \qquad (17)\\
&= \frac{1}{a_n}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\rho_{\tau_t}\left(e_{1\tau_t i}+\mathbf{x}_i'\beta_1^0(\tau_t)-\mathbf{x}_i'\beta_1(\tau_t)\right)-\rho_{\tau_t}\left(e_{1\tau_t i}\right)\right]1(q_i\le\gamma_0)\\
&+\frac{1}{a_n}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\rho_{\tau_t}\left(e_{2\tau_t i}+\mathbf{x}_i'\beta_2^0(\tau_t)-\mathbf{x}_i'\beta_2(\tau_t)\right)-\rho_{\tau_t}\left(e_{2\tau_t i}\right)\right]1(q_i>\gamma)\\
&+\frac{1}{a_n}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\rho_{\tau_t}\left(e_{2\tau_t i}+\mathbf{x}_i'\beta_2^0(\tau_t)-\mathbf{x}_i'\beta_1(\tau_t)\right)-\rho_{\tau_t}\left(e_{2\tau_t i}\right)\right]1(\gamma_0<q_i\le\gamma)\\
&\equiv T_1(\theta)+T_2(\theta)+T_3(\theta).
\end{aligned}
$$

$T_1(\theta)$ and $T_2(\theta)$ can be similarly analyzed, so take $T_1(\theta)$ as an example. From Knight (1998),

$$
\begin{aligned}
&\rho_{\tau_t}\left(e_{1\tau_t i}+\mathbf{x}_i'\beta_1^0(\tau_t)-\mathbf{x}_i'\beta_1(\tau_t)\right)-\rho_{\tau_t}\left(e_{1\tau_t i}\right)\\
&=\varphi_{\tau_t}\left(e_{1\tau_t i}\right)\mathbf{x}_i'\left(\beta_1(\tau_t)-\beta_1^0(\tau_t)\right)+\int_0^{\mathbf{x}_i'\left(\beta_1(\tau_t)-\beta_1^0(\tau_t)\right)}\left(1(e_{1\tau_t i}<s)-1(e_{1\tau_t i}<0)\right)ds.
\end{aligned}
$$

Note that

$$
n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\varphi_{\tau_t}\left(e_{1\tau_t i}\right)\mathbf{x}_i'1(q_i\le\gamma_0)=O_p(n^{-1/2}),
$$

and

$$
\begin{aligned}
&\tfrac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\int_0^{\mathbf{x}_i'\left(\beta_1(\tau_t)-\beta_1^0(\tau_t)\right)}\left(1(e_{1\tau_t i}<s)-1(e_{1\tau_t i}<0)\right)ds1(q_i\le\gamma_0)\\
&=\sum_{t=1}^{T}O_p\left(\left(\beta_1(\tau_t)-\beta_1^0(\tau_t)\right)'J_{\tau_t}\left(\beta_1(\tau_t)-\beta_1^0(\tau_t)\right)\right)=O_p\left(\left\|\beta_{1T}-\beta_{1T}^0\right\|^2\right)
\end{aligned}
$$

uniformly for $\left\|\beta_{1T}-\beta_{1T}^0\right\|\le\epsilon$, where J_{τ_t}, $t=1,\cdots,T$, is positive definite from Assumption D6. As a result,

$$
T_1(\theta)=O_p(n^{-1/2}\left\|\beta_{1T}-\beta_{1T}^0\right\|/\delta_n'\delta_n)+O_p(\left\|\beta_{1T}-\beta_{1T}^0\right\|^2/\delta_n'\delta_n),
$$

and the second part of $T_1(\theta)$ dominates on $\left\|\beta_{1T}-\beta_{1T}^0\right\|\ge M\left\|\delta_n\right\|$ for any $M>0$ given that $n^{1/2}\left\|\delta_n\right\|\to\infty$. So for any $M>0$, we can find a constant $C_M>0$ such that

$$
P\left(\inf_{\left\|\beta_{1T}-\beta_{1T}^0\right\|\ge M\left\|\delta_n\right\|}T_1(\theta)>C_M\right)\to 1.
$$

Similar results apply to $T_2(\theta)$. As to $T_3(\theta)$, by a similar analysis as in $T_1(\theta)$, we can show

$$
T_3(\theta)=O_p(n^{-1/2}\left\|\beta_{1T}-\beta_{2T}^0\right\|/\delta_n'\delta_n)+O_p(\left\|\beta_{1T}-\beta_{2T}^0\right\|^2/\delta_n'\delta_n)
$$

uniformly for $\left\|\beta_{1T}-\beta_{1T}^0\right\|\le\epsilon$, and $\gamma-\gamma_0\ge\epsilon$. $\left|\left\|\beta_{1T}^0-\beta_{2T}^0\right\|-\left\|\beta_{1T}-\beta_{1T}^0\right\|\right|\le\left\|\beta_{1T}-\beta_{2T}^0\right\|\le\left\|\beta_{1T}-\beta_{1T}^0\right\|+\left\|\beta_{1T}^0-\beta_{2T}^0\right\|$. Given that $\left\|\beta_{1T}^0-\beta_{2T}^0\right\|=\left\|\delta_n\right\|$, for any $M>0$, $\left\|\beta_{1T}-\beta_{2T}^0\right\|$ is either $O(\left\|\delta_n\right\|)$ or $o(\left\|\delta_n\right\|)$ when $\left\|\beta_{1T}-\beta_{1T}^0\right\|\ge M\left\|\delta_n\right\|$. In the former case, for any $\epsilon>0$ and $M>0$, we can find a constant $C_{M\epsilon}>0$ such that with probability approaching 1, $\inf_{\left\|\beta_{1T}-\beta_{1T}^0\right\|\ge M\left\|\delta_n\right\|,\gamma_0+\epsilon\le\gamma\le\overline{\gamma}}T_3(\theta)>C_{M\epsilon}$; in the later case, $T_3(\theta)$ is dominated by $T_1(\theta)$ or $T_2(\theta)$.

The above arguments can be applied to $\gamma < \gamma_0$, so in summary, for any $\epsilon > 0$ and $M > 0$, we can find a constant $C_{M\epsilon} > 0$ such that

$$P\left(\inf_{\left\|\beta_{\ell T}-\beta^0_{\ell T}\right\|\geq M\|\delta_n\|,|\gamma-\gamma_0|\geq\epsilon} Q_n(\gamma,\beta_T) > C_{M\epsilon}\right) \to 1,$$

which implies the results of interest. ■

Lemma 3 *Under Assumption D,* $n\left(\widehat{\gamma}-\gamma_0\right) = O_p(1)$, *and* $\sqrt{n}\left(\beta_T - \beta^0_T\right) = O_p(1)$.

Proof. This proof uses Corollary 3.2.6 of van der Vaart and Wellner (1996). First define A_τ, B_τ, C_τ and D_τ as in Lemma 1.

First, $Q(\theta) - Q(\theta_0) \geq Cd^2(\theta,\theta_0)$, where $Q(\theta)$ is the probability limit of $Q_n(\theta)$, $d(\theta,\theta_0) = \left\|\beta_T - \beta^0_T\right\| + \sqrt{|\gamma-\gamma_0|}$ for $\theta \in \mathcal{N}$ with $\mathcal{N}$ being an open neighborhood of θ_0.

$$\begin{aligned}
&Q(\theta) - Q(\theta_0)\\
&\overset{(1)}{=} \sum_{t=1}^T E\left[A_{\tau_t}\left(\mathsf{w}_{1\tau_t}|\theta\right) + B_{\tau_t}\left(\mathsf{w}_{2\tau_t}|\theta\right) + C_{\tau_t}\left(\mathsf{w}_{1\tau_t}|\theta\right) + D_{\tau_t}\left(\mathsf{w}_{2\tau_t}|\theta\right)\right]\\
&\overset{(2)}{=} \sum_{t=1}^T E\left[\left(\rho_{\tau_t}\left(e_{1\tau_t} + \mathbf{x}'\left(\beta^0_1(\tau_t) - \beta_1(\tau_t)\right)\right) - \rho_{\tau_t}\left(e_{1\tau_t}\right)\right)1\left(q \leq \gamma\wedge\gamma_0\right)\right]\\
&+ \sum_{t=1}^T E\left[\left(\rho_{\tau_t}\left(\mathbf{x}'\left(\beta^0_2(\tau_t) - \beta_2(\tau_t)\right) + e_{2\tau_t}\right) - \rho_{\tau_t}\left(e_{2\tau_t}\right)\right)1(q > \gamma\vee\gamma_0)\right]\\
&+ \sum_{t=1}^T E\left[\left(\rho_{\tau_t}\left(e_{1\tau_t} + \mathbf{x}'\left(\beta^0_1(\tau_t) - \beta_2(\tau_t)\right)\right) - \rho_{\tau_t}\left(e_{1\tau_t}\right)\right)1(\gamma\wedge\gamma_0 < q \leq \gamma_0)\right]\\
&+ \sum_{t=1}^T E\left[\left(\rho_{\tau_t}\left(e_{2\tau_t} + \mathbf{x}'\left(\beta^0_2(\tau_t) - \beta_1(\tau_t)\right)\right) - \rho_{\tau_t}\left(e_{2\tau_t}\right)\right)1(\gamma_0 < q \leq \gamma\vee\gamma_0)\right]\\
&\overset{(3)}{\geq} \sum_{t=1}^T \left(\beta_1(\tau_t) - \beta^0_1(\tau_t)\right)' E[f_{e_{1\tau_t}|x,q}(0|x,q)\mathbf{x}\mathbf{x}'1\left(q \leq \gamma\wedge\gamma_0\right)]\left(\beta_1(\tau_t) - \beta^0_1(\tau_t)\right)\\
&+ \sum_{t=1}^T \left(\beta_2(\tau_t) - \beta^0_2(\tau_t)\right)' E[f_{e_{2\tau_t}|x,q}(0|x,q)\mathbf{x}\mathbf{x}'1\left(q > \gamma\vee\gamma_0\right))]\left(\beta_2(\tau_t) - \beta^0_2(\tau_t)\right)\\
&+ \sum_{t=1}^T \left(\beta_2(\tau_t) - \beta^0_1(\tau_t)\right)' E[f_{e_{1\tau_t}|x,q}(0|x,q)\mathbf{x}\mathbf{x}'1\left(\gamma\wedge\gamma_0 < q \leq \gamma_0\right)]\left(\beta_2(\tau_t) - \beta^0_1(\tau_t)\right)\\
&+ \sum_{t=1}^T \left(\beta_1(\tau_t) - \beta^0_2(\tau_t)\right)' E[f_{e_{2\tau_t}|x,q}(0|x,q)\mathbf{x}\mathbf{x}'1\left(\gamma_0 < q \leq \gamma\vee\gamma_0\right)]\left(\beta_1(\tau_t) - \beta^0_2(\tau_t)\right)\\
&\overset{(4)}{\geq} C\left\{\sum_{t=1}^T\left[\left\|\beta_1(\tau_t) - \beta^0_1(\tau_t)\right\|^2 + \left\|\beta_2(\tau_t) - \beta^0_2(\tau_t)\right\|^2\right] + |\gamma-\gamma_0|\right\} = Cd^2(\theta,\theta_0),
\end{aligned}$$

where (1) and (2) are straightforward, and (3) is from the convexity of $E\left[\rho_{\tau_t}(\cdot)\right]$. The first part of (4) is from Assumptions D4 and D6, and the second part is from Assumptions D2, D4, D5 and D6.

Second, $E\left[\sup_{d(\theta,\theta_0)<\delta}\left|\mathbb{G}_n\left(m\left(w|\theta\right) - m\left(w|\theta_0\right)\right)\right|\right] \leq C\delta$ for any sufficiently small δ. $\{A_\tau\left(\mathsf{w}_{1\tau}|\theta\right) : d(\theta,\theta_0) < \delta\}$ is a VC subgraph class. This is because

$$\begin{aligned}
&A_\tau\left(\mathsf{w}_{1\tau}|\theta\right) = \left[\rho_\tau\left(e_{1\tau} + \mathbf{x}'\left(\beta^0_1(\tau) - \beta_1(\tau)\right)\right) - \rho_\tau\left(e_{1\tau}\right)\right]1\left(q \leq \gamma\wedge\gamma_0\right)\\
&= \tau\left[\left(e_{1\tau} + \mathbf{x}'\left(\beta^0_1(\tau) - \beta_1(\tau)\right)\right)1\left(y > \mathbf{x}'\beta_1(\tau)\right) - e_{1\tau}1\left(y > \mathbf{x}'\beta^0_1(\tau)\right)\right]1\left(q \leq \gamma\wedge\gamma_0\right)\\
&+ (\tau-1)\left[\left(e_{1\tau} + \mathbf{x}'\left(\beta^0_1(\tau) - \beta_1(\tau)\right)\right)1\left(y \leq \mathbf{x}'\beta_1(\tau)\right) - e_{1\tau}1\left(y \leq \mathbf{x}'\beta^0_1(\tau)\right)\right]1\left(q \leq \gamma\wedge\gamma_0\right),
\end{aligned}$$

where $\left\{\mathbf{x}'\left(\beta^0_1(\tau) - \beta_1(\tau)\right) : d(\theta,\theta_0) < \delta\right\}$ is VC subgraph from Example 2.9 of Pakes and Pollard (1989), and $\left\{1(q \leq \gamma\wedge\gamma_0) : d(\theta,\theta_0) < \delta\right\}$, $\left\{1(y > \mathbf{x}'\beta_1(\tau)) : d(\theta,\theta_0) < \delta\right\}$ and $\left\{1(y \leq \mathbf{x}'\beta_1(\tau)) : d(\theta,\theta_0) < \delta\right\}$ are VC subgraph from Lemma 2.4 of Pakes and Pollard (1989), so by Lemma 2.4(i) and (ii) of Pakes and Pollard

(1989), $\left\{\sum_{t=1}^{T} A_{\tau_t}\left(\mathsf{w}_{1\tau_t}|\theta\right) : d\left(\theta,\theta_0\right) < \delta\right\}$ is VC subgraph with the envelope

$$1\left(q \leq \gamma \wedge \gamma_0\right)\|\mathbf{x}\| \sup_{d(\theta,\theta_0)<\delta} \sum_{t=1}^{T} \left\|\beta_1^0\left(\tau_t\right) - \beta_1\left(\tau_t\right)\right\|.$$

Similarly, $\left\{\sum_{t=1}^{T} B_{\tau_t}\left(\mathsf{w}_{2\tau_t}|\theta\right) : d\left(\theta,\theta_0\right) < \delta\right\}$, $\left\{\sum_{t=1}^{T} C_{\tau_t}\left(\mathsf{w}_{1\tau_t}|\theta\right) : d\left(\theta,\theta_0\right) < \delta\right\}$ and $\left\{\sum_{t=1}^{T} D_{\tau_t}\left(\mathsf{w}_{2\tau_t}|\theta\right) : d\left(\theta,\theta_0\right) < \delta\right\}$ are VC subgraph with the envelope

$$1(q > \gamma \vee \gamma_0)\|\mathbf{x}\| \sup_{d(\theta,\theta_0)<\delta} \sum_{t=1}^{T} \left\|\beta_2^0\left(\tau_t\right) - \beta_2\left(\tau_t\right)\right\|,$$
$$1(\gamma \wedge \gamma_0 < q \leq \gamma_0)\|\mathbf{x}\| \sup_{d(\theta,\theta_0)<\delta} \sum_{t=1}^{T} \left\|\beta_1^0\left(\tau_t\right) - \beta_2\left(\tau_t\right)\right\|,$$
$$1(\gamma_0 < q \leq \gamma \vee \gamma_0)\|\mathbf{x}\| \sup_{d(\theta,\theta_0)<\delta} \sum_{t=1}^{T} \left\|\beta_2^0\left(\tau_t\right) - \beta_1\left(\tau_t\right)\right\|,$$

respectively, so by Lemma 2.4(i) of Pakes and Pollard (1989), $\{m\left(w|\theta\right) - m\left(w|\theta_0\right) : d\left(\theta,\theta_0\right) < \delta\}$ is VC subgraph with the envelope

$$F = \|\mathbf{x}\| \sup_{d(\theta,\theta_0)<\delta} \sum_{t=1}^{T} \left\{1\left(q \leq \gamma \wedge \gamma_0\right)\left\|\beta_1^0\left(\tau_t\right) - \beta_1\left(\tau_t\right)\right\| + 1(q > \gamma \vee \gamma_0)\left\|\beta_2^0\left(\tau_t\right) - \beta_2\left(\tau_t\right)\right\|\right.$$
$$\left. +1\left(\gamma \wedge \gamma_0 < q \leq \gamma_0\right)\left\|\beta_1^0\left(\tau_t\right) - \beta_2\left(\tau_t\right)\right\| + 1\left(\gamma_0 < q \leq \gamma \vee \gamma_0\right)\left\|\beta_2^0\left(\tau_t\right) - \beta_1\left(\tau_t\right)\right\|\right\}.$$

From Theorem 2.14.2 of van der Vaart and Wellner (1996),

$$E\left[\sup_{d(\theta,\theta_0)<\delta} \left|\mathbb{G}_n\left(m\left(w|\theta\right) - m\left(w|\theta_0\right)\right)\right|\right] \leq C\sqrt{PF^2}.$$

By Assumptions D2 and D4, $\sqrt{PF^2} \leq C\delta$. So $\phi\left(\delta\right) = \delta$ in Corollary 3.2.6 of van der Vaart and Wellner (1996) and δ/δ^α is decreasing for all $1 < \alpha < 2$. Since $r_n^2\phi\left(\frac{1}{r_n}\right) = r_n$, $\sqrt{n}d\left(\widehat{\theta} - \theta_0\right) = O_P\left(1\right)$. By the definition of d, the result follows. ■

Lemma 4 *Under Assumptions D1-D7 and* $\|\delta_n\| \to 0$, $\sqrt{n}\|\delta_n\| \to \infty$, $a_n\left(\widehat{\gamma} - \gamma_0\right) = O_p(1)$, *and* $\sqrt{n}\left(\beta_T - \beta_T^0\right) = O_p(1)$.

Proof. Since δ_n depends on n, Corollary 3.2.6 of van der Vaart and Wellner (1996) cannot be used. Nevertheless, we can apply the proof idea of Theorem 3.2.5 in van der Vaart and Wellner (1996) to prove this result. Define $d_n\left(\theta,\theta_0\right) = \max\left\{\sqrt{n}\left\|\beta - \beta_0\right\|, a_n\left|\gamma - \gamma_0\right|\right\}$ for θ in a neighborhood of θ_0. For each n, the parameter space (minus the point θ_0) can be partitioned into the "shells" $S_{j,n} = \left\{\theta : 2^{j-1} < d_n\left(\theta,\theta_0\right) \leq 2^j\right\}$ with j ranging over the integers. Given an integer J,

$$P\left(d_n\left(\widehat{\theta},\theta_0\right) > 2^J\right) \leq \sum_{j\geq J, \left\|\beta_T - \beta_T^0\right\| < M\|\delta_n\|, |\gamma-\gamma_0|<\eta} P\left(\inf_{\theta\in S_{j,n}}\left(Q_n(\theta) - Q_n(\theta_0)\right) \leq 0\right) \tag{18}$$
$$+ P\left(2\left\|\beta_T - \beta_T^0\right\| \geq M\|\delta_n\|, 2\left|\gamma - \gamma_0\right| \geq \eta\right),$$

where $Q_n(\theta)$ is defined in (17), and M and η are small positive numbers. The second term on the right hand side of (18) converges to zero as $n \to \infty$ for every $\eta > 0$ and $M > 0$ by the Lemma 2, so we can concentrate

on the first term.

$$
\begin{aligned}
&P\left(\inf_{\theta\in S_{j,n}}\left(Q_n(\theta)-Q_n(\theta_0)\right)\le 0\right)\\
&\le P\left(\sup_{\theta\in S_{j,n}}\left|Q_n(\theta)-Q_n(\theta_0)-E\left[Q_n(\theta)-Q_n(\theta_0)\right]\right|\ge \inf_{\theta\in S_{j,n}}\left|E\left[Q_n(\theta)-Q_n(\theta_0)\right]\right|\right)\\
&\le \sum_{k=1}^{3} P\left(\sup_{\theta\in S_{j,n}}\left|T_k\left(\theta\right)-E\left[T_k\left(\theta\right)\right]\right|\ge \inf_{\theta\in S_{j,n}}\left|E\left[T_k\left(\theta\right)\right]\right|\right)\\
&\le \sum_{k=1}^{3} E\left[\sup_{\theta\in S_{j,n}}\left|T_k\left(\theta\right)-E\left[T_k\left(\theta\right)\right]\right|\right]\Big/\inf_{\theta\in S_{j,n}}\left|E\left[T_k\left(\theta\right)\right]\right|,
\end{aligned}
$$

where the last equality is from Markov's inequality, and $T_k(\theta)$, $k=1,2,3$, is defined in (17).

From Lemma 2, it is not hard to see that for $k=1,2$, $\inf_{\theta\in S_{j,n}}\left|E\left[T_k\left(\theta\right)\right]\right|\ge C\frac{2^{2j-2}}{a_n}$, and $\inf_{\theta\in S_{j,n}}\left|E\left[T_3\left(\theta\right)\right]\right|\ge C\frac{2^{j-1}}{a_n}$ given that $\widehat{\beta}_{\ell T}-\beta^0_{\ell T}=o_p(\|\delta_n\|)$. From the last lemma, for $k=1,2$,

$$
E\left[\sup_{\theta\in S_{j,n}}\left|T_k\left(\theta\right)-E\left[T_k\left(\theta\right)\right]\right|\right]\le C\frac{2^j/\sqrt{n}}{\sqrt{n}\delta_n'\delta_n}=C\frac{2^j}{a_n}.
$$

As to $T_3(\theta)$, applying a maximal inequality (e.g., Theorem 2.14.2 of van der Vaart and Wellner (1996)) we can show that

$$
E\left[\sup_{\theta\in S_{j,n}}\left|T_3\left(\theta\right)-E\left[T_3\left(\theta\right)\right]\right|\right]\le C\frac{\sqrt{\delta_n'\delta_n 2^j/a_n}}{\sqrt{n}\delta_n'\delta_n}\le C\frac{2^{j/2}}{a_n}.
$$

In summary,

$$
\begin{aligned}
&\sum_{j\ge J,\left\|\beta_T-\beta_T^0\right\|<M\|\delta_n\|,|\gamma-\gamma_0|<\eta} P\left(\sup_{\theta\in S_{j,n}}\left(Q_n\left(\theta\right)-Q_n\left(\theta_0\right)\right)\ge 0\right)\\
&\le C\sum_{j\ge J}\left(\frac{2^{j/2}}{a_n}\Big/\frac{2^{j-1}}{a_n}+\frac{2^j}{a_n}\Big/\frac{2^{2j-2}}{a_n}\right)\le C\sum_{j\ge J}\left(\frac{1}{2^{j/2}}+\frac{1}{2^j}\right),
\end{aligned}
$$

which can be made arbitrarily small by letting J large enough. ■

Lemma 5 *Under Assumption D, uniformly for $h=(u_T',v)'$ in any compact set of $\mathbb{R}^{2dT+1}$,*

$$
\begin{aligned}
&nP_n\left(m\left(\cdot\left|\beta_T^0+\frac{u}{\sqrt{n}},\gamma_0+\frac{v}{n}\right.\right)-m\left(\cdot\left|\beta_T^0,\gamma_0\right.\right)\right)\\
&=\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\rho_{\tau_t}\left(e_{1\tau_t i}-\frac{u'_{1\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{1\tau_t i}\right)\right]1(q_i\le\gamma_0)+\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\rho_{\tau_t}\left(e_{2\tau_t i}-\frac{u'_{2\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{2\tau_t i}\right)\right]1(q_i>\gamma_0)\\
&+D_{Tn}\left(v\right)+o_p\left(1\right),
\end{aligned}
$$

where $u_T=\left(u'_{1\tau_1},\cdots,u'_{1\tau_T},u'_{2\tau_1},\cdots,u'_{2\tau_T}\right)'\equiv(u'_{1T},u'_{2T})'\in\mathbb{R}^{2dT}$, and

$$
D_{Tn}\left(v\right)=\sum_{i=1}^{n}\overline{z}_{1Ti}1\left(\gamma_0+\frac{v}{n}<q_i\le\gamma_0\right)+\sum_{i=1}^{n}\overline{z}_{2Ti}1\left(\gamma_0<q_i\le\gamma_0+\frac{v}{n}\right).
$$

Proof. Note that

$$
\begin{aligned}
&nP_n\left(m\left(\cdot\left|\beta_T^0+\frac{u}{\sqrt{n}},\gamma_0+\frac{v}{n}\right.\right)-m\left(\cdot\left|\beta_T^0,\gamma_0\right.\right)\right)\\
&=\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\rho_{\tau_t}\left(e_{1\tau_t i}-\frac{u'_{1\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{1\tau_t i}\right)\right]1(q_i\le\gamma_0\wedge\gamma_0+\frac{v}{n})\\
&+\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\rho_{\tau_t}\left(e_{2\tau_t i}-\frac{u'_{2\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{2\tau_t i}\right)\right]1\left(q>\gamma_0+\frac{v}{n}\vee\gamma_0\right)\\
&+\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\rho_{\tau_t}\left(e_{1\tau_t i}+\mathbf{x}_i'\beta_1^0\left(\tau_t\right)-\mathbf{x}_i'\beta_2^0\left(\tau_t\right)-\frac{u'_{2\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{1\tau_t i}\right)\right]1\left(\gamma_0+\frac{v}{n}<q_i\le\gamma_0\right)\\
&+\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\rho_{\tau_t}\left(e_{2\tau_t i}+\mathbf{x}_i'\beta_2^0\left(\tau_t\right)-\mathbf{x}_i'\beta_1^0\left(\tau_t\right)-\frac{u'_{1\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{2\tau_t i}\right)\right]1\left(\gamma_0<q_i\le\gamma_0+\frac{v}{n}\right),
\end{aligned}
$$

so we need to show for $t=1,\cdots,T$,

$$
\begin{aligned}
&\sum_{i=1}^{n}\left[\rho_{\tau_t}\left(e_{1\tau_t i}-\tfrac{u'_{1\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{1\tau_t i}\right)\right]1(q_i\le\gamma_0\wedge\gamma_0+\tfrac{v}{n})\\
&=\sum_{i=1}^{n}\left[\rho_{\tau_t}\left(e_{1\tau_t i}-\tfrac{u'_{1\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{1\tau_t i}\right)\right]1(q_i\le\gamma_0)+o_p(1),\\
&\sum_{i=1}^{n}\left[\rho_{\tau_t}\left(e_{2\tau_t i}-\tfrac{u'_{2\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{2\tau_t i}\right)\right]1(q_i>\gamma_0+\tfrac{v}{n}\vee\gamma_0)\\
&=\sum_{i=1}^{n}\left[\rho_{\tau_t}\left(e_{2\tau_t i}-\tfrac{u'_{2\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{2\tau_t i}\right)\right]1(q_i>\gamma_0)+o_p(1),\\
&\sum_{i=1}^{n}\left[\rho_{\tau_t}\left(e_{1\tau_t i}+\mathbf{x}_i'\beta_1^0\left(\tau_t\right)-\mathbf{x}_i'\beta_2^0\left(\tau_t\right)-\tfrac{u'_{2\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{1\tau_t i}\right)\right]1(\gamma_0+\tfrac{v}{n}<q_i\le\gamma_0)\\
&=\sum_{i=1}^{n}\left[\rho_{\tau_t}\left(e_{1\tau_t i}+\mathbf{x}_i'\beta_1^0\left(\tau_t\right)-\mathbf{x}_i'\beta_2^0\left(\tau_t\right)\right)-\rho_{\tau_t}\left(e_{1\tau_t i}\right)\right]1(\gamma_0+\tfrac{v}{n}<q_i\le\gamma_0)+o_p(1),\\
&\sum_{i=1}^{n}\left[\rho_{\tau_t}\left(e_{2\tau_t i}+\mathbf{x}_i'\beta_2^0\left(\tau_t\right)-\mathbf{x}_i'\beta_1^0\left(\tau_t\right)-\tfrac{u'_{1\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{2\tau_t i}\right)\right]1\left(\gamma_0<q_i\le\gamma_0+\tfrac{v}{n}\right)\\
&=\sum_{i=1}^{n}\left[\rho_{\tau_t}\left(e_{2\tau_t i}+\mathbf{x}_i'\beta_2^0\left(\tau_t\right)-\mathbf{x}_i'\beta_1^0\left(\tau_t\right)\right)-\rho_{\tau_t}\left(e_{2\tau_t i}\right)\right]1\left(\gamma_0<q_i\le\gamma_0+\tfrac{v}{n}\right)+o_p(1).
\end{aligned}
$$

Actually, for $\tau\in\mathcal{T}$,

$$
\begin{aligned}
&\sum_{i=1}^{n}\left[\rho_{\tau}\left(e_{1\tau i}-\frac{u'_{1\tau}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau}\left(e_{1\tau i}\right)\right]\left[1(q_i\le\gamma_0\wedge\gamma_0+\frac{v}{n})-1(q_i\le\gamma_0)\right]\\
&\overset{(1)}{\le}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\|\mathbf{x}_i\right\|\left\|u_{1\tau}\right\|1\left(\gamma_0<q_i\le\gamma_0\wedge\gamma_0+\frac{v}{n}\right)\\
&\overset{(2)}{=}o_p\left(\sum_{i=1}^{n}1\left(\gamma_0<q_i\le\gamma_0\wedge\gamma_0+\frac{v}{n}\right)\right)\overset{(3)}{=}o_p(1)
\end{aligned}
$$

where (1) is from the Lipschitzity of $\rho_\tau(\cdot)$, (2) is from Assumption D7 which implies $\sup_{i\leq n}\|\mathbf{x}_i\| = o_p(n^{1/2})$, and (3) is from Assumption D4. Similarly,

$$\sum_{i=1}^{n}\left[\rho_\tau\left(e_{2\tau i}-\frac{u'_{2\tau}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_\tau\left(e_{2\tau i}\right)\right]\left(1\left(q_i>\gamma_0+\frac{v}{n}\vee\gamma_0\right)-1(q_i>\gamma_0)\right)=o_p(1),$$

$$\sum_{i=1}^{n}\left[\rho_\tau\left(e_{1\tau i}+\mathbf{x}_i'\beta_1^0\left(\tau\right)-\mathbf{x}_i'\beta_2^0\left(\tau\right)-\frac{u'_{2\tau}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_\tau\left(e_{1\tau i}+\mathbf{x}_i'\beta_1^0\left(\tau\right)-\mathbf{x}_i'\beta_2^0\left(\tau\right)\right)\right]1\left(\gamma_0+\frac{v}{n}<q_i\leq\gamma_0\right)=o_p(1),$$

$$\sum_{i=1}^{n}\left[\rho_\tau\left(e_{2\tau i}+\mathbf{x}_i'\beta_2^0\left(\tau\right)-\mathbf{x}_i'\beta_1^0\left(\tau\right)-\frac{u'_{1\tau}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_\tau\left(e_{2\tau i}+\mathbf{x}_i'\beta_2^0\left(\tau\right)-\mathbf{x}_i'\beta_1^0\left(\tau\right)\right)\right]1\left(\gamma_0<q_i\leq\gamma_0+\frac{v}{n}\right)=o_p(1).$$

■

Lemma 6 *Under Assumptions D1-D7 and $\|\delta_n\|\to 0$, $\sqrt{n}\|\delta_n\|\to\infty$, uniformly for $h=(u_T',v)'$ in any compact set of $\mathbb{R}^{2dT+1}$,*

$$\begin{aligned}
&nP_n\left(m\left(\cdot\left|\beta_T^0+\frac{u}{\sqrt{n}},\gamma_0+\frac{v}{a_n}\right.\right)-m\left(\cdot\left|\beta_T^0,\gamma_0\right.\right)\right)\\
&=\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\rho_{\tau_t}\left(e_{1\tau_t i}-\frac{u'_{1\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{1\tau_t i}\right)\right]1(q_i\leq\gamma_0)+\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\rho_{\tau_t}\left(e_{2\tau_t i}-\frac{u'_{2\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{2\tau_t i}\right)\right]1(q_i>\gamma_0)\\
&+C_{Tn}\left(v\right)+o_p\left(1\right),
\end{aligned}$$

where u_T is defined in the last lemma, and

$$C_{Tn}\left(v\right)=\begin{cases}\sum_{t=1}^{T}\delta'_{tn}\left[\sum_{i=1}^{n}\mathbf{x}_i\psi_{\tau_t}\left(e_{1\tau_t i}\right)1\left(\gamma_0+\frac{v}{a_n}<q_i\leq\gamma_0\right)\right]+\frac{f_q(\gamma_0)}{2}\frac{\pi_{1Tn}}{\delta'_n\delta_n}\left|v\right|, & \text{if } v\leq 0,\\ -\sum_{t=1}^{T}\delta'_{tn}\left[\sum_{i=1}^{n}\mathbf{x}_i\psi_{\tau_t}\left(e_{2\tau_t i}\right)1\left(\gamma_0<q_i\leq\gamma_0+\frac{v}{a_n}\right)\right]+\frac{f_q(\gamma_0)}{2}\frac{\pi_{2Tn}}{\delta'_n\delta_n}v, & \text{if } v>0.\end{cases}$$

Proof. We decompose $nP_n\left(m\left(\cdot\left|\beta_T^0+\frac{u}{\sqrt{n}},\gamma_0+\frac{v}{n}\right.\right)-m\left(\cdot\left|\beta_T^0,\gamma_0\right.\right)\right)$ as

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\rho_{\tau_t}\left(e_{1\tau_t i}-\frac{u'_{1\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{1\tau_t i}\right)\right]1(q_i\leq\gamma_0)\\
&+\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\rho_{\tau_t}\left(e_{2\tau_t i}-\frac{u'_{2\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{2\tau_t i}\right)\right]1\left(q>\gamma_0\right)\\
&+\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\rho_{\tau_t}\left(e_{1\tau_t i}+\mathbf{x}_i'\beta_1^0\left(\tau_t\right)-\mathbf{x}_i'\beta_2^0\left(\tau_t\right)-\frac{u'_{2\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\left(e_{1\tau_t i}-\frac{u'_{1\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)\right]1\left(\gamma_0+\frac{v}{n}<q_i\leq\gamma_0\right)\\
&+\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\rho_{\tau_t}\left(e_{2\tau_t i}+\mathbf{x}_i'\beta_2^0\left(\tau_t\right)-\mathbf{x}_i'\beta_1^0\left(\tau_t\right)-\frac{u'_{1\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)-\rho_{\tau_t}\left(e_{2\tau_t i}-\frac{u'_{2\tau_t}}{\sqrt{n}}\mathbf{x}_i\right)\right]1\left(\gamma_0<q_i\leq\gamma_0+\frac{v}{n}\right).
\end{aligned}$$

We need only to show the last two terms have an approximation of $C_{Tn}\left(v\right)$. Given that $\|\delta_n\|\to 0$ and $\sqrt{n}\|\delta_n\|\to\infty$, $u_{\ell\tau_t}/\sqrt{n}$ can be neglected. Now, from Knight (1998),

$$\begin{aligned}
&\rho_{\tau_t}\left(e_{1\tau_t i}+\mathbf{x}_i'\beta_1^0\left(\tau_t\right)-\mathbf{x}_i'\beta_2^0\left(\tau_t\right)\right)-\rho_{\tau_t}\left(e_{1\tau_t i}\right)\\
&=\psi_{\tau_t}\left(e_{1\tau_t i}\right)\delta'_{tn}\mathbf{x}_i+\int_0^{-\mathbf{x}_i'\delta_{tn}}\left(1(e_{1\tau_t i}<s)-1(e_{1\tau_t i}<0)\right)ds,
\end{aligned}$$

so the first term matches the first term of $C_{Tn}(v)$. As to the second term, note that

$$\sum_{t=1}^{T}\sum_{i=1}^{n}\int_{0}^{-\mathbf{x}_i'\delta_{tn}}\left(1(e_{1\tau_t i}<s)-1(e_{1\tau_t i}<0)\right)ds1\left(\gamma_0+v/a_n<q_i\leq\gamma_0\right)$$

has a mean

$$\sum_{t=1}^{T}\frac{f_q(\gamma_0)}{2}\frac{\delta_{tn}'}{\|\delta_n\|}E\left[f_{e_{1\tau_t}|x,q}(0|x,q)\mathbf{x}\mathbf{x}'|q=\gamma_0\right]\frac{\delta_{tn}}{\|\delta_n\|}+o(1),$$

and the deviation from the mean is uniformly small. ■

Lemma 7 *Under Assumption D, $D_{Tn}(v)\rightsquigarrow D_T(v)$ on any compact set of v.*

Proof. This proof includes two parts: (i) the finite-dimensional limit distributions of $D_{Tn}(v)$ are the same as specified in the theorem; (ii) the process $D_{Tn}(v)$ is stochastically equicontinuous.

Part (i): This is a direct corollary of the asymptotic limit of the characteristic function in the proof of Theorem 3, so omitted here to avoid repetition.

Part (ii): Without loss of generality, we prove the result only for $v>0$. Suppose $0<v_1<v_2$ are stopping times in a compact set; then for any $\epsilon>0$,

$$\begin{aligned}&P\left(\sup_{|v_2-v_1|<\delta}|D_{Tn}(v_2)-D_{Tn}(v_1)|>\epsilon\right)\overset{(1)}{\leq}P\left(\sum_{i=1}^{n}|\overline{z}_{2Ti}|\cdot\sup_{|v_2-v_1|<\delta}1\left(\gamma_0+\tfrac{v_1}{n}<q_i\leq\gamma_0+\tfrac{v_2}{n}\right)>\epsilon\right)\\&\overset{(2)}{\leq}\sum_{i=1}^{n}E\left[|\overline{z}_{2Ti}|\sup_{|v_2-v_1|<\delta}1\left(\gamma_0+\tfrac{v_1}{n}<q_i\leq\gamma_0+\tfrac{v_2}{n}\right)\right]\Big/\epsilon\overset{(3)}{\leq}\tfrac{C\delta}{\epsilon},\end{aligned}$$

where (1) is obvious, (2) is from Markov's inequality, and C in (3) can take $\overline{f}_q\sup_{\gamma_0<\gamma\leq\gamma_0+\epsilon}E\left[|\overline{z}_{2Ti}||\,q_i=\gamma\right]<\infty$ from Assumptions D4 and D7. ■

Lemma 8 *Under Assumptions D1-D7 and $\|\delta_n\|\to 0$, $\sqrt{n}\|\delta_n\|\to\infty$, $D_{Tn}(v)\rightsquigarrow C_T(v)$ on any compact set of v, where*

$$C_T(v)=\begin{cases}\sqrt{f_q(\gamma_0)}\sigma_T W_1(-v)+\frac{f_q(\gamma_0)}{2}\pi_{1T}|v|, & \text{if } v\leq 0,\\ \sqrt{f_q(\gamma_0)}\sigma_T W_2(v)+\frac{f_q(\gamma_0)}{2}\pi_{2T}|v|, & \text{if } v>0,\end{cases}$$

where $\pi_{\ell T}=\lim_{n\to\infty}\pi_{\ell Tn}/\delta_n'\delta_n$, and $\sigma_T^2=\lim_{n\to\infty}\sigma_{Tn}^2/\delta_n'\delta_n$.

Proof. Note that D_{Tn} has the same weak limit as C_{Tn} by Lemma 6. By similar arguments as in the last lemma, we need only check the stochastic equicontinuity of $D_{Tn}(v)$. Without loss of generality, we prove the result only for $v>0$. Suppose $v_1<v_2$ are positive numbers in a compact set; then for any $\epsilon>0$,

$$\begin{aligned}&P\left(\sup_{|v_2-v_1|<\delta}|D_{Tn}(v_2)-D_{Tn}(v_1)|>\epsilon\right)\overset{(1)}{\leq}P\left(\sum_{i=1}^{n}|\overline{z}_{2Ti}|\cdot\sup_{|v_2-v_1|<\delta}1\left(\gamma_0+\tfrac{v_1}{a_n}<q_i\leq\gamma_0+\tfrac{v_2}{a_n}\right)>\epsilon\right)\\&\overset{(2)}{\leq}\sum_{i=1}^{n}E\left[|\overline{z}_{2Ti}|^2\sup_{|v_2-v_1|<\delta}1\left(\gamma_0+\tfrac{v_1}{a_n}<q_i\leq\gamma_0+\tfrac{v_2}{a_n}\right)\right]\Big/\epsilon^2\\&\overset{(3)}{\leq}Cn\delta_n\delta_n'|v_2-v_1|E\left[\|\mathbf{x}_i\|^2|q_i=\gamma_0\right]f_q(\gamma_0)/\left(a_n\epsilon^2\right)\overset{(4)}{=}C|v_2-v_1|/\epsilon^2,\end{aligned}$$

where (1) is obvious, (2) is from Markov's inequality, (3) is from the Lipschitz continuity of $\rho_{\tau_t}(\cdot)$, $t=1,\cdots,T$, and (4) is from Assumptions D4 and D7. ■

Supplementary Materials

1. Asymptotics When q is the Only Covariate

In this section, we discuss the validity of Assumption D8 in the simple threshold regression with q being the only covariate. Suppose the population model is

$$y = \beta 1(q \leq \gamma) + e,\ q \sim U[0,1], e \sim N(0,1), \tag{19}$$

where e is independent of q, $\beta_0 = 1$ and $\gamma_0 = 0.5$. To simplify notations, denote the pdf and cdf of e as $f(\cdot)$ and $F(\cdot)$, respectively. We first consider the asymptotic distribution of $\widehat{\gamma}_\tau$. From Theorem 1,

$$n\left(\widehat{\gamma}_\tau - \gamma_0\right) \xrightarrow{d} \arg\min_v D_\tau(v),$$

where

$$D_\tau(v) = \begin{cases} \sum_{i=1}^{N_1(|v|)} z_{1\tau i} = \sum_{i=1}^{N_1(|v|)} \left|e_{\tau i}^- + \beta_0\right| - \left|e_{\tau i}^-\right| + (2\tau - 1)\beta_0, \text{ if } v \leq 0; \\ \sum_{i=1}^{N_2(v)} z_{2\tau i} = \sum_{i=1}^{N_2(v)} \left|e_{\tau i}^+ - \beta_0\right| - \left|e_{\tau i}^+\right| - (2\tau - 1)\beta_0, \text{ if } v > 0. \end{cases}$$

Here $\left\{e_{\tau i}^-, e_{\tau i}^+, i = 1, \cdots, N_1(\cdot), N_2(\cdot)\right\}$ are independent of each other, $e_{\tau i}^-$ and $e_{\tau i}^+$ follow the same distribution as $e_i - \xi_\tau$ with $\xi_\tau \equiv F^{-1}(\tau)$, and $N_1(\cdot)$ and $N_2(\cdot)$ are standard Poisson processes. This asymptotic distribution is also valid in a little more general model, $y = (1,q)\beta_1 1(q \leq \gamma) + (1,q)\beta_2 1(q > \gamma) + e$. We need only redefine $\beta_0 = (1,\gamma_0)(\beta_1^0 - \beta_2^0)$.

1.1. Distributions of $z_{1\tau}$ and $z_{2\tau}$

We now check the distributions of $z_{1\tau i}$ and $z_{2\tau i}$. First, the distribution of $z_{1\tau i}$ when $\beta = \beta_0$ is the same as that of $z_{2\tau i}$ when $\beta = -\beta_0$, so we need only consider the case of $\beta_0 > 0$.

$$z_{1\tau i} = \begin{cases} 2(\tau - 1)\beta_0, & \text{if } e_{\tau i}^- \leq -\beta_0; \\ 2\tau\beta_0 + 2e_{\tau i}^-, & \text{if } -\beta_0 < e_{\tau i}^- \leq 0; \\ 2\tau\beta_0, & \text{if } e_{\tau i}^- > 0; \end{cases}$$

and

$$z_{2\tau i} = \begin{cases} 2(1-\tau)\beta_0, & \text{if } e_{\tau i}^+ \leq 0; \\ 2(1-\tau)\beta_0 - 2e_{\tau i}^+, & \text{if } 0 < e_{\tau i}^+ \leq \beta_0; \\ -2\tau\beta_0, & \text{if } e_{\tau i}^+ > \beta_0; \end{cases}$$

have bounded supports. So the distribution of $z_{1\tau i}$ is

$$P\left(z_{1\tau i} \leq t\right) = \begin{cases} 0, & \text{if } t < 2(\tau-1)\beta_0; \\ P\left(e_i \leq \xi_\tau - \beta_0\right), & \text{if } t = 2(\tau-1)\beta_0; \\ P\left(e_i \leq \xi_\tau + \frac{t}{2} - \tau\beta_0\right), & \text{if } 2(\tau-1)\beta_0 < t < 2\tau\beta_0; \\ 1, & \text{if } t \geq 2\tau\beta_0; \end{cases}$$

with the density

$$f_{z_{1\tau}}(t) = \begin{cases} P\left(e_i \leq \xi_\tau - \beta_0\right), & \text{if } t = 2(\tau-1)\beta_0; \\ \frac{1}{2} f\left(\xi_\tau + \frac{t}{2} - \tau\beta_0\right), & \text{if } 2(\tau-1)\beta_0 < t < 2\tau\beta_0; \\ P\left(e_i > \xi_\tau\right) = 1 - \tau, & \text{if } t = 2\tau\beta_0; \end{cases}$$

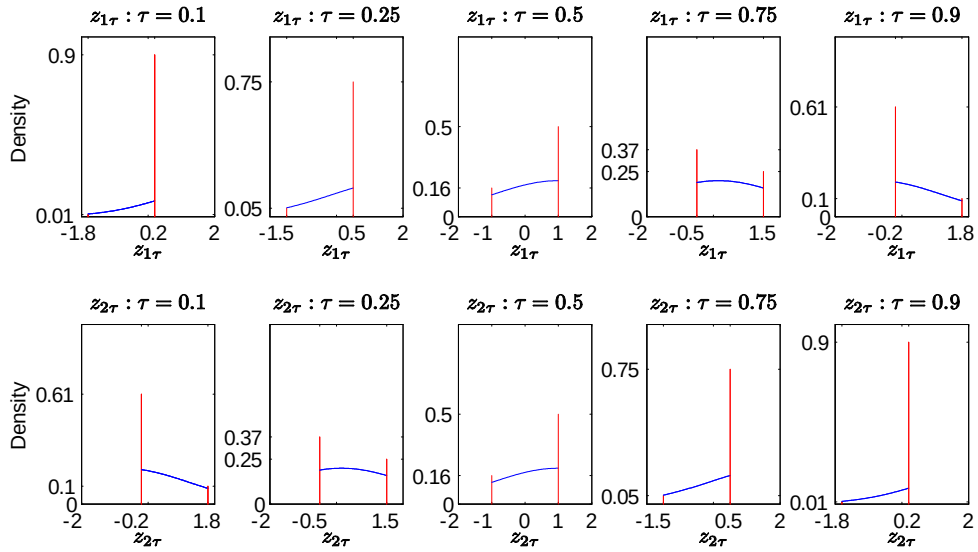

Figure 4: Densities of $z_{1\tau}$ and $z_{2\tau}$ for Different τ's

and the distribution of $z_{2\tau i}$ is

$$P\left(z_{2\tau i}\leq t\right)=\begin{cases}0, & \text{if } t<-2\tau\beta_0;\\ P\left(e_i>\xi_\tau+\beta_0\right), & \text{if } t=-2\tau\beta_0;\\ P\left(e_i\geq\xi_\tau-\frac{t}{2}+(1-\tau)\beta_0\right), & \text{if } -2\tau\beta_0<t<2(1-\tau)\beta_0;\\ 1, & \text{if } t\geq 2(1-\tau)\beta_0;\end{cases}$$

with the density

$$f_{z_{2\tau}}\left(t\right)=\begin{cases}P\left(e_i>\xi_\tau+\beta_0\right), & \text{if } t=-2\tau\beta_0;\\ \frac{1}{2}f\left(\xi_\tau-\frac{t}{2}+(1-\tau)\beta_0\right), & \text{if } -2\tau\beta_0<t<2(1-\tau)\beta_0;\\ P\left(e_i\leq\xi_\tau\right)=\tau, & \text{if } t=2(1-\tau)\beta_0.\end{cases}$$

Note that there are two point masses in the distributions of $z_{1\tau}$ and $z_{2\tau}$, so Assumption D8 does not hold. For different τ's, the distances between the two point masses are the same: $2\beta_0$. The differences are the locations and magnitudes of the point masses. The distributions of $z_{1\tau i}$ and $z_{2\tau i}$ are shown in Figure 4. From Figure 4, $z_{1\tau i}$ has a different distribution from $z_{2\tau i}$ unless $\tau=0.5$. The distributions of $z_{1\tau}$ and $z_{2\tau}$ are not symmetric even if the distribution of e is symmetric and when $\tau=0.5$. When e is symmetric, the distribution of $z_{1\tau}$ is the same as $z_{2(1-\tau)}$ but different from $z_{1(1-\tau)}$.

1.2. Distributions of z_{1T} and z_{2T}

$z_{\ell T}=\sum_{t=1}^{T}z_{\ell\tau_t}$. Given that the distribution of $z_{\ell\tau_t}$ has two point masses which are at different locations for different τ_t's, we expect the distribution of $z_{\ell T}$ to have $2T$ point masses. To avoid the arbitrariness in

the selection of τ_t, we consider the limit case as $T \to \infty$. In this limit case,

$$z_\ell \equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} z_{\ell\tau_t} \to \frac{1}{|\mathcal{T}|}\int_{\mathcal{T}} z_{\ell\tau} d\tau$$

where $\tau_{t+1} - \tau_t = |\mathcal{T}|/T$, $\mathcal{T} = [\underline{\tau}, \overline{\tau}]$, and $|\mathcal{T}| = \overline{\tau} - \underline{\tau}$ is the length of $\mathcal{T}$.

Based on the distributions of $z_{1\tau}$ and $z_{2\tau}$, we can derive the distributions of z_1 and z_2. Note that the distributions of $z_{\ell\tau}$ for different τ's are correlated, and the only randomness is from e. As a result,

$$P(z_1 \le t) = P\left(\int_{\mathcal{T}} \{|e - \xi_\tau + \beta_0| - |e - \xi_\tau| + (2\tau - 1)\beta_0\} d\tau \le |\mathcal{T}|\, t\right).$$

Note that as $e \le \xi_\tau - \beta_0$, the integrand equals $2(\tau-1)\beta_0$; as $\xi_\tau - \beta_0 < e \le \xi_\tau$, the integrand equals $2\tau\beta_0 + 2(e - \xi_\tau)$; as $e > \xi_\tau$, the integrand equals $2\tau\beta_0$. So we can divide the domain of e into five areas: $e \le \xi_{\underline{\tau}} - \beta_0, \xi_{\underline{\tau}} - \beta_0 < e \le \xi_{\underline{\tau}}, \xi_{\underline{\tau}} < e \le \xi_{\overline{\tau}} - \beta_0$, $\xi_{\overline{\tau}} - \beta_0 < e \le \xi_{\overline{\tau}}$ and $e > \xi_{\overline{\tau}}$, and the integrations for e on the five areas are

$$\begin{aligned}
&\textstyle\int_{\mathcal{T}} 2(\tau-1)\beta_0 d\tau = \beta_0\left[\overline{\tau}^2 - \underline{\tau}^2 - 2(\overline{\tau} - \underline{\tau})\right],\\
&\textstyle\int_{\underline{\tau}}^{F(e+\beta_0)} [2\tau\beta_0 + 2(e - \xi_\tau)]\, d\tau + \int_{F(e+\beta_0)}^{\overline{\tau}} 2(\tau-1)\beta_0 d\tau\\
&= \beta_0\left(F(e+\beta_0)^2 - \underline{\tau}^2\right) + 2e\left(F(e+\beta_0) - \underline{\tau}\right) - 2\textstyle\int_{\underline{\tau}}^{F(e+\beta_0)} \xi_\tau d\tau + \beta_0\left[\overline{\tau}^2 - F(e+\beta_0)^2 - 2(\overline{\tau} - F(e+\beta_0))\right]\\
&= \beta_0\left(\overline{\tau}^2 - \underline{\tau}^2\right) - 2(\underline{\tau} e + \overline{\tau}\beta_0) + 2F(e+\beta_0)(\beta_0 + e) - 2\textstyle\int_{\underline{\tau}}^{F(e+\beta_0)} \xi_\tau d\tau,\\
&\textstyle\int_{\underline{\tau}}^{F(e)} 2\tau\beta_0 d\tau + \int_{F(e)}^{F(e+\beta_0)} [2\tau\beta_0 + 2(e - \xi_\tau)]\, d\tau + \int_{F(e+\beta_0)}^{\overline{\tau}} 2(\tau-1)\beta_0 d\tau\\
&= \beta_0\left(F(e+\beta_0)^2 - \underline{\tau}^2\right) + 2e\left(F(e+\beta_0) - F(e)\right) - 2\textstyle\int_{F(e)}^{F(e+\beta_0)} \xi_\tau d\tau + \beta_0\left[\overline{\tau}^2 - F(e+\beta_0)^2 - 2(\overline{\tau} - F(e+\beta_0))\right]\\
&= \beta_0\left(\overline{\tau}^2 - \underline{\tau}^2\right) - 2(eF(e) + \overline{\tau}\beta_0) + 2F(e+\beta_0)(\beta_0 + e) - 2\textstyle\int_{F(e)}^{F(e+\beta_0)} \xi_\tau d\tau,\\
&\textstyle\int_{\underline{\tau}}^{F(e)} 2\tau\beta_0 d\tau + \int_{F(e)}^{\overline{\tau}} [2\tau\beta_0 + 2(e - \xi_\tau)]\, d\tau = \beta_0\left(\overline{\tau}^2 - \underline{\tau}^2\right) + 2e(\overline{\tau} - F(e)) - 2\textstyle\int_{F(e)}^{\overline{\tau}} \xi_\tau d\tau,\\
&\textstyle\int_{\mathcal{T}} 2\tau\beta_0 d\tau = \beta_0\left(\overline{\tau}^2 - \underline{\tau}^2\right),
\end{aligned}$$

respectively, where the integration on the third area degenerates to

$$\int_{\mathcal{T}} [2\tau\beta_0 + 2(e - \xi_\tau)]\, d\tau = \beta_0\left(\overline{\tau}^2 - \underline{\tau}^2\right) + 2e(\overline{\tau} - \underline{\tau}) - 2\int_{\mathcal{T}} \xi_\tau d\tau$$

if $\xi_{\overline{\tau}} - \beta_0 < \xi_{\underline{\tau}}$. It is easy to see that z_1 has a point mass $F\left(\xi_{\underline{\tau}} - \beta_0\right)$ at $\beta_0\left[\overline{\tau}^2 - \underline{\tau}^2 - 2(\overline{\tau} - \underline{\tau})\right]/|\mathcal{T}|$ and a point mass $1 - \overline{\tau}$ at $\beta_0\left(\overline{\tau}^2 - \underline{\tau}^2\right)/|\mathcal{T}|$, and is continuously distributed on other area as a complicated function of e. Similarly, we can get the five areas for z_2 and the integrations on these areas. It is easy to see that z_2 has a point mass $\underline{\tau}$ at $-\beta_0\left[\overline{\tau}^2 - \underline{\tau}^2 - 2(\overline{\tau} - \underline{\tau})\right]/|\mathcal{T}|$ and a point mass $1 - F(\xi_{\overline{\tau}} + \beta_0)$ at $-\beta_0\left(\overline{\tau}^2 - \underline{\tau}^2\right)/|\mathcal{T}|$, and is continuously distributed on other area. For both z_1 and z_2, the point masses go to zero when $\underline{\tau}$ goes to zero and $\overline{\tau}$ goes to 1.

Figure 5 shows the distributions of z_1 and z_2 when $\mathcal{T} = [0.1, 0.9]$ based on 100000 simulated draws of e. It seems that z_1 and z_2 has the same distribution. Compared with $z_{\ell\tau}$, z_ℓ has less point masses and more density on the positive axis, which implies that $\widehat{\gamma}$ is more efficient than $\widehat{\gamma}_\tau$. To compare with the MLE, we also impose the densities of z_1 and z_2 in the MLE on Figure 5. Since the MLE is maximizing the objective function while the IQTRE is minimizing the objective function, the densities of z_1 and z_2 associated with the MLE in Figure 5 are actually the densities of $-z_1$ and $-z_2$ in the MLE. Note also that the MLE is the same as the LSE in this simple case, so $-z_1$ and $-z_2$ in the MLE are the same as z_1 and z_2 in the LSE. From Figure 5, the distributions of z_1 and z_2 in the MLE are more spreading than those in the IQTRE.

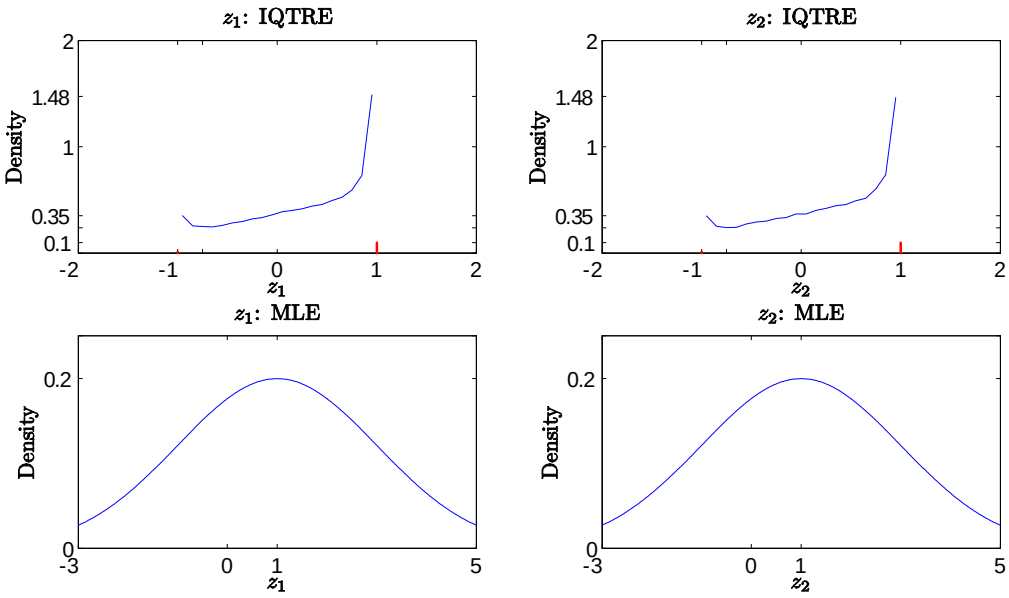

Figure 5: Comparison of the Densities of z_1 and z_2 in the IQTRE and MLE

2. Conditions to Guarantee the Uniqueness of $\arg\min_v D(v)$

From the above discussion, either $z_{\ell\tau}$ or z_ℓ has two point masses in their distribution when q is the only covariate, so $\arg\min_v D(v)$ may not be unique in this case. To simplify notations, we discuss the uniqueness of $\arg\min_v D(v)$ for a generic compound Poisson process

$$D(v) = \begin{cases} \sum_{i=1}^{N_1(|v|)} z_{1i}, \text{ if } v \leq 0; \\ \sum_{i=1}^{N_2(v)} z_{2i}, \text{ if } v > 0. \end{cases}$$

We first show that when the distribution of z_ℓ is absolutely continuous, $\arg\min_v D(v)$ is unique. Note that

$$\begin{aligned} P(E) &\equiv P\left(D(v) \text{ has at least two minimums}\right) \\ &= P\left(\begin{array}{c} \sum_{i=1}^{K} z_{1i} = \sum_{i=1}^{L} z_{1i}, \ \sum_{i=1}^{K} z_{2i} = \sum_{i=1}^{L} z_{2i} \text{ for some } K = 0, 1, \cdots, L = 0, 1, 2, \cdots, \text{ and } K \neq L, \\ \sum_{i=1}^{K} z_{1i} = \sum_{i=1}^{L} z_{2i} \text{ for some } K = 1, 2, \cdots, \text{ and } L = 1, 2, \cdots \end{array} \right) \\ &= P\left(E_1 \cup E_2 \cup E_3\right), \end{aligned}$$

where

$$E_1 = \left\{ \sum_{i=K}^{L} z_{1i} = 0 \text{ for some } K = 1, \cdots, L = 1, 2, \cdots, \text{ and } K \leq L \right\},$$
$$E_2 = \left\{ \sum_{i=K}^{L} z_{2i} = 0 \text{ for some } K = 1, \cdots, L = 1, 2, \cdots, \text{ and } K \leq L \right\},$$
$$E_3 = \left\{ \sum_{i=1}^{K} z_{1i} = \sum_{i=1}^{L} z_{2i} \text{ for some } K = 1, 2 \cdots \text{ and } L = 1, 2, \cdots \right\}.$$

A sufficient condition for this probability being zero is that the probabilities of all three events are zero. When the distribution of $z_{\ell i}$ is absolutely continuous, the distribution of $\sum_{i=K}^{L} z_{\ell i}$ is absolutely continuous for any $K = 1, 2, \cdots, L = 1, 2, \cdots$, and $K \leq L$, so either of these three events are union of countable zero-probability events and has probability zero.

If $z_{\ell i}$ has discrete components in its distribution, a sufficient condition for $P(E) = 0$ is more messy. Let's start from easier cases. Suppose first that $z_{\ell i}$ has only one discrete component, say, d_ℓ. For E_1 and E_2, unless $d_\ell = 0$, their probabilities are zero. $d_\ell = 0$ is the relevant case in bootstrapping threshold regression, see Yu (2013a). As to E_3, we require $Kd_1 \neq Ld_2$ for any $K = 1, 2 \cdots$ and $L = 1, 2, \cdots$, that is, $d_1/d_2 \neq r$ for r being any positive rational number. This obviously excludes the case that both d_1 and d_2 are rationals. In summary, a sufficient condition for $P(E) = 0$ when $z_{\ell i}$ has only one discrete component is that

$$d_\ell \neq 0 \text{ and } d_1/d_2 \neq r \text{ for } r \text{ being any positive rational number.}$$

Now, suppose $z_{\ell i}$ has two discrete components, say, $d_\ell^{(k)}$, $k = 1, 2$. This case is relevant in the simple example above and in regression discontinuity designs with unknown discontinuity points (see footnote 35 of Porter and Yu (2011)). For E_1 and E_2, we require $Ad_\ell^{(1)} + Bd_\ell^{(2)} \neq 0$ for any $A, B = 0, 1, \cdots$ and both A and B are not zero. This is equivalent to $d_\ell^{(k)} \neq 0$, $k = 1, 2$ and $d_\ell^{(1)}/d_\ell^{(2)} \neq r$ for r being any negative rational number. For E_3, we require $Kd_1^{(k)} \neq Ld_2^{(l)}$ for any $K = 1, 2 \cdots$, $L = 1, 2, \cdots$, $k = 1, 2$, and $l = 1, 2$, which is equivalent to $d_1^{(k)} \neq rd_2^{(l)}$ for r being any positive rational number, $k = 1, 2$, and $l = 1, 2$. In summary, a sufficient condition for $P(E) = 0$ when $z_{\ell i}$ has two discrete components is that

$$d_\ell^{(k)} \neq 0,\ d_\ell^{(1)}/d_\ell^{(2)} \neq r \text{ for } r \text{ being any negative rational number,}$$
$$d_1^{(k)}/d_2^{(l)} \neq r \text{ for } r \text{ being any positive rational number, } k = 1, 2 \text{ and } l = 1, 2.$$

For the IQTRE and all setups in Figure 4, this condition does not hold. It is quite possible to derive general conditions to guarantee $P(E) = 0$ following the logic above, but we do not delve into it here. Nevertheless, we mention that $\arg\max_v \widetilde{S}_1(v)$ in Theorem 6.1 of Lee and Seo (2008) is not unique.

3. Symmetry of the Distribution of $\arg\min_v D_\tau(v)$

In this section, we briefly discuss the symmetry of Z_τ when $\mathbf{x}$ may include other nonconstant covariates, where $Z_\tau \equiv \arg\min_v D_\tau(v)$. Suppose $\mathbf{x}$ does not include q; otherwise, the effect of q in $z_{\ell\tau}$ is absorbed in the constant term as $z_{\ell\tau}$ is defined as the conditional distribution given $q = \gamma_0$. When q is independent of $(x', e_{\ell\tau})'$,

$$z_{1\tau} = \overline{z}_{1\tau} = |e_{1\tau} + \mathbf{x}'(\beta_{10} - \beta_{20})| - |e_{1\tau}| + (2\tau - 1)\mathbf{x}'(\beta_{10} - \beta_{20}),$$
$$z_{2\tau} = \overline{z}_{2\tau} = |e_{2\tau} - \mathbf{x}'(\beta_{10} - \beta_{20})| - |e_{2\tau}| - (2\tau - 1)\mathbf{x}'(\beta_{10} - \beta_{20}).$$

Define $\mathbf{x}'(\beta_{10} - \beta_{20})$ as ζ, and suppose the joint distribution of $(x', e_{\ell\tau})'$ is continuous; then the distribution of $z_{1\tau}$ is

$$\begin{aligned}
&P(z_{1\tau} \le t) \\
&= P\left(e_{1\tau} + \zeta > 0, e_{1\tau} > 0, \zeta \le \tfrac{t}{2\tau}\right) + P\left(e_{1\tau} + \zeta > 0, e_{1\tau} \le 0, e_{1\tau} \le \tfrac{t}{2} - \tau\zeta\right) \\
&\quad + P\left(e_{1\tau} + \zeta \le 0, e_{1\tau} > 0, e_{1\tau} \ge (\tau - 1)\zeta - \tfrac{t}{2}\right) + P\left(e_{1\tau} + \zeta \le 0, e_{1\tau} \le 0, \zeta \ge \tfrac{t}{2(\tau-1)}\right) \\
&\overset{(*)}{=} P\left(-e_{1\tau} < \zeta \le \tfrac{t}{2\tau}, e_{1\tau} > 0\right) + P\left(-e_{1\tau} < \zeta \le \tfrac{t}{2\tau} - \tfrac{e_{1\tau}}{\tau}, e_{1\tau} \le 0\right) \\
&\quad + P\left(\tfrac{e_{1\tau}}{\tau-1} + \tfrac{t}{2(\tau-1)} \le \zeta \le -e_{1\tau}, e_{1\tau} > 0\right) + P\left(\tfrac{t}{2(\tau-1)} \le \zeta \le -e_{1\tau}, e_{1\tau} \le 0\right), \\
&\overset{(**)}{=} P\left(e_{1\tau} > \max\{-\zeta, 0\}, \zeta \le \tfrac{t}{2\tau}\right) + P\left(-\zeta < e_{1\tau} \le \min\left\{0, \tfrac{t}{2} - \tau\zeta\right\}\right) \\
&\quad + P\left(\max\left\{0, (\tau - 1)\zeta - \tfrac{t}{2}\right\} < e_{1\tau} \le -\zeta\right) + P\left(e_{1\tau} \le \min\{0, -\zeta\}, \zeta \ge \tfrac{t}{2(\tau-1)}\right),
\end{aligned}$$

and the distribution of $z_{2\tau}$ is

$$\begin{aligned}
&P(z_{2\tau} \le t) \\
&= P\left(e_{2\tau} - \zeta > 0, e_{1\tau} > 0, \zeta \ge -\tfrac{t}{2\tau}\right) + P\left(e_{2\tau} - \zeta > 0, e_{2\tau} \le 0, e_{2\tau} \le \tfrac{t}{2} + \tau\zeta\right) \\
&\quad + P\left(e_{2\tau} - \zeta \le 0, e_{2\tau} > 0, e_{2\tau} \ge -(\tau - 1)\zeta - \tfrac{t}{2}\right) + P\left(e_{2\tau} - \zeta \le 0, e_{2\tau} \le 0, \zeta \le \tfrac{t}{2(1-\tau)}\right) \\
&\overset{(*)}{=} P\left(-\tfrac{t}{2\tau} \le \zeta < e_{2\tau}, e_{2\tau} > 0\right) + P\left(\tfrac{e_{2\tau}}{\tau} - \tfrac{t}{2\tau} \le \zeta < e_{2\tau}, e_{2\tau} \le 0\right) \\
&\quad + P\left(e_{2\tau} \le \zeta \le \tfrac{e_{2\tau}}{1-\tau} + \tfrac{t}{2(1-\tau)}, e_{2\tau} > 0\right) + P\left(e_{2\tau} \le \zeta \le \tfrac{t}{2(1-\tau)}, e_{2\tau} \le 0\right). \\
&\overset{(**)}{=} P\left(e_{2\tau} \le \min\{0, \zeta\}, \zeta \le \tfrac{t}{2(1-\tau)}\right) + P\left(\max\left\{0, (1 - \tau)\zeta - \tfrac{t}{2}\right\} < e_{2\tau} \le \zeta\right) \\
&\quad + P\left(\zeta < e_{2\tau} \le \min\left\{0, \tfrac{t}{2} + \tau\zeta\right\}\right) + P\left(e_{2\tau} > \max\{0, \zeta\}, \zeta \ge -\tfrac{t}{2\tau}\right).
\end{aligned}$$

To simplify these distributions, suppose $e_{1\tau} = e_{2\tau} = e_\tau$ in the following discussion.

From Appendix D of Yu (2012), Z_τ is symmetric if and only if $P(z_{1\tau} \le t) = P(z_{2\tau} \le t)$ for all t. From $(*)$, if $\zeta|e_\tau$ is symmetric about zero, then $P(z_{1\tau} \le t) = P(z_{2\tau} \le t)$. From $(**)$, if $e_\tau|\zeta$ is symmetric about zero, then $P(z_{1\tau} \le t) = P\left(z_{2(1-\tau)} \le t\right)$; especially, $P(z_{1,0.5} \le t) = P(z_{2,0.5} \le t)$. If we further assume that ζ is independent of e, then $\zeta|e_\tau$ and $e_\tau|\zeta$ can be replaced by ζ and e_τ, respectively. Based on these facts, we can understand the distributions in Figure 4. For $\tau \neq 0.5$, since ζ is a point mass at $\beta \neq 0$, the distributions of $z_{1\tau}$ and $z_{2\tau}$ can not be the same. When $\tau = 0.5$, since $e_{0.5}$ follows $N(0, 1)$ which is symmetric, the distributions of $z_{1\tau}$ and $z_{2\tau}$ are the same.

Bai (1995) claims when $\mathbf{x}$ includes a constant, symmetry of $Z_{0.5}$ requires the symmetry of $e_{0.5}$. This is not right. For example, suppose $\mathbf{x} = (1, \varepsilon)'$ where ε follows $N(-\frac{\beta_{11}-\beta_{21}}{\beta_{12}-\beta_{22}}, \frac{1}{(\beta_{12}-\beta_{22})^2})$; then $\zeta = (\beta_{11} - \beta_{21}) + (\beta_{12} - \beta_{22})\varepsilon$ which follows $N(0, 1)$. From the above analysis, $P(z_{1\tau} \le t) = P(z_{2\tau} \le t)$, which guarantees the symmetry of $Z_{0.5}$.

4. Construction of the SEBE of γ and the NPI

The following algorithm is adapted from Yu (2008).

Step 1: Get the IQTRE of γ ($\widehat{\gamma}$), the LADE of β ($\widehat{\beta}_{0.5}$), and the corresponding residuals $\{\widehat{e}_i\}_{i=1}^n$ in model (1).

Step 2: Get a uniformly consistent estimator of the joint density of $(e_\ell, x', q)'$, $f_\ell(e_\ell, x, q)$, based on $\{\widehat{e}_i, x_i, q_i\}_{i=1}^n$ by kernel smoothing, and denote the estimator as $\widehat{f}_\ell(e_\ell, x, q)$.

Step 3: Define the SEBE of γ as

$$\widehat{\gamma}_{SEB} = \arg\min_t \int_\Gamma l_n(t-\gamma)\widehat{\mathcal{L}}_n(\gamma)\pi(\gamma)\, d\gamma.$$

where $l_n(t-\gamma) = l(n(t-\gamma))$ is the loss function of γ, $\pi(\gamma)$ is the prior of γ, e.g., $\pi(\gamma)$ can be the uniform distribution on $(q_{\min}, q_{\max})$ with $q_{\min}$ ($q_{\max}$) being the minimum (maximum) of $\{q_i\}_{i=1}^n$, and

$$\begin{aligned}\widehat{\mathcal{L}}_n(\gamma) &= \prod_{i=1}^n \left[\widehat{f}_1\left(y_i - \mathbf{x}_i'\widehat{\beta}_1, x_i, q_i\right) 1(q_i \le \gamma) + \widehat{f}_2\left(y_i - \mathbf{x}_i'\widehat{\beta}_2, x_i, q_i\right) 1(q_i > \gamma)\right] \\ &= \exp\left\{\sum_{i=1}^n 1(q_i \le \gamma) \ln \widehat{f}_1\left(y_i - \mathbf{x}_i'\widehat{\beta}_1, x_i, q_i\right) + \sum_{i=1}^n 1(q_i > \gamma) \ln \widehat{f}_2\left(y_i - \mathbf{x}_i'\widehat{\beta}_2, x_i, q_i\right)\right\} \\ &\equiv \exp\left\{\widehat{L}_n(\gamma)\right\},\end{aligned}$$

is the estimated likelihood function.

Step 4: Based on a MCMC algorithm, draw a Markov chain

$$S = \left(\gamma^{(1)}, \cdots, \gamma^{(B)}\right)$$

whose marginal density is approximately the posterior distribution

$$\widehat{p}_n(\gamma) = \frac{\exp\left\{\widehat{L}_n(\gamma)\right\}\pi(\gamma)}{\int_\Gamma \exp\left\{\widehat{L}_n(\widetilde{\gamma})\right\}\pi(\widetilde{\gamma})\, d\widetilde{\gamma}},$$

Then $\widehat{\gamma}_{SEB}$ is the mean of S when $l(v) = v^2$ and $\widehat{\gamma}_{SEB}$ is the median of S when $l(v) = |v|$. Also, the $100(1-\alpha)\%$ NPI of γ can be constructed by picking out the $\alpha/2$ and $1-\alpha/2$ quantile of S.

In the SEB procedure above, a key step is to estimate the likelihood function. For the simulation study in Section 6, we use the following algorithm.

Step 1: Obtain $\widehat{\gamma}$, and the associated $\widehat{\beta}_{\ell,0.5}$. Then the residuals $\widehat{e}_{1,0.5,i} = y_i - \mathbf{x}_i'\widehat{\beta}_{1,0.5}$ when $q_i \le \widehat{\gamma}$ and $\widehat{e}_{2,0.5,i} = y_i - \mathbf{x}_i'\widehat{\beta}_{2,0.5}$ when $q_i > \widehat{\gamma}$.

Step 2: Since $e_{\ell,0.5,i} = \sigma_\ell(e_i - \xi_{0.5})$, estimate $\sigma_{\ell 0}^2$ by

$$\begin{aligned}\widehat{\sigma}_1^2 &= \sum\nolimits_{i=1}^n \left(\widehat{e}_{1,0.5,i} - \overline{\widehat{e}}_{1,0.5}\right)^2 1(q_i \le \widehat{\gamma}) \Big/ \sum\nolimits_{i=1}^n 1(q_i \le \widehat{\gamma}), \\ \widehat{\sigma}_2^2 &= \sum\nolimits_{i=1}^n \left(\widehat{e}_{2,0.5,i} - \overline{\widehat{e}}_{2,0.5}\right)^2 1(q_i > \widehat{\gamma}) \Big/ \sum\nolimits_{i=1}^n 1(q_i > \widehat{\gamma}).\end{aligned}$$

where $\overline{\widehat{e}}_{1,0.5} = \sum_{i=1}^n \widehat{e}_{1,0.5,i} 1(q_i \le \widehat{\gamma}) / \sum_{i=1}^n 1(q_i \le \widehat{\gamma})$, and $\overline{\widehat{e}}_{2,0.5} = \sum_{i=1}^n \widehat{e}_{2,0.5,i} 1(q_i > \widehat{\gamma}) / \sum_{i=1}^n 1(q_i > \widehat{\gamma})$.

Step 3: Get empirical counterparts of $e_i - \xi_{0.5}$,

$$\widehat{e}_i - \xi_{0.5} = \widehat{e}_{1,0.5,i}/\widehat{\sigma}_1 \text{ if } q_i \leq \widehat{\gamma},$$
$$\widehat{e}_i - \xi_{0.5} = \widehat{e}_{2,0.5,i}/\widehat{\sigma}_2 \text{ if } q_i > \widehat{\gamma}.$$

Step 4: Estimate the density of $e_i - \xi_{0.5}$ by kernel smoothing based on $\widehat{e}_i - \xi_{0.5}$ and denote the estimator as $\widehat{f}_e(\cdot)$.

Step 5: The estimated likelihood is

$$\widehat{\mathcal{L}}_n(\gamma) = \prod_{i=1}^{n} \left[\frac{1}{\widehat{\sigma}_1} \widehat{f}_e \left(\frac{y_i - \mathbf{x}_i' \widehat{\beta}_{1,0.5}}{\widehat{\sigma}_1} \right) 1(q_i \leq \gamma) + \frac{1}{\widehat{\sigma}_2} \widehat{f}_e \left(\frac{y_i - \mathbf{x}_i' \widehat{\beta}_{2,0.5}}{\widehat{\sigma}_2} \right) 1(q_i > \gamma) \right].$$

In kernel smoothing of Step 4, we use the Mablab function kde.m provided in Botev et al. (2010).

Additional References

Botev, Z.I. et al., 2010, Kernel Density Estimation via Diffusion, *The Annals of Statistics*, 38, 2916-2957.

Lee, S. and M.H. Seo, 2008, Semiparametric Estimation of a Binary Response Model with a Change-point Due to a Covariate Threshold, *Journal of Econometrics*, 144, 492-499.

Porter, J. and P. Yu, 2011, Regression Discontinuity with Unknown Discontinuity Points: Testing and Estimation, Mimeo, Department of Economics, University of Auckland.